THE CAPCO INSTITUTE

JOURNAL

OF FINANCIAL TRANSFORMATION

CLOUD

#55 MAY 2022

CAPCO
a wipro company

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

CLOUD

CRYPTO

CYBER

DEAR READER,

Welcome to edition 55 of the Capco Institute Journal of Financial Transformation. Our central theme is cloud computing, which has transformed from an efficiency initiative for our clients, to an indispensable growth driver for financial services.

The pandemic has changed consumer expectations, with consumers now demanding 24/7 access to their financial resources from anywhere, as well as hyper-personalized products that reflect their lifestyle choices.

In this edition of the Journal, we explore the power of cloud and its potential applications through the lens of a joint Capco and Wipro global study, and take a deeper look at the financial services data collected in Wipro FullStride Cloud Services' 2021 Global Survey. The survey was focused on perceptions of cloud and its importance to business strategy from over 1,300 C-level executives and key decision-makers across 11 industries.

The study indicates that cloud is becoming ever more intelligent, hyperconnected, and pervasive, and enables companies to offer their end users the personalized, user-centric experience that they have come to expect. It's clear that only the financial services firms that can successfully leverage cloud, will thrive.

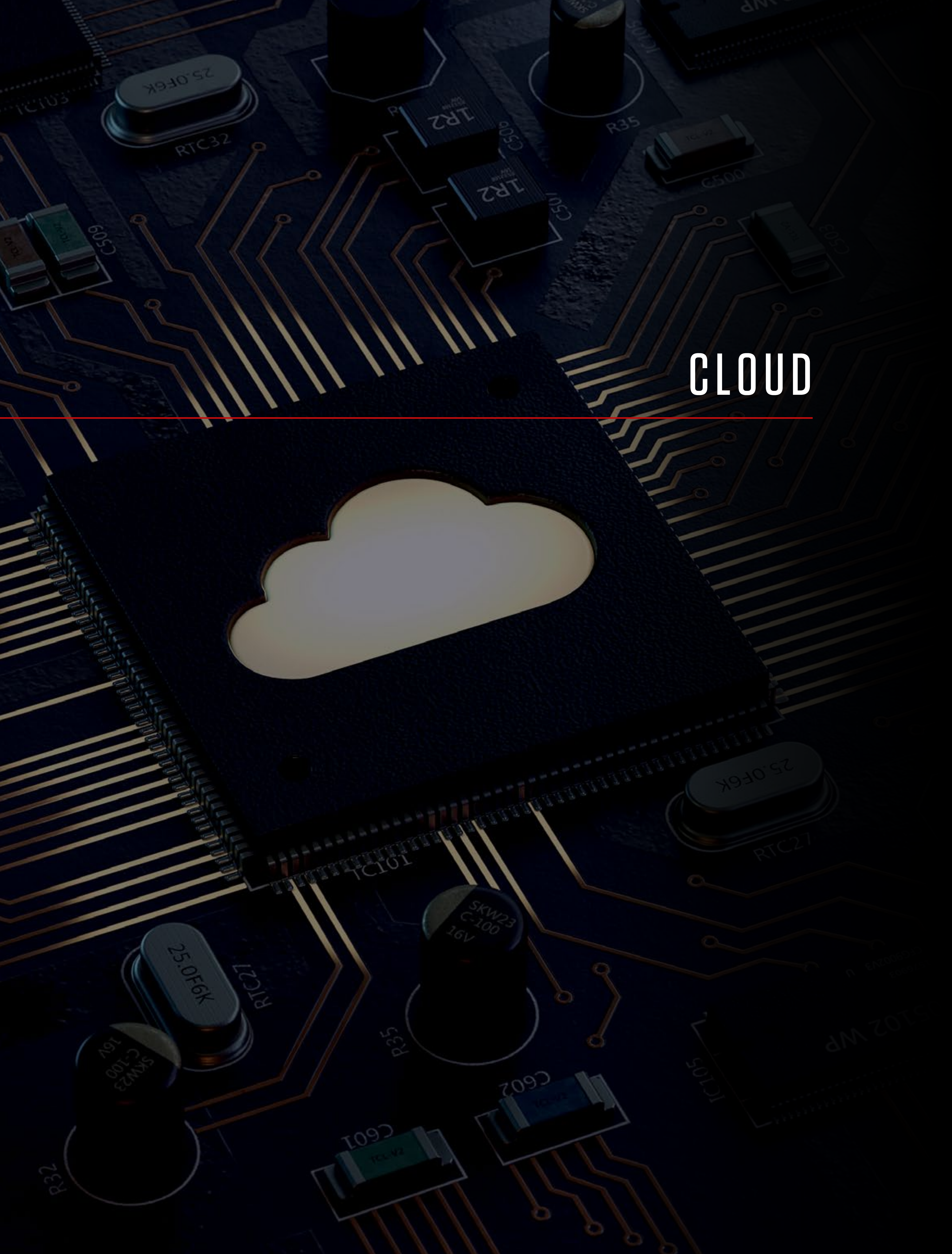
In addition, this edition of the Journal examines important topics around digital assets and decentralized finance, including central bank digital currencies, and bitcoin's impact on the environment, and cybersecurity and resilience.

As ever, you can expect the highest calibre of research and practical guidance from our distinguished contributors, and I trust that this will prove useful in informing your own thinking and decision-making.

Thank you to all our contributors and thank you for reading. I look forward to sharing future editions of the Journal with you.

Lance Levy, **Capco CEO**

CLOUD


25.0F6K
IC103
RTC32
1R2
1R2
C506
C507
R35
C500
C509
C508
IC101
25.0F6K
RTC27
SKW23
C-100
16V
25.0F6K
RTC27
R35
SKW23
C-100
16V
R32
C601
C602
IC105
102 WP

CLOUD'S TRANSFORMATION OF FINANCIAL SERVICES: HOW COVID-19 CREATED OPPORTUNITIES FOR GROWTH ACROSS THE INDUSTRY

PETER KENNEDY | Partner (U.K.), Capco
ANIELLO BOVE | Partner (Switzerland), Capco
VIKAS JAIN | Managing Principal (U.S.), Capco
CHESTER MATLOSZ | Managing Principal (U.S.), Capco
AJAYKUMAR UPADHYAY | Managing Principal (U.S.), Capco
FRANK WITTE | Managing Principal (Germany), Capco

ABSTRACT

The financial services sector is undergoing unprecedented disruption, thanks to a combination of the digital revolution and COVID-19's social and business upheavals. The collision of these two forces in 2020 and 2021 quickly altered the competitive landscape. Financial services institutions had to become smarter and more nimble, working in new ways with unfamiliar technologies at an unparalleled pace to meet escalating digital demands of clients. Firms also had to deal with increased competition, as fintechs, as well as technology giants, looked to take advantage of uncertain macro and micro economic environments. The phrase "inflection point" is often misapplied to characterize various competitive shifts, but we believe it accurately describes what leaders in financial services face today. In this paper, we lay out the industry's current state as seen through the eyes of practitioners, how cloud technologies are being used as an accelerant to drive growth and return on investment (ROI), and what lies ahead for our clients over the next few years.

1. INTRODUCTION

The advent of COVID-19 forced most industries to speed up their plans to more effectively service their clients or end customers – and in many cases, this meant accelerating their digitization strategies. The financial services industry has been no different, and during the pandemic we witnessed an industry that responded rapidly to new demands and has emerged stronger and more nimble.

One of the key technologies that assisted the financial services industry to address clients' shifting expectations during the pandemic has been cloud services, neatly summed up by Microsoft as "the delivery of computing services – including servers, storage, databases, networking, software, analytics and intelligence – over the Internet ('the cloud') to offer faster innovation, flexible resources and economies of scale."

Drawing upon data from Wipro FullStride Cloud Services' global report – entitled "Making Business Thrive: A Cloud Leader Roadmap for Achieving 10x ROI" – we looked to identify key trends and opportunities that financial services firms should be considering as they seek to become cloud leaders. We found that the senior executives who responded to the survey believed that: (1) the health crisis accelerated technology roadmaps, prompting more than half of respondents globally to invest more in cloud to increase resilience, reduce risk, enable more flexible working alternatives, and create new ways to interact with customers; (2) financial services firms have started their journey toward full digital implementation in the cloud, but are likely three to five years away from achieving that vision; and (3) cloud investments pay off on the bottom line in operational cost reduction but, more importantly, by generating new revenue (ROI generated from the cloud multiplies tenfold as firms advance on their cloud journey). Other benefits include making organizations quicker to market, enhancing their customer relations, and contributing to smarter decisions; and within two years, 40 percent of firms are expected to be advanced or fully optimized in respect of cloud data centers, migration and modernization of core processes, and cloud-native app development.

2. THE STUDY

The Wipro FullStride Cloud Services study was compiled using responses from 1,300 C-level executives and key decision-makers across 11 industries, of which 26 percent were financial services related – specifically, banking, insurance, and capital markets, such as wealth advisory and asset management firms. This article uses data compiled from close to 340 financial services executives to get a better understanding of their perspectives regarding how cloud computing impacted their businesses during the pandemic.

One of the key questions those executives were asked was how COVID-19 affected their operations. Four key implications of COVID-19 were cited by financial service leaders (Table 1):

1. Elevated the priority to use cloud to improve customer experience.
2. Realized the importance of cloud usage to make processes more efficient and agile.
3. Increased willingness to make cloud investments.
4. Elevated the priority of cloud to improve business continuity and resilience.

A larger picture emerged as we combed through the data survey. As companies emerge from the pandemic, management teams are embedding the cloud into their growth platform for the future. For digital leaders, the cloud provides a data-enabled, interconnected foundation to support enterprise-wide business activities and workflow solutions, while leveraging use of artificial intelligence (AI), internet of things (IoT), and other transformative technologies.

3. STATE OF PLAY: WHAT FIRMS ARE DOING NOW

Executives were asked to look back two years and ahead two years and reflect on what have been the most significant areas of investment, both technologically and organizationally. Reviewing the previous two years, here is what they said:

Table 1: How has COVID impacted your organization?

	BANKS	CAPITAL MARKETS	INSURANCE
Accelerated our timetable for moving more activities to the cloud.	34%	27%	25%
Caused us to increase investment in the cloud.	49%	52%	45%
Caused rethinking our cloud organization and skills training plans.	4%	6%	7%
Has elevated the priority of cloud usage to improve business continuity and resilience.	46%	53%	50%
Elevated the priority of cloud usage to improve the customer experience.	70%	63%	55%
Elevated the priority of cloud usage to support remote working and collaboration.	43%	40%	32%
Increased cybersecurity and compliance issues from use of the cloud.	43%	42%	42%
Opened new opportunities for cloud-enabled products and services.	24%	35%	25%
Shown the importance of cloud usage to make processes more efficient and agile.	54%	67%	59%

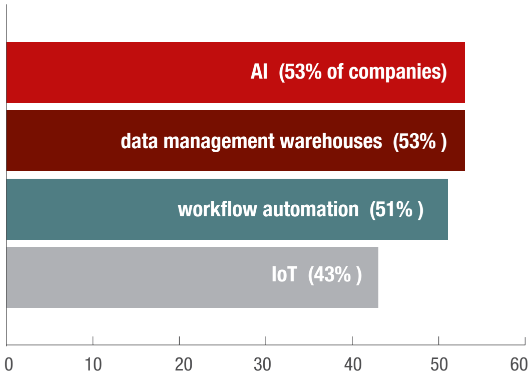
Table 2: Where have you made significant progress using cloud?

	BANKS	CAPITAL MARKETS	INSURANCE
Customer management and experience	43%	38%	26%
Cybersecurity and risk management	37%	35%	35%
Data and customer analysis	59%	58%	53%
Financial management, reporting, and auditing	40%	37%	34%
Human resources and employee experiences	23%	14%	19%
IT management and operations	49%	53%	38%
Marketing and distribution	16%	14%	28%
Middle- and back-office processes	19%	9%	14%
Procurement and supply chain	10%	8%	16%
Product development, R&D, and innovation	16%	12%	23%
Sales and business development	20%	18%	23%
Strategic planning and market analysis	23%	39%	29%

3.1 Spending

Banks' cloud spending was expected to have averaged U.S.$36 million in 2021; capital markets, U.S.$41 million; and insurance, U.S.$55 million. Average cloud spend across all 11 industries was estimated at U.S.$37 million. Despite those investments, financial services organizations realize that they are nowhere near fully optimized around the cloud. Fewer than 20 percent consider themselves either in advanced implementation stages or fully optimized, and less than 10 percent expect to be fully optimized in two years. Currently, firms operate, on average, 38 percent of their business applications through the cloud, and they anticipate that percentage will increase to 55 percent in two years. (Table 2).

Figure 1: Significant technologies employed



3.2 Projects

Similar to investments made by other industries, cloud spending in the financial services sector is spread evenly between data center projects, cloud-native development, migration of processes, and modernizing those processes. Spending is also balanced between upfront investment, implementation costs, and ongoing maintenance and fees.

Almost half (48 percent) of financial services firms report using digital enterprise platforms for enterprise resource planning, a slightly higher rate than other industries. Centralizing operations on one platform can be a critical benefit. As one banking executive said, "With the help of perfect cloud integration, the company's 253 banks, six branches, and the central bank will now be operated on one platform." Cloud has given them the most advances in the areas of data and customer analysis; IT management and operations; customer management and experience; financial management, reporting, and auditing; and cybersecurity and risk management.

3.3 Strategies

In the early years of cloud deployments, companies across industries brought in digital technologies to solve specific problems. The survey data reflects that organizations are now maturing in using cloud-based digital solutions and adopting a cloud-across-the-enterprise strategy.

For example, in the marketing department, financial services firms can now use powerful tools, such as AI and data analytics, to differentiate their brand and drive enduring loyalty and increased customer satisfaction. In insurance, 74 percent of respondents listed "customer and management experience" as the top business area where they have used cloud resources most effectively (banks reported 53 percent, and capital markets 46 percent in that area).

In addition to using cloud to drive business strategy, organizations must decide on a strategy for managing their cloud resources – storage, security, applications, maintenance, and the like. According to the survey, most firms expect growth in the use of both hybrid cloud and public cloud, suggesting hybrid is not just a holding pattern while customers get comfortable with public cloud. Hybrid is here to stay.

"By deploying hybrid cloud models, we have seen that it allows us to integrate risk data within its environment, which helps us against data breaches and thefts," an insurance executive reported.

Who are firms turning to provide those services? Financial services firms envision relative growth in Google Cloud Platform and Microsoft Azure to be on a par with Amazon Web Services in the next two years. (In other industries, Google looks to have the pole position.)

Respondents also predict growth in the use of platform as a service (PaaS) and software as a service (SaaS) tools, which allow them to rapidly scale up projects as needed without adding significant capital expense. "The implementation of cloud solutions, like SaaS and PaaS," an insurance executive wrote, "helps in managing IT resources (including) upgrading storage, memory, and scaling processing speed."

4. FUTURE STATE: WHERE FIRMS SEE THEMSELVES IN TWO YEARS

While IT spending, in general, is expected to remain static, cloud spending is likely to increase. Firms are broadly expecting cloud costs to rise in the 1-10 percent range.

The cloud is fast becoming more intelligent, hyperconnected, and pervasive. For the immediate future, financial services firms are focusing their cloud goals more on business growth and revenue generation than reducing expenses.

Leaders said they expect to make their most significant cloud investments over the next two years in product development/R&D (62 percent), cybersecurity (48 percent), business development and sales (42 percent), and procurement/supply chain (44 percent).

Although the survey does not provide specific project details, work with our clients offers more information about where investments might be made.

4.1 Business development and sales

Financial services firms are often rich in client data, the seed corn for powerful targeted marketing campaigns. Cloud platforms provide relatively cheap processing power, storage, and data analytics tools that enable firms to segment customers to a finer degree than previously possible and create customized offerings based on customer behavior. "With the help of Cloudera," a bank executive offered, "we created a predictive analytics platform that delivers targeted recommendations for trading ideas, research material, investment strategies, and much more."

4.2 Cybersecurity

Cyber crime is soaring as digital criminals become more sophisticated. The FBI's Internet Crime Complaint Center reported 2,084 ransomware complaints from January through July 2021,[1] a 62 percent jump over the previous year. This means that firms must continuously update their defense programs and reduce threat surfaces, such as by employing DevOps practices to build in security as code is written. Moving to cloud helps here because cloud vendors provide proven security measures for clients and continually update technologies, processes, and recovery mechanisms. In addition, global regulators are showing more interest in protecting consumer safety and privacy online, requiring firms to focus on these areas or face substantial penalties.

4.3 Procurement/supply chain

Financial services firms are developing multiservice cloud strategy approaches that mix and match tools and services from multiple vendors. For example, cloud enables banks to better use a myriad of services available from fintechs and other vendors. In capital markets, some vendor services can be hugely resource intensive, especially those involving pricing and Monte Carlo risk simulations using large market

[1] https://bit.ly/3hr1mSt.

datasets. A cloud platform allows firms to test and install these services much more quickly and effectively because they no longer depend on a lengthy hardware procurement process. From a procurement perspective, this allows a degree of experimentation and switching around of services, which is not otherwise possible.

4.4 Product development/R&D

In the cloud, firms can take advantage of specialized computing resources to model potential products, calculate returns from new business lines, and spur innovation by using agile development practices to "fail fast" by getting products to market quickly for testing. Nearly half of all firms will increase their investments in each of these four business functions and aim to get to 40-50 percent of advanced implementation status in two years. Significant dividends are expected from their cloud investments.

In two years, 62 percent of the financial executives who responded to the survey expect that the cloud will increase their revenues, 55 percent expect cloud to increase market share, 46 percent expect it to improve customer retention, and 30 percent expect the cloud to boost shareholder value (Table 3). Over half of the firms plan to use cloud capabilities to leverage artificial intelligence (AI), data analytics, workflow automation, and digital enterprise platforms.

The survey suggests that cloud especially matters when financial services organizations need ready access to data and processing power for analysis, such as in financial management and reporting, employee and customer experiences, and IT management and operations. Banking and capital markets firms plan to make their most considerable advances in payment processing and services and retail branch management. Insurance firms are prioritizing underwriting and product/policy design. As noted previously, financial services organizations are at least three years away from utilizing digital services to the fullest. We imagine these future products could include using AI and machine learning to provide customers with proactive services that fit their in-the-moment lifestyle needs, on-demand insurance and rapid claims adjusting, or cryptocurrency management using blockchain.

Table 3: Expected cloud benefits in two years

		BANKS	CAPITAL MARKETS	INSURANCE
FINANCIAL	Better use of capital	7%	10%	12%
	Decreased costs	48%	56%	41%
	Greater shareholder value	31%	30%	29%
	Improved profitability	47%	59%	50%
	Increased revenue	66%	66%	55%
OPERATIONAL	Accelerated time to market	24%	28%	28%
	Greater and faster innovation	25%	23%	26%
	Greater teamwork and stronger corporate culture	35%	29%	35%
	Improved employee engagement/productivity	42%	40%	28%
	Increased customer satisfaction and retention	47%	48%	42%
	More effective risk management and compliance	35%	44%	34%
	Reduced carbon footprint	14%	24%	16%
	Streamlined operations/improved quality	25%	30%	27%
STRATEGIC	Greater ability to scale across business/global markets	32%	29%	22%
	Greater innovation/new business models	29%	23%	28%
	Greater market share/expanded client base	62%	56%	46%
	Improved planning and decision-making	47%	49%	39%
	No benefits	0%	0%	0%
	Stronger reputation and brand equity	14%	19%	15%
	Stronger reputation and brand equity	12%	7%	11%

5. BENEFITS: FINANCIAL AND OTHER DIVIDENDS

In a crucial shift, the cloud is moving from an efficiency play to a growth driver. The survey results give financial services leaders real numbers around the benefits they can expect from a well-executed move to digital.

Almost all clients report revenue benefits over three years due to using cloud to create new products, services, and business models. These gains average at around 4 percent, although about a third of firms anticipate revenue increases of up to approximately 15 percent (Table 4).

As one banking executive commented: "Customer relationship management using a PaaS solution via cloud technology is our organization's most successful revenue-generating cloud initiative."

Other executives pointed specifically to automation as a cloud enabler that provides quick returns. They touted wins with automation in policy governance (insurance), claim settlement and operational efficiency (insurance), and market settlements and clearing processes (capital markets) as effective ways to generate more returns.

Financial services organization tend to see cloud investment paybacks over 24 months or less from a data center perspective and for migration of legacy systems. For modernization and cloud-native development, results tend to be more variable and spread over a more extended period.

The biggest eye-popper: return on investment (RoI) generated from the cloud multiplies tenfold as firms advance on their cloud journey. While beginners see a 6 percent annualized cloud-related RoI, this grows to 44 percent for advancers, and 59 percent for leaders. "Cloud adoption," an insurance leader said, "gives [us] the ability to convert fixed infrastructure costs into variable costs."

The survey shows that when it comes to the cloud, practice makes profits. Cloud usage drives significant cumulative bottom-line gains. While beginners see a lift of about 2.6 percent in total revenue gains and cost reductions when starting cloud computing, cloud leaders get a boost of up to 12 percent.

Here is another financial benefit: hyperscalers – the leading cloud platform providers – are helping fund this digital future. They pick up all or some of the upfront and implementation cloud costs for about three-quarters of firms, covering 78 percent of related software costs on average.

We believe companies often do not consider the total cost benefits when measuring the RoI on the cloud. Only 40 percent include benefits from decreased non-IT costs, and even fewer measure reduced carbon footprint, accelerated time to market, or improved productivity.

Table 4: Cloud promotes revenue increases three years after implementation

	BANKS	CAPITAL MARKETS	INSURANCE
No change in revenue	7%	7%	2%
1%	13%	20%	16%
2%	13%	15%	27%
3%	17%	17%	20%
4%	18%	12%	4%
5%	12%	7%	9%
6% to 10%	15%	15%	16%
11% to 15%	3%	7%	7%
16% to 20%	2%	2%	–
Average (of firms that reported a revenue impact)	**4.03%**	**4.10%**	**4.00%**

Asked to identify the main benefits from using cloud beyond revenue and profitability gains, the top five were (Table 5):

1. Increased market share/expanded client base
2. Improved customer satisfaction and retention
3. Greater teamwork and stronger corporate culture
4. Improved planning and decision-making
5. More effective risk management and compliance.

Table 5: Main benefits of using cloud

		BANKS	CAPITAL MARKETS	INSURANCE
FINANCIAL	Better use of capital	5%	9%	7%
	Decreased costs	50%	51%	40%
	Greater shareholder value	18%	16%	17%
	Improved profitability	62%	55%	59%
	Increased revenue	55%	50%	46%
OPERATIONAL	Accelerated time to market	28%	25%	21%
	Greater and faster innovation	24%	18%	27%
	Greater teamwork and stronger corporate culture	42%	30%	39%
	Improved employee engagement/productivity	41%	34%	37%
	Increased customer satisfaction and retention	48%	48%	37%
	More effective risk management and compliance	41%	44%	35%
	Reduced carbon footprint	5%	3%	5%
	Streamlined operations/improved quality	26%	24%	22%
STRATEGIC	Greater ability to scale across business/global markets	37%	27%	25%
	Greater innovation/new business models	34%	28%	40%
	Greater market share/expanded client base	55%	55%	38%
	Improved planning and decision-making	45%	40%	45%
	Stronger reputation and brand equity	12%	7%	11%

6. OBSTACLES: NAVIGATING THE NEGATIVES

The survey provides insights into where companies are most likely to fly into headwinds on their digital journeys.

Lack of a cloud strategy is the biggest challenge facing companies across the board, financial executives say. Some 44 percent cite this as a common obstacle hindering their transition to the cloud. This remains a hurdle throughout the cloud journey for beginners (45 percent) and leaders (36 percent), as companies strive to take their strategy to a higher level.

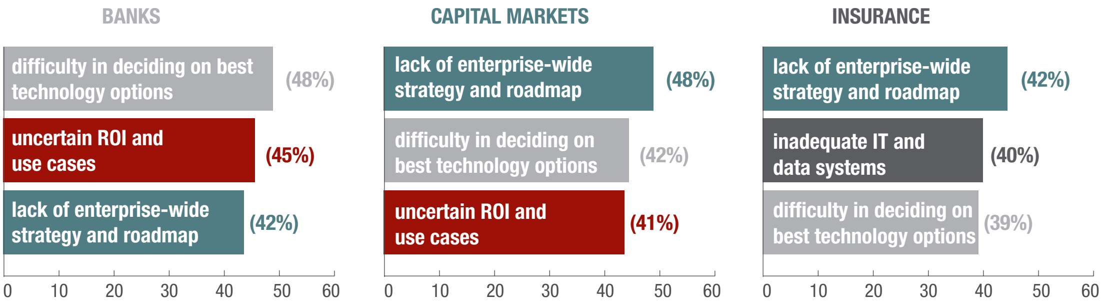
Why does this happen? Our clients' experiences show that cloud technologies have traditionally been applied at the department level to fix specific issues, such as improving data entry speed and accuracy or automating back-office workflows. Internal IT groups favored keeping these digital solutions on-premises or in a hybrid mix. Thus, cloud initiatives became buried in departmental silos.

The result: roadmaps that chartered a digital journey for the entire organization were slow to happen, if at all. In our recommendations below, we stress the importance of developing an enterprise-wide cloud strategy and roadmap early in the transformation process, one that details technology choices, governance measures, spending priorities, and that moderates other potential battlegrounds that can dilute implementation.

The survey also underscores that training, recruitment, and retention are fundamental competencies that need careful and early planning or risk snagging a cloud shift. An average of 25 percent of financial services firms said "limited access to cloud skills and talent and need for training" were serious impediments to successful cloud implementation.

Cloud leaders are already increasing their digital capabilities. 52 percent of cloud leaders develop teams and skills to drive cybersecurity in the cloud, 48 percent provide cloud training to IT and line of business staff, and 43 percent roll out change management strategies to facilitate cloud transformation.

Figure 2: Top three obstacles to cloud implementation listed by financial sector



Among the list of roadblocks emerged a couple of surprising and somewhat cheerful findings.

Once a significant sore spot for our clients, managing multiple technology providers was mentioned as a substantial obstacle by just 14 percent of respondents. In addition, surprisingly, firms said shifting from a capex model to an opex one, which requires some complexity in how costs and chargebacks are handled, was not a major difficulty.

7. FINANCIAL SERVICES DIFFER SIGNIFICANTLY FROM OTHER INDUSTRIES

The survey, which spans 11 industries, suggests that financial services in general still lag other industries in their use of digital, but shows investments are underway to catch up and even lead in some sectors.

On a global level, banks and capital markets firms lag those in other industries in their cloud maturity. The financial services sector is much further ahead in the U.S., where 42 percent of banks and 30 percent of capital markets companies are cloud leaders (33 percent combined). This is not surprising, given the country's overall edge in technology adoption and cloud usage. This follows the pattern seen in other industries, where companies headquartered in the U.S. also tend to be more advanced. Moreover, financial services firms in the U.S. are making progress quickly, with 51 percent in the advancer stage.

Even if financial services organizations aren't generating the RoI experienced by more advanced users in other industries (average cloud RoI for capital markets firms is 10 percent versus 66 percent for oil and gas companies, for example) the survey shows an industry knee-deep in its transition to digital.

Table 6: How different industries use cloud to generate revenue

	BANKS	CAPITAL MARKETS	INSURANCE	CONSUMER PACKAGED GOODS	HEALTHCARE PROVIDERS	LIFE SCIENCES	MANU-FACTURING	OIL/GAS	RETAIL	TRANS-PORTATION	UTILITIES	TOTAL
Create new products and services	8%	7%	14%	6%	21%	19%	29%	18%	19%	16%	16%	16%
Develop new business models	25%	13%	25%	35%	21%	28%	22%	16%	24%	23%	22%	23%
Grow revenue through greater productivity	65%	60%	46%	62%	62%	56%	46%	54%	51%	56%	68%	57%
Improved market positioning and branding	27%	30%	30%	29%	36%	43%	23%	29%	37%	24%	18%	30%
Increase customer retention and upselling	38%	38%	38%	46%	32%	30%	35%	27%	44%	40%	30%	36%
Penetrate new client segments	55%	60%	61%	38%	38%	33%	40%	54%	53%	42%	38%	47%
Reach new global markets	67%	60%	41%	44%	53%	59%	54%	64%	53%	53%	60%	55%
Speed up time to market	63%	63%	54%	62%	38%	59%	75%	52%	59%	65%	60%	59%

Practically, no firm views itself as either in advanced implementation or fully optimized at this point. Financial service organizations are near the cross-industry average in this regard, with banking and capital markets slightly higher and insurance lower. Life sciences, manufacturing, and retail industries are significantly ahead.

There is a stronger focus than in other industries on improving productivity and reaching new global markets, more so with banking and capital markets than insurance. Banks see the cloud, more than other industries, as a way to become lean and standardized (Table 6).

Within financial services, an interesting tale is told that separates, in a significant way, the insurance sector's cloud journey from banking and capital markets firms. Insurance organizations target 18 percent RoI on cloud initiatives versus 14 percent for banking and capital markets. The insurance RoI number was also slightly higher than all other industries. This statistic is striking; generally, the use of cloud in insurance has different priorities and economics than the rest of financial services and is much more in common with other industries. This begs the question: is insurance that much better at designing and implementing cloud projects, or is it capturing returns on projects today that other industries earned long ago while insurance was digitally disengaged?

8. DIGITAL TRANSFORMATION IS DRIVEN FROM THE TOP

Given the vast strategy shifts, financial resources, and organizational changes required for a successful digital transformation, it is perhaps not surprising that survey-takers understand the importance of senior leadership to drive digital journeys. (Remember, a quarter of firms listed lack of senior executive support as a major obstacle.)

When asked who is responsible for overall cloud strategy, the most common response across industries was the chief operating officer (COO), closely followed by the CTO, although there was a wide spread of answers. The CFO, CDO (Chief Digital Officer), CSO, and CIO figured often. However, around a quarter of banking and capital market organizations see this as the CEO's responsibility.

When asked which executives play critical roles in developing and implementing that strategy, 69 percent listed COOs and 63 percent listed CEOs.

9. RECOMMENDATIONS: DESIGNING THE SUCCESSFUL CLOUD JOURNEY

We will not blame financial services leaders if they feel pinned in the middle of a riptide roiling the entire industry. Steps they took over the last two years to move to cloud, as traumatic (and transformative) to their businesses as they have been, could be washed away if their transition stalls in the next two years.

This is not the time for timidity. Just keeping pace with competitors will not cut it. Here are key areas to consider for expediting your digital journey to the cloud.

- **Organizational commitment:** buy-in must be broad and deep across all lines of business. The mandate for progress must come from the top but be bought into by every employee: (a) instead of organizing around products or services, today's financial services organizations must put the customer at the center of it all (leaders drive a simple yet powerful message to all employees: if you take care of the customer, the rest of the business will take care of itself); (b) be transparent that although robots will take some jobs, displaced workers will be trained for new positions that will be more rewarding and fulfilling; and (c) recognize and support the CTO, who plays a critical role in executing the transformation strategy and evangelizing benefits to the entire organization.
- **Prioritize:** stage implementation steps in a way that boosts confidence: (a) build momentum early (as a capital markets executive expressed in the survey: "Engage the business through some early flagship big wins"); (b) identify key areas to convert to cloud-ready across the existing tech landscape, platform, application, tools, and processes (emphasize operating model changes that can create competitive wins over competitors, such as fast claims resolution or wider accessibility to loan products); and (c) convert manual, time-consuming, and sequential processes to simpler and nimbler tasks that automation can effortlessly handle.
- **Plan:** create an aggressive but balanced roadmap for cloud adoption, focusing on agile development methodologies and skill-building: (a) ensure first and foremost that digital initiatives align with business goals; (b) deploy evaluation mechanisms at regular periods to verify the program is on track, that technical debt is minimized, and better user experiences are emphasized;

and (c) build in robust programs to train employees on the new technologies and business practices. In addition, involve key workers early in the design phase – they will confirm whether the new processes will do the job required. Training should be completed before the conversion switch is thrown.

- **Upskilling and recruiting:** financial services organizations will quickly need to crack the human resource equation of upskilling existing and hiring talent to manage the digital operation and cloud-native development going forward. Finding talented workers in the digital sphere – programmers, designers, UX, data analysts, security experts – represents a major potential stumbling block to your digital aspirations. In the cloud, capabilities need to reach a new level: (a) survey results emphasize that many companies already invest in training IT and line of business staff, and recruit deployment specialists to lead cloud deployment; (b) use automation to free up and upskill workers to take on more valuable roles, such as customer service or auditing; and (c) develop deep relationships with universities and training schools. Regular campus visits to recruit talented new graduates will be key for keeping your personnel pipeline full.
- **Customer experience:** keep customer experience as the core theme of cloud strategy and adoption: (a) creating compelling customer experiences pays off in lower churn and greater loyalty to the brand; (b) customer experience (CX) should engage at every customer touchpoint with your organization, from opening an account to navigating your mobile website; and (c) in the survey, financial executives said customer management and experience will be one of their top focus areas over the next two years.
- **Partner:** invest in assessing and leveraging third-party tools and capabilities to expedite the cloud journey and enable repeatability and scale: (a) the survey reveals that outside contractors make up about a quarter of the workforce. Finding, managing, and collaborating with partners will be a crucial capability to develop; (b) most financial services organizations do not have the in-house resources necessary to provide all the products and services customers want. Partner with adjacent companies to offer clients one-stop financial shopping, such as integrating financial planning, insurance, and financing; and (c) partners should be viewed as strategic collaborators, not just parts providers. Consult them as you develop products and services, build capabilities, and evaluate technologies. Chances are they have been there before and have wisdom to share. A banking executive shared that they signed a five-year strategic partnership with their technology partners, a longer-term commitment intended to foster stability and collaboration in an uncertain environment.

10. CONCLUSION

The next few years will be critical for determining leadership in many industries, including financial services.

COVID will eventually fade, but the remainder of this decade will only see disruption escalate. That is because the pandemic unleashed new consumer expectations. They want 24/7 access to their financial resources from anywhere, and products hyper-personalized to their lifestyle choices. For financial services organizations to meet those requirements and grow, they will need to be driven by data, faster to market, and agile and resilient in execution.

Some industry giants will fade. Some unknown companies will ascend. Products not even conceived before today will win the hearts and wallets of financial consumers. The companies best designed to compete during this turmoil will be crowned new industry leaders.

And those companies will be thriving in the cloud.

CLOUD FINANCE: A REVIEW AND SYNTHESIS OF CLOUD COMPUTING AND CLOUD SECURITY IN FINANCIAL SERVICES

MICHAEL B. IMERMAN | Associate Professor of Finance, Peter F. Drucker and Masatoshi Ito Graduate School of Management, Claremont Graduate University; Visiting Scholar, Federal Reserve Bank of San Francisco

RYAN PATEL | Senior Fellow, Peter F. Drucker and Masatoshi Ito Graduate School of Management, Claremont Graduate University

YOON-DO KIM | Quantitative Analyst, Federal Reserve Bank of Minneapolis; Ph.D. Student in Financial Engineering, Claremont Graduate University

ABSTRACT

Cloud computing is hardly a new concept, although its embracement by the financial services industry has mostly occurred in the past few years. Unlike traditional computing infrastructure used by financial services firms, such as data centers and mainframes, cloud computing relies on the internet to access storage hardware as well as software applications from anywhere at any time. This is proving to be of tremendous value for many firms especially as remote work becomes more common and on-the-fly data access is expected by stakeholders. However, it is not without its risks and challenges. In this article, we review the current state of cloud computing as it applies to financial service firms and outline both the benefits and challenges, including cybersecurity issues for data and applications based in the cloud. Further complicating matters for incumbents in the financial services industry is the fact that fintech challengers are "cloud native", in that they are built upon a cloud-based computing infrastructure and are, therefore, able to more easily adapt to changes with the technology.

1. INTRODUCTION

Cloud computing, defined as the use of computing services that are accessed over the internet rather than via onsite hardware and software, went from being an emerging technology used by only the earliest adopters just over a decade ago to now being ubiquitous in almost every organization from higher education to healthcare as well as financial services. In this review article, we discuss the evolution of cloud computing paradigms with particular emphasis on their application to financial services and fintech.

We start with a very brief literature review. A quick Google Scholar search of the terms "cloud computing" and "financial services" returns over 17,000 hits just since 2018! That being said, rigorous studies that analyze the implementation of cloud platforms and implications for business strategies are few and far between. As previously noted, with cloud computing becoming ubiquitous in financial operations – from the legacy firms (or incumbents) to the fintech startups – more analysis, especially from a risk management perspective, is warranted.

As a review article, we then proceed to cover the state of cloud computing. Topics such as public, private, and hybrid cloud models are discussed in enough detail to familiarize the reader but without getting overly technical. We then proceed to discuss the importance of cloud computing technology to financial services and fintech. The following section goes on to address cybersecurity issues and their importance to cloud computing in financial services. Finally, we conclude with some remarks for investors, regulators, startups, and incumbents about how they may want to approach cloud computing in financial services and fintech going forward.

2. LITERATURE REVIEW

While there has been quite a bit of scholarly attention on cloud computing, until recently few studies have focused on its applications in financial services. Most of the previous research on the application of cloud computing note its benefits to financial services firms. One of the earlier papers that discusses how cloud computing can optimize financial services is Ghule et. al (2014), who specifically look at banking activities. One of the primary benefits that is highlighted is automation in many of the bank's processes. Going further, the authors list cost savings, business continuity, business agility, and environmental friendliness as other benefits of cloud computing applications. These benefits, other than environmental friendliness, are reiterated by Yan (2017).

However, it has been noted that the applications of cloud computing in financial services are not without challenges. Yan (2017) notes that information security issues can be one of the biggest risks, which are associated with data breaches and cloud destruction. Furthermore, utilizing cloud computing in banking can lead to more general business continuity issues. This is because cloud computing providers might lack capacities that the banks require, thereby forcing the bank to go to yet another external vendor that may not be compatible with the bank's existing systems. At the furthest extreme, if the cloud services provider declares bankruptcy and liquidates, this could have massive implications for the bank's business operations. Lastly, this article points out that the lack of technical standards on its regulatory rules and policies on the application of cloud computing represent both a risk and challenge. A more recent paper by Sampson and Chowdhury (2021) highlights data breaches as the biggest concern for financial institutions such as banks. For instance, in a high-profile well-publicized case, Capital One was victim of a data breach in summer 2019. This breach included data from over 100 million of its customers, including personal information such as names, addresses, phone numbers, birth dates, social security numbers, and bank account numbers.

In order to address these challenges, a few articles have suggested the need for standardization and regulation of cybersecurity in cloud computing, although further studies are certainly needed. A very well-done paper by Scott et. al. (2019) points to the existing regulatory frameworks for cloud computing applications in financial services – one designed by Federal Financial Institutions Examination Council (FFIEC) for the use in the U.S., and the other by European Banking Authority (EBA) in Europe. These frameworks require financial services firms and their regulators to perform a preliminary risk assessment on the cloud computing service providers as well as monitor and audit them. We agree with the authors' assessment that this is an area that is going to require more resources from regulatory agencies for ongoing monitoring and risk control when it comes to financial institutions' use of cloud computing.

The recent study by Tissir et. al. (2021) proposed that cybersecurity for cloud computing be standardized according to the frameworks offered by International Organization for Standardization (ISO) and National Institute of Standards and Technology (NIST). They note that the purpose of such standardization would be to achieve improved levels of security with stronger controls in place in a cost-effective and reliable cloud environment.

3. CLOUD COMPUTING: A REVIEW

Cloud computing is a technology that is being used for development and deployment of a variety of fintech solutions. The technology has evolved so dramatically over the past decade that anything written about cloud computing in 2012 would be out of date in describing applications today in 2022. In this section, we examine the current state of cloud computing and discuss its importance to applications in fintech and, more broadly, financial services. First, we will provide a definition of what cloud computing is and then we will make a distinction between private and public cloud.

3.1 Definition of cloud computing

Cloud computing generally refers to the model where computing services are accessed over the internet rather than from in-house, onsite hardware and software. The hardware to which we refer may include storage or processing. These used to be synonymous with cloud computing in years past, but now much of the value added comes from software, which could include database management systems (DBMS), business intelligence and analytics platforms, customer relationship management (CRM), enterprise resource planning (ERP), ML algorithms and AI tools (e.g., TensorFlow and sentiment analysis, respectively), and cybersecurity solutions.

The economic model and accounting processes for cloud computing are dramatically different from traditional IT management in financial services. With cloud computing, access to the hardware and/or software is based on a pay-as-you-go or pay-as-you-use model. Traditionally, when it comes to systems, financial institutions have relied on massive physical servers based out of their own data centers as well as legacy mainframe-based systems that are built on top of half-century-old technology. These require substantial upfront investments, which, from an accounting perspective, would be depreciated over time.

Cloud technology plays a vital role in the fintech space by providing a more flexible and agile business model that is more readily able to adapt to changing market demands. Sometimes, when dealing with cloud computing technology, you will hear about IaaS (infrastructure-as-a-service) and SaaS (software-as-a-service), respectively, as cloud computing providers market their hardware and software solutions. Increasingly, most cloud computing platforms incorporate both the hardware and/or software components, depending on the client's needs; consequently, a model that falls in between IaaS and SaaS is platform-as-a-service (PaaS).

Many fintech companies are "cloud native", meaning that they are built "in the cloud" and have been cloud-based from their inception. This is particularly important, as the inherent flexibility that cloud models provide is conducive to the agile framework that allows startups and challengers (in any industry but especially in the finance industry) to fail fast, pivot, and move in a new direction much faster than the incumbents. However, it does necessitate the reliance on public cloud providers, which can have a complicated cost structure and introduce potential risks. Consequently, before going further we will define what is considered "public cloud" versus "private cloud".

3.2 Different types of clouds

3.2.1 PUBLIC CLOUD

Public cloud refers to situations where the cloud computing technology is maintained by a third party. The public cloud market is dominated by the big three providers: Amazon (with Amazon Web Services or AWS), Google (with Google Cloud Platform or GCP), and Microsoft (with Azure). Another player in this space is IBM, a case we will come back to later in the section. In 2020, 6 percent of companies who had embraced cloud computing used a single public cloud [Flextra (2020)]. Reliance on the public cloud is a bit like using a utility. In that respect, from an accounting perspective, it is a part of your IT overhead but with a variable cost component, since you pay for what you use. Consider a fictitious company's hypothetical electricity bill. Management may know and expect that there will be a $1000 distribution fee per month regardless of usage. However, as they use more kilowatt hours (kWh) per month, the monthly charge will increase proportionately. If the utility charges 20 cents per kWh, then there could be an extra $600 for 3,000 kWh or $30,000 for 150,000 kWh, or anywhere in between. Using a public cloud provider is similar in that respect. You may pay a nominal periodic subscription fee, but the costs will increase proportionately with usage. The more apps that are used or the more storage that is required the higher the cost.

The firm can control costs – to some extent – by scaling up or down their cloud service needs. Hence, the economics and accounting associated with traditional financial services IT costs changes dramatically when moving from in-house computing to cloud computing. Rather than a large initial upfront cost that is then depreciated over time, there is this pay-as-you-go or pay-as-you-use model, which not only introduces flexibility in terms of how and what the software and/or infrastructure is used, but also introduces flexibility in terms of investment. This can be crucial for a startup with limited funds. But it can also lend a paradigm shift to the incumbents and their cost structures, if they make the leap.

3.2.2 PRIVATE CLOUD

Private cloud can be classified into several different categories, including virtual private cloud and on-premises private cloud. Virtual private cloud refers to situations where the cloud computing technology is maintained by a third party, but only for a single entity or a single organization. It provides higher security by constructing a firewall and only grants access to in-network users through virtual private network (VPN). The downside of a virtual private cloud is that the cost of services is significantly higher than that of public cloud.

On-premises private cloud requires an organization to completely build their own cloud infrastructure. It is an in-house private cloud that offers greater flexibility as well as higher security. However, there are many downsides to this type of private cloud. Firstly, it requires the users to be physically in the network, which limits accessibility. This was particularly a problem during the pandemic when remote work (work-from-home) became a mainstay in many industries. Secondly, it requires cloud professionals and consistent maintenance for higher security. Lastly, on-premises private cloud requires an enormous amount of equipment, which includes data centers. The data centers that make up on-premises private clouds are much bigger than they were 20 years ago, which makes sense given that we are in the age of "big data". When there were just a handful of servers in the data center, each one would have a name and would be referred to as a "pet". The IT folks knew their pets well. In today's data centers, which make up the so-called private cloud, they are no longer pets, but rather "cattle"; nameless, rather assigned a number, each server is a replaceable member of the herd.[1]

3.2.3 MULTI-CLOUD AND HYBRID CLOUD

To wrap up our discussion on the types of clouds, we address one issue that perhaps requires more attention from practitioners who are dealing with cloud computing in a financial services firm: relying entirely on one public cloud provider can be risky. This is a concept that financial services firms, especially investment companies, know very well and it has to do with diversification. Committing to one cloud provider opens the company up to both cyber risk and financial risk. Suppose that the cloud provider is the victim of a data breach or hack. Relying entirely on that one provider could result in being fully compromised. From a financial standpoint, if that one cloud provider fails for whatever reason (think Lehman Brothers in 2008), you are back to square one shopping for a new cloud service provider but with one less competitor (ergo giving them more pricing power).

To address the issue of risk, most companies these days – 93 percent in fact – are using a multi-cloud strategy [Flexera (2020)]. This could be splitting business across the big three, mentioned above, using other specialty cloud services, or a "hybrid cloud" model, which is becoming increasingly popular.[2] Hybrid cloud refers to a combination of using a public cloud provider and still using some private cloud, which offers protection for classified data from public cloud security breaches. Apart from the de-risking and diversification elements, there is the fact that some cloud service providers may be better for certain tasks than others and that is part of the decision that has to be made in constructing the multi-cloud strategy.

3.3 Cloud architecture and deployment models

In terms of cloud architecture, there are two trends that we feel are relevant to the reader. The first is serverless computing, or function-as-a-service (FaaS), and the other is the movement from virtual machine (VM)-based cloud platforms to distributed cloud computing architechture. Serverless computing does not require any infrastructure management, is highly scalable, and makes the most efficient use of resources. With serverless computing, the servers are still running the code, but the developer has no direct interaction with it, which allows their teams to focus on innovation and creating more value for the organization.[3] In addition to the big three, to which we have repeatedly referred – AWS, GCP, and Microsoft Azure – two other companies to consider in this FaaS space are IBM and Oracle.

When discussing modern cloud architecture and deployment, in any industry, a company that often comes up is Kubernetes. Kubernetes uses "containers", which effectively breaks up and distributes software across multiple systems simultaneously (i.e., in parallel). This is different to the previous deployment model, which used a single virtual machine (VM) to run all software on the cloud. Containers are modularized units on which apps can be developed and deployed. This allows for more efficient utilization of resources, which is particularly important when the app or cloud-based program uses a massive amount of data (think Netflix, which uses its own container deployment solution called "Titus").[4]

The idea of containers is not new and easily goes back decades to the advent of UNIX, Linux, and Solaris when server-based computing rose to prominence for larger organizations. The modern commercialization of containers can be attributed to Docker, which was released in 2013. This made for a new deployment model that could be used by organizations large and small, including startups whose entire value proposition is predicated on cloud-based app development. And, in fact, this brings us full circle to why cloud computing is so important to the growth and success of fintech.

3.4 Why is cloud computing so important for fintech?

Finally, we come back to the question: why is cloud computing so important for fintech? Well, we have already mentioned that fintech companies are "cloud native" and that the incumbents are scrambling to "migrate to the cloud" to remain competitive, but that does not really answer the question.

In order to answer the question (in part) we need to define APIs or "application programming interfaces". APIs have become the lifeblood of fintech apps. APIs allow data from multiple sources to come together on one platform, seamlessly, and

[1] This is, in fact, the analogy that practitioners in the area of cloud computing use [Menchaca (2018)].

[2] In fact, Flexera (2020) indicates that of the 93 percent of companies that are using a multi-cloud strategy, 87 percent are using a hybrid cloud strategy.

[3] Microsoft, "Serverless computing: an introduction to serverless technologies," https://bit.ly/3pYlqjS.

[4] https://bit.ly/3i6QBVW.

analyzed as if the data resided in the app itself. Plaid, the banking infrastructure company, was valued at U.S.$5 billion (USD) in the M&A transaction with Visa in February 2020.[5] Perhaps it is a bit of an overstatement to say that all of that value comes from its clever use of APIs in connecting banks and other financial service providers to various different technologies and software, but that is a large part of it.

In the fintech industry, as with many other industries that rely on digital solutions, data is a valuable commodity, and one way in which this commodity can be monetized is through APIs that provide access to data from different sources. Another important reason is that much of the growth of fintech has been due to the increasing reliance on mobile technology. Apps that run on mobile devices, such as tablets and smartphones, could not be developed or deployed without going through the cloud. To be a bit more technical, the containers that were previously discussed are conducive to microservices, which have become a standard component in developing apps such as the ones that fintech companies create and market.

4. CYBERSECURITY ISSUES

One area where we have seen, and will continue to see, attention being paid to in financial services is cybersecurity (as well as business strategies that prioritize good data management and cybersecurity). As with any kind of innovation, investors associate enhanced cybersecurity with a premium. On the flip side, firms that have lapses in cybersecurity will be penalized by the market. This is similar to the trend with corporate social responsibility (CSR).[6] It took time for companies to realize that this is something that customers demand and investors want.

We have seen increasing innovation and investment in cybersecurity in recent years, and perhaps nowhere is this topic more relevant than in discussing cloud computing and its application to financial services. As the financial services industry becomes increasingly more digitized, it is almost a veritable certainty that the industry will encounter more cyberattacks. Financial institutions, technology providers, and fintech companies alike need to provide a message to their stakeholders, be proactive, and, importantly, have solid recovery plans in place. They have to carefully consider (and reconsider) those recovery plans, with those strategies continually being updated as situations change.

When it comes to cybersecurity, firms need buy-in from senior management. This is another area where fintech firms may also have an advantage given their digital upbringing. Incumbents in financial services may find cultural frictions between the cybersecurity teams and the C-suite. This is where it becomes critical that the cybersecurity experts at banks and other financial institutions really understand their audience. They must spend time breaking down complex issues into simple, digestible terms. Additionally, cybersecurity teams should identify allies among their leadership teams and their boards to help encourage and drive a better environment where there is not fear, but rather a mutual understanding. This high-level strategy needs to come out of a real conversation between the technical experts and leadership.

As the internet of things (IoT) becomes increasingly intertwined with fintech services, and fintech providers and digitally enabled financial services incumbents collect exponentially more data from users, it is natural to worry about what is being done to protect that data. This worry has an added layer when that data resides with a third party, as is the case with cloud computing. It is also important to realize that consumers often have the right to "opt out" of sharing data. The question is, then, are consumers aware of what data is being collected and how that data is being used? This comes back to the idea of education with respect to cybersecurity and fintech. As the financial services industry invests more in cybersecurity solutions, they also need to realize that empowering customers will help lead to data trust, brand loyalty, and better customer experience.

It is important that the regulation of cybersecurity among financial incumbents, bigtech, and fintech firms ensures the private data is being protected. Service Organization Control 2 (SOC-2), a procedure developed by the American Institute of Certified Public Accountants (AICPA), examines the standardized technical audits for security, availability, processing integrity, confidentiality, and privacy.[7] It is specifically designed for SaaS providers to minimize the risk and exposure to confidential data. The certification of SOC-2 might demonstrate that the certified firm has high security against the cybersecurity risk, but it does not necessarily mean that the firm is risk-free. We can think of situations where a person has a driver's license but is still a bad driver.

[5] This deal was blocked in January 2021 by the U.S. Department of Justice. See "Visa and Plaid abandon merger after antitrust division's suit to block," https://bit.ly/3q1sbRY.

[6] Though not directly related to the topic of the present paper, the idea of market participants rewarding CSR compliant firms and penalizing firms with businesses that are at odds with CSR principles is a very active area of research. See, for example, Mackey et al. (2022).

[7] https://bit.ly/3laPpey.

For financial services firms and fintech firms alike to become more resilient in terms of their cybersecurity and establish a level of digital trust with their customers, it is essential to have validation processes at the federal level. Indeed, in the U.K., the Bank of England proposed something similar in 2019 [Jones (2019)].

One area that industry participants and regulators should be keenly aware of is the development and deployment of artificial intelligence (AI) from cloud-based platforms. As AI becomes increasingly more prevalent in financial services, and those algorithms are running off a cloud-based platform, validation and governance of these models, their data, and the underlying infrastructure will become paramount. Many of the AI algorithms being used by financial institutions and fintech companies are black-boxes to the employees of these firms let alone their customers.

What insights are the algorithms providing the companies about the users? This is an important question that needs to be addressed as well. Perhaps this is an area where fintech and technology companies can learn from the financial services incumbents. Banks and securities firms are required by their respective regulators to have rigorous model documentation and validation processes in place. Such documentation must highlight the assumptions, weaknesses, and limitations of the models that are used by the firm. Inputs must be "stressed" (i.e., taken to the most extreme values) to see if the models function properly. Any changes to the model over time must be catalogued and documented. With fintech companies largely flying under the regulatory radar (for now), many of them are not required to engage in these processes.[8] However, it is not a bad idea to begin a practice of validation, documentation, and governance with respect to machine learning (ML) models and AI algorithms at fintech companies. When things go wrong in the firm – whether it is a data breach, cyberattack, or algorithm misbehaving – investors and regulators demand transparency and accountability.

> *"Cloud technology plays a vital role in the fintech space by providing a more flexible and agile business model that is more readily able to adapt to changing market demands."*

As financial apps are increasingly being run off mobile devices, and residing on the cloud, biometric protections should also be an area in which financial services firms and fintech companies need to continue to improve. It is bad enough for customers to try to remember 14 different passwords across all of their accounts and financial services providers, but when these passwords are stolen, it is very easy for criminals to access their sensitive data. Whereas passwords can be hacked through brute force or stolen, biometrics leverage unique features that are physically unremovable from the customer and can be used across platforms, accounts, and service providers. When combined with multi-factor authorization or other biometric authentications, these protections can be very powerful. When facial recognition or someone's fingerprint is used to access data, an account, or any sensitive service, a simple text to the user's mobile phone or private email asking them to verify access can provide not only additional piece of mind but also added security.

5. CONCLUSION

5.1 For investors

Cloud computing still represents a major opportunity for investors despite the technology becoming increasingly mainstream. There are several approaches that investors could take to gain exposure to cloud computing technology. These are covered in greater detail in Imerman and Fabozzi (2020), who discuss investing in fintech innovations using their conceptual framework of a fintech ecosystem. One strategy is to find pure plays in the cloud computing space. This could be the aforementioned bigtech companies that control a large portion of the public cloud market and hybrid cloud strategies or going for niche cloud software companies that are developing more tailored, specific solutions for financial

[8] One exception might be robo-advisors and their automated investment tools, which are considered registered investment advisors (RIAs) by the Securities and Exchange Commission (SEC) and, therefore, "must describe the criteria and methodology used, including the tool's limitations and key assumptions," https://bit.ly/34FNEZf.

applications. To drill down even more into a particular sector of financial services – what Imerman and Fabozzi (2020) refer to as fintech verticals – investors can look for startup companies that are developing cloud-based solutions for digital banking, insurance, or wealth management. For investors looking to make a broad play on the overall cloud computing technology and its long-term growth, they can seek out an ETF that tracks indexes on cloud computing companies.

5.2 For regulators

This is actually a very exciting time for regulators to be exploring applications of new technologies to financial services. Cloud computing aside, for the moment, the next 10 years are going to see major advances in applications of quantum computing, blockchain and distributed ledger technology, IoT, as well as augmented reality and virtual reality being applied to financial services. Returning to the topic of cloud computing, many of the aforementioned emerging technologies rely on, or are fully integrated with, cloud computing platforms. And, as we noted earlier, of all the emerging technologies making their way into financial services, cloud computing is one of the more mature in terms of adoption and utilization. For both of these reasons, financial regulators need to remain vigilant in their ongoing monitoring of how cloud computing is being used by financial services firms, from banks to insurance companies to broker-dealers. Understanding how data is managed, handled, and stored is important for ensuring the integrity of the models that are using the data as well as to protect said data from cyberattacks and breaches. For this reason, cybersecurity in the cloud is likely to continue to be an important issue going forward. Furthermore, as AI models run off the cloud, having a framework for validating not only the models but the processes and the data (inputs and outputs) will be increasingly important for regulators to monitor in their supervisory efforts.

5.3 For startups

Any entrepreneur looking to provide innovations in the fintech space ought to be familiar with the paradigms of cloud computing. That is because fintech startups – unlike the incumbents in the financial services industry – are cloud native. This has many benefits over the incumbents, who quite frankly can learn from their startup competition. One benefit is the agility and flexibility that cloud-based solutions provide the company. The cost-benefit of pay-as-you-use storage is also beneficial to a startup that needs to be careful with every invested dollar of capital. Decisions must be made about whether a private, public, or hybrid cloud should be used; however, again, the ability to pivot from one strategy to another is much easier in a cloud environment than it would be with a data center filled with servers or a basement of mainframe computers. Then deciding what software and/or models are going to be run on in-house hardware versus off the cloud becomes both a strategic and an economic decision. We are likely to see the trend of increasing amounts of software and models run off the cloud. But with that point we should remind startups to consider the risks – operational, cybersecurity, systemic, etc. – associated with being fully dependent on the cloud.

5.4 For incumbent financial institutions

The time is now to migrate from mainframe and server-based system to cloud-based storage and software. In this market environment, where innovation moves at the speed of now, it is imperative to embrace a more agile mentality when it comes to IT systems so as to not lose more ground to startups, which are cloud native and have agility in their proverbial business DNA. That being said, such migrations are not without their risks. Cybersecurity issues, which have been highlighted in this article, must be addressed with contingency plans in place in the event of a breach. Furthermore, relying on one vendor for cloud services is risky from the standpoint that if something happens to that provider it could dramatically affect the institution's operations potentially for a long period of time. There is also the issue of systemic risk, which was not a main focus of this article but is certainly an area that warrants much more examination from academic researchers and regulators alike. Given that the public cloud is essentially an oligopoly – made up of Amazon's AWS, Google's GCP, and Microsoft's Azure (with IBM as a close fourth though their recent strategy seems more focused on a hybrid cloud) – should something happen to one of these companies or their respective products, it could represent a massive shock to the global financial system to the extent that the world's largest banks and clearinghouses are relying on those specific cloud products.

REFERENCES

Flextra, 2020, "2020 state of the cloud report," https://bit.ly/3MGSMxz

Ghule, S., R. Chikhale, and K. Parmar, 2014, "Cloud computing in banking services," International Journal of Scientific and Research Publications 4:6, 1-8

Imerman, M. B., and F. J. Fabozzi, 2020, "Cashing in on innovation: a taxonomy of fintech," Journal of Asset Management 21, 167–177

Jones, H., 2019, "Bank of England calls for 'super shield' against cyber attacks," Reuters, May 14, https://reut.rs/3J5v6k6

Mackey, T. B., A. Mackey, L. J. Christensen, and J. J. Lepore, 2022, "Inducing corporate social responsibility: should investors reward the responsible or punish the irresponsible?" Journal of Business Ethics 175:1, 59-73

Menchaca, J., 2018, "DevOps concepts: pets vs cattle," May 6, https://bit.ly/36fzUVl

Sampson, D., and M. M. Chowdhury, 2021, "The growing security concerns of cloud computing," in 2021 IEEE International Conference on Electro Information Technology (EIT), pp. 050-055

Scott, H. S., J. Gulliver, and H. Nadler, 2019, "Cloud computing in the financial sector: a global perspective," Program on International Financial Systems 2019, https://bit.ly/37sxnYO

Tissir, N., S. El Kafhali, and N. Aboutabit, 2021, "Cybersecurity management in cloud computing: semantic literature review and conceptual framework proposal," Journal of Reliable Intelligent Environments 7:2, 69-84

Yan, G., 2017, "Application of cloud computing in banking: advantages and challenges," Advances in Economics, Business and Management Research (AEBMR) 23, 29-32

MULTI-CLOUD: THE WHY, WHAT, AND HOW OF PRIVATE-PUBLIC CLOUD SETUPS AND BEST PRACTICE MONITORING

FLORIAN NEMLING | Senior Consultant (Austria), Capco
MARTIN REHKER | Managing Principal (Germany), Capco
ALAN BENSON | Managing Principal (Germany), Capco

ABSTRACT

Banking, as any business, is complex, and there are many choices and decisions to be made – all the time. Underlying many businesses today is expensive technology that adds to the endless decision complexity. When mapping their IT strategy, companies need to evaluate opportunities and challenges presented by cloud technology, and many businesses find themselves with a mixed private and public cloud setup. This article explores important issues with multi-cloud scenarios, with a focus on the monitoring of multi-cloud solutions.

1. INTRODUCTION

The uptake of cloud technology by large, high-profile global companies, its popularity within IT departments, its promise of "almost free" (or at least low cost), and the rise of cloud-based solutions have predictably forced companies to scrutinize and reconsider their traditional legacy systems and setups.

Every significant company must develop a cloud strategy defining if and how they should move their existing setup to the cloud. They will likely start the transformation process implementing a private cloud solution on on-premises infrastructure to achieve a scalable and more manageable system.

However, as business activities expand, it becomes inevitable that the need for more computing resources or external software services grows to the point where the company faces a dilemma – expand the private cloud resources or acquire resources and services from a public cloud provider. Companies must consider compelling arguments linked to the benefits of public cloud use. Once a company starts to utilize a public cloud provider, the result is a hybrid-cloud solution, a combination of private and public cloud services. A complete cut-over to public cloud is rarely possible as some IT infrastructure usually needs to remain onsite or in a private environment.

There are many benefits of using a public cloud, like the flexibility in deploying new environments within minutes or quick reaction to peak workloads, to name just a couple. This results in better employee experience and efficiency. The quicker you can deploy and scale, the quicker people can get to projects using cloud. Public cloud providers offer solutions that enable quick business transformation journeys and faster go-to-market requirements.

Computing resources have become a commodity. Businesses are aware of the fact that they cannot differentiate solely based on technological maturity linked to the usage of their infrastructure. Public cloud providers are generating and offering more and more new cloud based and cloud native services that reinforce technological innovations and modernization. Adopting these companies' services brings additional value to businesses.

Public cloud providers have also already addressed many regulatory and legal requirements and continue to do so. This is paving the way for a broader adoption of cloud services – increasingly also in highly regulated industries.

Driven by these very compelling arguments, many companies will chose to expand their private cloud infrastructure with public cloud offerings. This results in an ever-increasing number of multi-cloud setups.

2. MULTI-CLOUD COMBINES THE BEST OF BOTH WORLDS

The scenario described above results in the use of multi-cloud solutions, as there are compelling arguments and regulatory restrictions to retain some of the application landscape in a private cloud infrastructure.

Key attributes of both private and public clouds are discussed below.

2.1 Private cloud

- **Security aspects, regulatory, and legal requirements:** certain systems and data registers may not be placed into a public cloud environment. Possible reasons are constraints regarding geographical placement and the required physical access (e.g., emergency infrastructure, co-location requirements, bandwidth, and data restrictions).
- **Financial aspects:** hardware that is already owned and still not depreciated (and software if it cannot be reused) can be used for running a private cloud design.
- **Services or functionalities:** some services cannot be provided by a public cloud provider or are uneconomical to adapt for a public cloud scenario (e.g., legacy and custom-built applications).

2.2 Public cloud

- **Scalability:** complete flexibility to react to resource needs instantly without requiring oversized hardware for everyday business.
- **Dynamic creation of environments:** development and test activities benefit from unrestricted and automated creation and destruction of environments on demand.
- **Innovative services:** bringing in and leveraging innovative, external cloud-based services is easy.
- **Evaluation and proof of concept (PoC):** evaluation of products in cloud setups is very easy and can usually be done without financial commitments outlasting the duration of a PoC.

These examples show why a multi-cloud solution is often used, as it is necessary for companies for compliance reasons and they can also get the best of both worlds. However, even though a multi-cloud solution can bring considerable benefits for a company, it also brings inherent risks, specifically with monitoring.

3. CURRENT AND TARGET STATE OF MULTI-CLOUD MONITORING: Q&A ON THE MOST IMPORTANT ASPECTS

3.1 How do customers currently manage their multi-cloud environments? What is still left in their legacy infrastructure management landscape?

On the one hand, public cloud providers offer dedicated monitoring solutions like AWS CloudWatch, Azure Monitor, or GCP's operations suite. Monitoring of legacy infrastructure and private cloud domains, on the other hand, is dominated by on-premises hosted products such as Zabbix, Nagios, or Dynatrace. Both options are usually combined with third-party cloud and local services used for application monitoring. The use of distinct and isolated monitoring solutions, however, comes with operational risks and difficulties linked to the management of multi-cloud workloads.

3.2 What are the current challenges with multi-cloud management and their effect on resource usage?

The challenges related to monitoring multi-cloud solutions include operational factors such as:

- The support and operations teams must manage multiple monitoring platforms – this has a direct impact on resource utilization.
- Non-standardized monitoring frameworks result in different monitoring specifics. When used in parallel, the different metrics and threshold levels lead to confusion and loss of control. This can result in non-compliance with "service level agreements" (SLAs).
- Building a central event management system, an "IT service management" (ITSM), with ticketing or emailing, and multiple monitoring solutions linked to it using webhooks creates complex integration overhead.

- The lack of a centralized visualization interface results in an increased need of resources.
- The lack of dependency links between the different monitoring solutions obstructs a holistic view. A raised alarm is followed by complicated and time-consuming processes of incident troubleshooting and problem root-cause identification. It is expensive and prone to missing important factors that may significantly contribute to the issue's origination.
- The lack of dependency information can lead to wrong decisions by operations engineers because a metric or an alarm cannot be put into the context and thus loses meaning. The consequence is an increased risk of service unavailability with the potential of system-wide crashes.
- Meeting end-to-end business and customer SLAs becomes more challenging as the dependencies between systems monitored by different solutions are not taken into account.

Even though this is just a short, non-exhaustive list of challenges, they need to be addressed.

3.3 What should be the target state of multi-cloud monitoring? How does this effectively support customers on their journey (innovation, transformation, digitalization) with a hybrid cloud solution?

An ideal and solid multi-cloud monitoring system needs to have attributes of high availability and centralization. A quick win might be to choose a "monitoring-as-a-service" (MaaS) solution where a predefined, preconfigured, and scalable monitoring service is hosted in a public cloud. Solutions from all major monitoring providers can be found as instantly deployable services within public cloud providers' marketplace sections.

Another solution consists of creating and managing your own centralized monitoring system installation where either your local infrastructure or private/public cloud hosted infrastructure stack supports performance monitoring of infrastructure and application layers. Additional attributes include a fully customizable visualization interface and flawless compatibility with mobile platforms as a must. Finally, a wide support of API and webhook connectivity options towards mostly used and popular ITSM tools is an important attribute to facilitate an effective alert/event management.

The decision to choose between MaaS or own installation of monitoring solution should follow the standard cloud service adoption approach and analysis. It should consider factors already mentioned: regulatory requirements, company's infrastructure standards, availability, capacity of internal computing, human resources, willingness to invest and modernize, number of monitored services, as well as employee skillset and prior experience linked to legacy monitoring solutions.

3.4 What is a typical approach of implementing a multi-cloud monitoring solution?

The generic approach of implementing a new multi-cloud monitoring solution should contain the following steps and milestones:

- Identify the pain points and missing information that would protect the business.
- Together with relevant business units, determine what needs to be monitored. Outline how they benefit.
- Select a suitable monitoring solution based on requirements.
- Set up an implementation strategy – MaaS or own installation.
- List metrics, key performance indicators (KPIs) to monitor and events to log.
- Develop an incremental implementation strategy based on criticality and value.
- Deploy the solution for a trial run and set up alerts for high-priority events and metrics. Establish and test connections to external tools used as part of the ITSM and initiate business reporting.
- Check the functionality and reliability of the new monitoring system in a redundant set-up of the "old" and the newly deployed solution. Confirm that the new solution is reporting events correctly.
- Complete the setup of metrics monitoring and event logging.
- Control the metrics for accuracy and tune thresholds to the desired targets.
- Decommission legacy monitoring systems.

3.5 How to choose the right multi-cloud solution

There are many solutions on the market and it is difficult to make the right choice. During the selection process, you should consider the following questions:

- What type of monitoring are we seeking? The most common types are network and security, infrastructure availability and capacity, application performance and availability, web performance and user interaction monitoring, and business performance metrics (e.g., the number of processed documents, transactions, sales results, etc.).

- What is the preferred management technology and required information granularity?
  - **Agent monitoring:** typically designed for a specific platform and vendor, can collect more detailed information.
  - **Agentless monitoring:** uses standardized protocols and log analysis. In general, ready-to-deploy solutions are hosted in private/public clouds and usually have great built-in agentless monitoring possibilities.
- How many devices and applications are to be monitored?
- What is the company's licensing strategy? Would it rather invest, rent, or use an open-source solution, potentially contributing own development to implement specific functionalities?
- How likely is the IT landscape to grow and will the scope for monitoring need to be extended? Is the candidate solution flexible enough to support the growth strategy?

3.6 What is the right adoption and implementation strategy?

Regarding the adoption and implementation strategy, it is important to also consider some factors that are in favor of a MaaS solution. Generally, product licensing is already included in the operational costs, so there is no need to procure additional software licenses. Ready-to-use tools can be deployed instantly and provide flexibility in terms of features testing and creation of proofs of concept (PoCs).

Additionally, MaaS providers offer support 24/7. SaaS solutions allow more time to focus on business demands rather than worrying about the effort needed for monitoring maintenance and reliability, as well as offering the quick adoption of new services, components, and assets into the monitoring solution.

3.7 How to establish quick wins

By selecting and implementing a suitable multi-cloud monitoring solution, one of the earliest benefits a company can profit from is the possibility to review and manage the system utilization patterns. This allows for optimal resource allocation and manging the running costs effectively. The resulting increased availability of services, greater performance and cost transparency, as well as reduced critical incidents will improve business experience and strengthen relationships between IT operations and business.

If a business runs a correctly and purposefully set up multi-cloud monitoring solution, it will provide additional and significant benefits to different business areas:

FinOps:
Unmanaged utilization and deployment of new services using public providers can become very expensive. Monitoring is necessary to avoid cloud cost leaks.

SecOps:
Early alerting and effective response to security threats results in decreased costs linked to security issues. This will improve the security of applications and network layers in the long run.

CloudOps:
Simple and effective scaling when needed, either manually or automatically, using resource scaling scripts triggered by events, optimizes service availability and ensures positive end-user experience.

DevOps:
Embedding monitoring into the early stages of the DevOps cycle helps DevOps engineers to identify and fix issues at a much earlier stage. That way they can achieve and maintain optimal application performance.

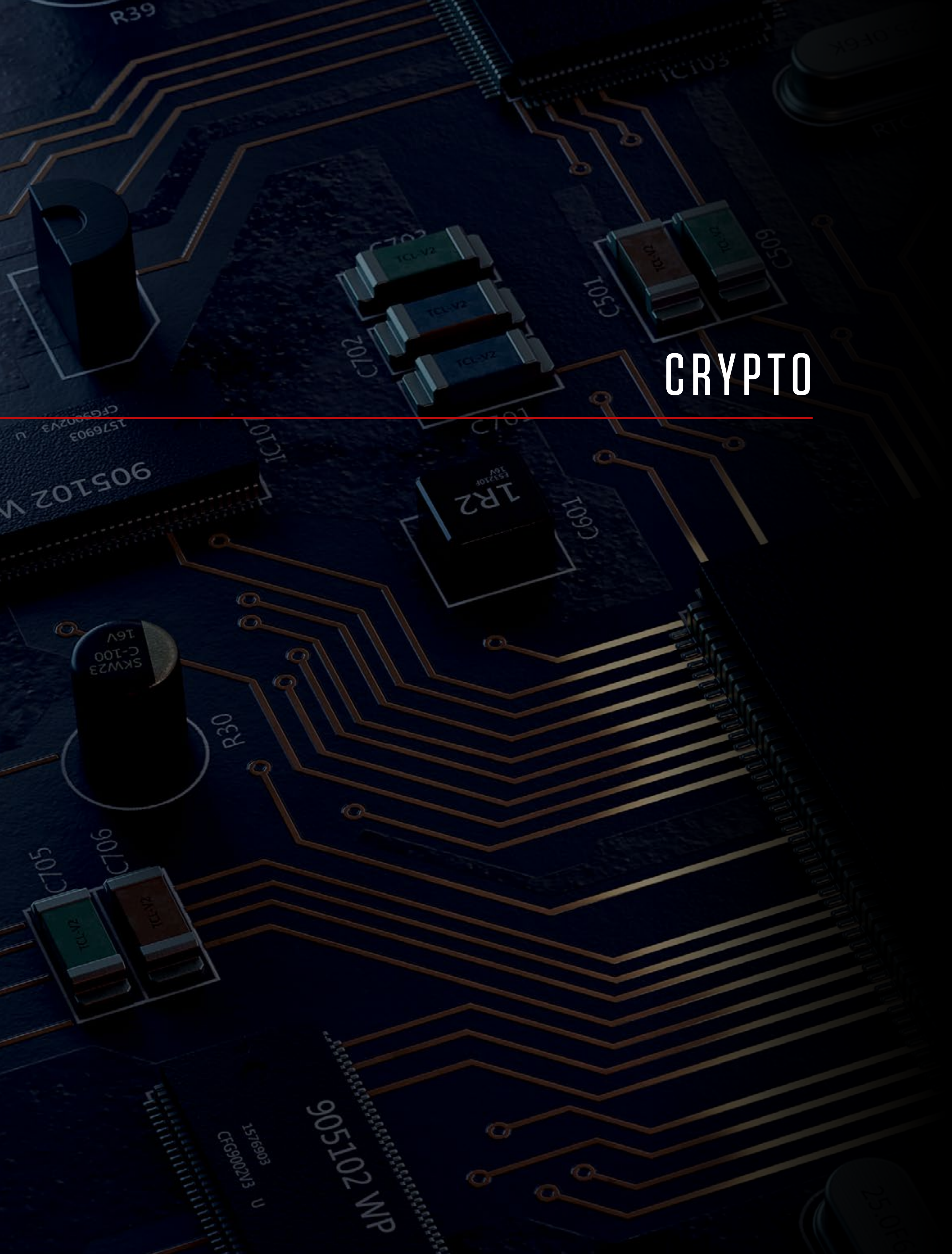
4. CONCLUSION

Many companies are running workloads, applications, and systems in a distributed landscape with a combination of public cloud providers, private clouds, and legacy on premises infrastructure. Without the right solution, the monitoring of all these environments can become a useless overhead in the best case and a danger to the business in the worst.

Consequently, we regard a multi-cloud monitoring solution as an important infrastructure item in this scenario. Implementing multi-cloud monitoring will require investment, but it is a crucial step towards operating a stable business and achieving lower total costs.

Monitoring systems are "behind the scenes" technology and are often only thought of when systems fail. Subsequently, getting investment and interest from the business can be challenging.

However, all business stakeholders need to be involved in the discussions about monitoring solutions, so that their needs and requirements are understood and considered. That way, they can also appreciate the value monitoring brings for them and how it facilitates effective growth and profitability for the overall business.


R39
IC103
C702
C701
C501
C509
TCL-V2
1R2
16V
C601
905102 W
IC107
1576903
CFG9002V3
SKW23
C-100
16V
R30
C705
C706
905102 WP

CRYPTO

DIGITAL ASSETS AND THEIR USE AS LOAN COLLATERAL: HEADLINE LEGAL CONSIDERATIONS

PHOEBUS L. ATHANASSIOU | Senior Lead Legal Counsel, European Central Bank[1]

ABSTRACT

Although substantial in terms of market capitalization, the economic potential of digital assets remains locked, inter alia, on account of their still limited use as loan collateral. The wider use of digital assets as security for credit would both help their holders to capitalize on their digital asset holdings and contribute towards easing liquidity conditions in the market by allowing market actors at both ends of a lending agreement to tap into a substantial, but largely unutilized, repository of collateral. This article explores some of the legal parameters relevant to the use of digital assets as collateral, with an emphasis on how a security interest in digital assets can be created, the modalities for the realization of digital assets accepted as loan collateral, and the ways in which collateral takers (but also collateral givers) can be protected from fluctuations in the value of some of the more volatile types of digital assets tendered as loan collateral.

1. INTRODUCTION

Entrepreneurs routinely need access to credit, and one of the most obvious ways for them to obtain it is by securing their borrowing obligations with something of value, by way of collateral. Apart from being a precondition for entrepreneurs to secure credit in the first place, the availability of good collateral will also determine the affordability of the interest rate. Conversely, without access to good collateral, the funding of research, development, and business growth can be challenging for many a business owner. That being the case, it can be argued that it makes sense for entrepreneurs to only invest in assets that they can readily tender as collateral, should the need to secure extra liquidity arise in the regular course of business.

Recent years have seen substantial growth in the market capitalization of digital assets and, although this has not been linear, the value of the global digital assets market has increased significantly over time.[2] Because of the considerable financial value they represent, digital assets should also be expected to embody substantial potential as a source of good collateral. If digital assets have yet to be "integrated" into the mainstream financial system, and if their appeal remains relatively limited, this is also because of the relatively limited degree to which they were being used, at the time of writing, as loan collateral. The result is the "immobilization" of a substantial repository of value that, if ever exploited to the fullest degree possible, could support the injection into the real economy of much needed liquidity. What largely accounts for the hitherto reluctance of collateral providers and takers to

[1] The views expressed in this article are those of the author and are not in any way representative of the views of the ECB or the Eurosystem. This article builds on Athanassiou, P. L., 2019, "Cryptocurrencies and their use as loan collateral: core legal considerations," 34 Butterworth's Journal of International Banking and Financial Law, and on a draft Report of the European Law Institute, entitled "Use of digital assets as security" (forthcoming), of which the author of this article is the main author.

[2] In November 2021, the market capitalization of the global cryptocurrency market alone was estimated at about U.S.$3 trillion: Sanyal, S., 2021, "Global cryptocurrency market cap reaches the U.S.$3 trillion mark," Analytics Insight, November 10, https://bit.ly/3rhbVx8.

tender and to accept, respectively, digital assets as security is their relative novelty, and the attendant legal uncertainty surrounding the possibility of, and the conditions for, their use as loan collateral. Although I take no position on the advisability of using digital assets as collateral nor, indeed, on the suitability of all of the different types of digital assets as security, I do, nevertheless, see scope for non-partisan reflection on the use digital assets as loan security, so that their full economic potential can be exploited by those who, for whatever reason, choose to invest in and to hold them in their portfolio.

The aim of this article is to explore the basic legal conditions to be fulfilled for digital assets to be used as loan collateral. The analysis in this article revolves around three core questions. The first is how to create a valid security interest in digital assets, and what conditions may need to be fulfilled to facilitate such creation. The second is how collateral takers can concretely realize digital assets they have accepted as collateral in the event of the collateral giver's default on the loan obligations secured against the use of digital assets as collateral. The third is how collateral takers and collateral givers alike can be protected from fluctuations in the value of digital assets that they have accepted or tendered, respectively, as collateral.

Before turning to the substance of my analysis, four remarks are apposite by way of clarification regarding the ambition and scope of this article. The first is that, except where the context requires otherwise, the terms "security" and "collateral" are used interchangeably throughout this article. The second remark is that the scope of the analysis in this article is limited to digital assets that are amenable for use as collateral. What the main characteristics of those digital assets are is explained in Section 2. The third is that none of the three questions explored below can be definitively answered in a legal vacuum, in other words without reference to a specific system of rules to govern the private (mostly property) law effects of holding and transferring control or ownership of digital assets. To avoid tying the analysis in this article to the rules of any particular jurisdiction we draw attention, below, to the main lines of inquiry that collateral providers and their lawyers would need to pursue, regardless of jurisdiction, to determine whether and how digital assets could be used as loan collateral. Finally, the emphasis of this article is on the legal parameters of the mobilization of digital assets as collateral; accordingly, more practical considerations, such as the vulnerability of digital assets to cyberattacks or to market manipulation, important as they are (also in terms of the safe use of digital assets as collateral and their price stability), will not be considered in this article.

2. DIGITAL ASSETS SUITABLE FOR USE AS COLLATERAL

The notion of "digital assets" is closely associated with the relatively recent emergence of distributed data storage technologies and platforms. A survey of the field testifies both to the considerable breadth of that notion and, no less significantly, to the objective difficulty of defining "digital assets" in a monolithic way, given their many variants, the substantive differences amongst them, and the constant evolution in this space, which has, over time, seen new categories of digital assets added to those already in existence.

Not all digital assets are suitable for use as collateral. A digital asset should have at least three qualities before it can be used as loan security. First, it should embody "value", which is to be understood in economic terms (this would, for instance, exclude social media accounts, which, although digital, need not always embody economic value, whatever their emotional worth for their holders). Economic value may either be associated with the asset itself (for instance, in the case of a cryptocurrency or a digital-only security) or be derived from a tangible, real-world asset, which the digital asset either reflects (as in the case of an asset-backed token) or which is there to guarantee the digital asset's price stability (for instance, in the case of so-called "stablecoins"[3]). Second, there should subsist in it a "de facto right of exclusive use", defined as the right to access and enjoy the economic value that a digital asset embodies. Despite their intangible nature, and the uncertainty surrounding their legal characterization as subjects of property law, digital assets can be the object of exclusive control and so-called "rivalrous"[4] enjoyment, which are preconditions for the creation of security interests in them.[5] Finally, a digital asset should have the attribute of "certainty", which is a prerequisite both for the exercise of control over it and for its assignability, in accordance with the terms of a security agreement.

[3] The reference is to class of privately-issued means of payment designed to maintain a stable value relative to fiat currencies by being linked to a "safe" asset or to an external pool of liquid "reserve assets", including cash deposits.

[4] The reference is to the economic quality of certain assets or goods that can only be used or consumed by specific people if their supply or value are not to be adversely affected. It is the risk of the depletion of their supply and the depreciation in their value that accounts for competition (rivalry) with regard to their exclusive use and consumption.

[5] Unlike digital assets, digital data may lack the attribute of certainty, with an impact on assignability and the exclusivity of control over them.

The types of digital assets meeting those qualities include cryptocurrencies and stablecoins, uncertificated (i.e., electronic-only) financial assets (such as security tokens), non-financial asset type tokens (including utility[6] and certain payment[7] tokens), and hybrid tokens (i.e., digital assets that share some of the characteristics of more than one digital asset class).[8] It follows from the foregoing examples of digital assets amenable to use as loan collateral that these encompass both "pure" digital assets (denoting those that have been created and only exist in the digital world, in the form of tokens representing a unique set of valuable attributes, such as cryptocurrencies or security tokens) and "asset-backed tokens" (i.e., digital representations of already existing, physical assets, such as tokenized securities or bonds, tokenized gold bullion, tokenized real estate or patents), as well as so-called "non-fungible tokens" (NFTs), such as tokenized works of art or collectibles.[9]

3. CREATING SECURITY INTERESTS IN DIGITAL ASSETS

Except where it is the object of specific regulation, the creation of security interests in digital assets is, at present, an area of considerable legal uncertainty. This is because of the relative novelty of digital assets as an asset class, as well as their cutting-edge technological underpinnings that make it difficult to "localize" many of them in any particular jurisdiction, the laws of which would govern their use as security. As many digital assets tend to lack a physical location, it is fair to speak, in their case, of a "notional" location. This will depend on a multitude of factors, including the manner of their holding.

The determination of the modalities for the creation of a security interest in digital assets will therefore require, on the one hand, the determination of their notional location at the time of their use as security and, on the other hand, the analogous application to them of national law rules applicable to more conventional asset types. Put differently, to determine the conditions subject to which a security interest can be created in a digital asset, two questions need to be answered. The first is: "What is the law applicable to a particular digital asset?" In other words, within which national legal system's remit a digital asset unit is deemed to fall. The second is: "What type of asset does a given national legal order consider a particular digital asset to be?" These two questions are clearly linked to one another: to work out the requirements for the creation of a security interest in an asset, one must first determine the law applicable to creation. In turn, the type of the asset in question and, in particular, its legal characterization in a given jurisdiction will play a key role both in determining the applicable law (i.e., the law of the creation of security interests in that asset) and in applying it, by helping to identify the types of security interest that can be created in an asset as well as the applicable requirements for creation (e.g., in writing and/or by way of registration).

The answer to the first question (i.e., what is the law applicable to a digital asset) goes to the core of what digital assets are, and, unless definitively answered, it is apt to derail the efficiency of any attempt to establish an enforceable security interest in a digital asset. Taking the example of cryptocurrencies, these can be held in one of three different ways: either directly on the relevant distributed ledger, through an online "wallet" (custodian or non-custodian), or in a "cold storage" device (typically, in the cryptocurrency holder's personal computer – one that is not connected to the internet – or in another "remote" hardware storage device, such as a USB memory stick or an external hard drive). As suggested above, the way the cryptocurrency is held will largely determine the answer to the first question. While the jurisdiction of a given cryptocurrency unit may be easy to determine when held in a wallet (on- or off-line), the same will not be true if the same unit is held directly on the blockchain, which resides, simultaneously, everywhere and nowhere. It follows that only some cryptocurrency holdings may lend themselves to being used as loan collateral, since their jurisdiction (and, by implication, also the law governing the creation and establishment of an enforceable security interest over them) will only be ascertainable in some cases, but not in others.

[6] The reference is to a class of programmable digital asset that grants to its holder the right to exchange it in the future for products or services, actual or under development, digital or physical, which are provided (or are intended to be provided) by the token's issuer. Utility tokens both enhance their issuer's ability to quantify the value of the right that is the object of the token-issuance transaction and facilitate its transfer.

[7] The concept of payment (or currency) tokens refers to digital, non-financial assets aiming to fulfill the properties of "fiat" money, but without amounting to legal tender.

[8] One example of a hybrid token would be a digital asset that both represents a share of ownership in a company and entitles its holder to the right to receive the first product or service that the said company manufactures.

[9] NFTs are cryptographic, digital tokens that represent objects in the real (or the digital) world, such as underlying works of art or collectibles, and may (but need not) embody ownership rights. Their creation and authentication rely mostly on the use of the Ethereum blockchain, utilizing digital signatures to guarantee their uniqueness and indivisibility (hence, also, their non-fungibility).

The answer to the second question (i.e., "What type of asset does a given national legal order consider a particular digital asset to be?") depends mostly on national law considerations, and, without being infinite, the range of options is considerable. Different regulatory and supervisory authorities in different jurisdictions have, at different times, declared different types of digital assets to be "money" or "currency", "securities" or "investment contracts", "commodities", or sui generis digital (intangible) assets.

If the question of the categorization of digital assets arises in the context of a jurisdiction where no regulatory, supervisory, or judicial pronouncements exist, or in a jurisdiction where conflicting judicial or regulatory pronouncements have been made in respect of the categorization of particular types of digital assets, it is only with the benefit of prior legislative intervention that this question could definitively be answered. In any event, which asset class a particular type of digital asset is deemed to fall into is a key question: the answer to that question will determine the "form" of security interest that can be created over it[10] and the "manner" of its creation (different formality requirements will typically apply to the creation and perfection of different forms of security interest, such as registration, the taking of physical control over collateral, or the exercise of possession thereof).

To conclude, in order to create a security interest over a digital asset, clarity is indispensable, both in terms of its "location" and in terms of the "type of asset" that this is deemed to be for the purposes of the national legal order whose laws govern the creation of security interests in it. On the question of location, prior regulatory intervention would appear necessary, at least in those jurisdictions where the legal status of digital assets is either unregulated or otherwise unclear on account of conflicting regulatory, supervisory, or judicial pronouncements. Absent such intervention, many national legal systems are poorly equipped, at the present juncture, to address the issue of the creation of enforceable security interests in digital assets. On the question of the asset categorization of particular types of digital assets, which is crucial to determine the process of, and the modalities for the creation of a security interest in them, legal clarity is also desirable, at least in those jurisdictions where contradictory pronouncements have been made in respect of the legal characterization of digital assets.

> *"The creation of security interests in digital assets is an area of considerable legal uncertainty but, also, great commercial promise."*

Finally, it bears noting that some legal systems approach the question of the creation of security interests separately from that of their third-party effectiveness (i.e., their legal enforceability on third parties with claims over the same asset).[11] For those legal systems, these two questions would need to be addressed independently from one another, with a view to assessing whether or not a security interest created in a digital asset would also enjoy priority over any subsequent claim over the same asset.

4. REALIZATION OF DIGITAL ASSETS USED AS COLLATERAL

Once there is clarity on the asset categorization and location of a digital asset to be used as collateral, the collateral provider and the collateral taker can enter into a security agreement, inter alia, describing the loan collateral by reference to an accurate description (to distinguish it from other digital asset holdings of the collateral provider) and to establish the extent of the collateral taker's security interest in it. The security agreement will only be of value if the collateral taker can realize the collateral in the event of the collateral provider's default on their payment obligations.

Because digital assets are intangibles, they cannot be seized and enforced upon as one might do with tangibles. The modalities for the enforcement of a secured creditor's

[10] The choice will typically be amongst an assignment, a pledge, a mortgage, or a charge, fixed or floating (more than one of these collateralization techniques may also be applicable).

[11] Examples also exist of jurisdictions that apply the same set of requirements to the creation of security interests as well as to their third-party effectiveness.

rights in them will depend on their attributes. For instance, if the digital asset used as security is a token, given as (non-possessory) security to a secured creditor, the latter can only realize such token if they have access to the debtor's private key. One way to overcome the debtor's refusal to grant access to their private key is for the security agreement to foresee the debtor's entry into an escrow agreement with a trusted third party, transferring to that third party, for safekeeping, the private key to the token. Acting as escrow agent, the third party would then cooperate with the secured creditor in the event of the debtor's insolvency, to enforce the creditor's security right (e.g., through a sale of the token to satisfy the secured creditor's claim). Whatever the particular attributes of the digital assets used as security, it will be clear from the aforementioned example that their realization as loan collateral will call for the exercise of "effective control" over them, whether by the creditor or by a third party that both parties trust to hold the collateral for the duration of the creditor-debtor relationship.

What it takes to exercise such effective control over a digital asset is not a question that can be addressed without reference to its features. Using the example of cryptocurrencies, such as bitcoin, it should be noted that their effective owner is the holder of the private key to the account where that cryptocurrency is held.

Considering that, for instance, bitcoin units are linked through one public and one private key to a bitcoin address – or "account" – through which they can be sent, received, or stored, their transfer involves moving those units from one electronic address (within or outside the bitcoin blockchain ledger) into another. It follows that what is essential for the exercise of effective control over a bitcoin unit held in the bitcoin blockchain ledger is control over the private key to the account where this is held, whether directly on the bitcoin blockchain ledger or indirectly through a wallet (on-line or off-line). It also follows that, for as long as collateral providers keep their private keys private, they continue to exercise control over their bitcoin holdings, which they can transfer at will, without the creditor's knowledge (the bitcoin blockchain ledger will record bitcoin transfers but, crucially, it will not record borrowings or security interests in bitcoin units).

As the reader will have deduced from the comments above, borrowers who use digital assets as collateral have an incentive to maintain control over the private key to their account. For their part, collateral takers have an interest in monitoring their borrower's ability to dispose of their cryptocurrency holdings, to ensure that the protection they enjoy, as collateral takers, will not prove illusory in the event of the collateral giver's default on their repayment obligations. Mutual distrust is bound to have an adverse effect on the readiness of both

parties to a lending transaction to part with their funds (in the case of a creditor) or with their digital asset collateral (in the case of a debtor who is a holder of digital assets).

As suggested earlier, one way to "square the circle" is by involving a trusted third party in the process. One such suitable third party is a wallet provider, willing to act as escrow agent for the duration of the creditor-debtor relationship between the collateral provider and the collateral taker. Crucially, provision would need to be made in the security agreement against the risk of the wallet provider's insolvency, which could result in a situation where either the cryptocurrency units tendered as collateral or the private keys to the account(s) where these are held become part of the wallet provider's bankruptcy estate.

There are other ways in which trust can be established between a creditor and a debtor, even without the involvement of a trusted third party. One is through the physical delivery, by the collateral giver to the collateral taker, of control over the collateral, but in a form that protects the collateral provider against the risk of its non-return after the loan has been repaid in full. This could, for instance, be achieved by collateral providers handing-over possession of their digital assets and, more specifically, of the private key to the account where these are held in the form of an encrypted storage device. Although practical, this solution would not protect either party from the risk of the physical loss of the encrypted storage device (this would entail the definitive loss of control over the digital assets tendered as collateral).

Another alternative is through recourse to a smart contract[12] between a lender and a borrower, written on a blockchain or another DLT-run platform (including that of a wallet provider). The aim of the smart contract would be to automate the process of the realization of collateral in the event of the borrower's default on their repayment obligations. Alternatively, the smart contract could be used to release the collateral after the borrower has complied with their repayment obligations, without any possibility for the parties to the security agreement to tamper with the collateral for the duration of the creditor-debtor relationship, and without the need for third-party intermediation, provided the lender's and the borrower's technology and processes are consistent with their participation in a shared platform to host the smart contract.

To conclude, because the process of realizing collateral in the form of a digital asset will involve the exercise of effective control over it, and because effective control over a digital asset necessitates control over the private key to the account where that digital asset is held, the parties to a lending transaction will need to devise ways in which to protect their legitimate interests in the loan collateral, without intruding too much into those of their counterparty.

The three ways in which this can be achieved are by involving in the process a trusted third party (e.g., a wallet provider willing to act as escrow agent), by arranging for the physical delivery of control over the collateral, but in a way that shields the borrower against the risk of its non-return, or by resorting to a smart contract. The first avenue could prove workable, but it is, arguably, difficult to square with the disintermediation goals of digital financial innovation. The second option is vulnerable to the loss of collateral, while the third one presupposes the use of a third-party intermediary or the use, by both the collateral provider and the collateral taker, of technology and processes that are compatible with the use of smart contracts.

5. PROTECTING THE PARTIES TO A SECURITY AGREEMENT FROM FLUCTUATIONS IN THE VALUE OF DIGITAL ASSETS

The valuation of assets offered as security may present certain challenges, especially where these are intangible, as in the case of digital assets. What is more, certain types of digital assets are notoriously volatile. The flagship type of digital assets that is prone to volatility are cryptocurrencies. To draw on the example of bitcoin, its price fluctuated between U.S.$19,783.21 on December 17, 2017, and U.S.$3,874 in early March 2019, following the crash of 2018 (itself preceded by a massive wave in appreciation in the course of 2017; it is telling that, in December 2016, bitcoin's price stood at a mere U.S.$930). On December 31, 2021, the price of bitcoin stood at a staggering U.S.$45,800. Other cryptocurrencies, including Ethereum, have also displayed a similar pattern of volatility. Because of their volatility, which tends to exceed that of more "traditional" assets, cryptocurrencies used as security may appreciate or depreciate substantially in value during the lifetime of a security agreement.

[12] The reference is to a software protocol (i.e., computer code), which is executed automatically (hence, without human intermediation), as soon as certain pre-programmed conditions, agreed upon between the parties to the smart contract, have been satisfied.

The volatility of cryptocurrencies need not prove fatal to their use as collateral, provided the parties to a lending arrangement have factored in the risk of their eventual appreciation or depreciation. One way in which this can be achieved is by the parties making use of a smart contract to track fluctuations in the value of cryptocurrencies tendered as collateral, and to either trigger a "margin call", in the event of a depreciation in the value of the collateral, or to automatically release some of the collateral tendered, in the event of its appreciation. Although theoretically practicable, the use of smart contracts for this purpose is contingent on the technology and processes of the parties to a security agreement being consistent with their participation in a shared platform, where smart contracts can be hosted and applied to the monitoring of fluctuations in the value of cryptocurrencies tendered and accepted as collateral.

Another way in which the parties to a security agreement may cater for the risk of volatility peculiar to cryptocurrencies is by making provision for fluctuations in the value of the cryptocurrency units tendered and accepted as collateral. Security agreements will typically specify the asset or property being held as collateral under the agreement, including its description by type, quantity, and, crucially, value. Absent any contrary provisions or doctrine under the law of contract governing the security agreement, the inclusion in a security agreement of a mechanism for the valuation of the digital asset or assets tendered and accepted as collateral, to cater for potential fluctuations in value, should not vitiate the legal effect and the enforceability of that agreement by rendering it ambiguous, vague, or indefinite.

6. CONCLUSION

Applied to the world of finance, digital innovation holds several promises. These include creating entirely new, investable asset classes, free of the costs, delays, and complications that surround the issuance of traditional assets, such as securities and bonds, and the trading of such traditional assets, especially across borders; facilitating the transfer of ownership in digital assets without the need for intermediaries or a "paper trail", and with immediate finality (at least in operational terms); and simplifying the issuance of, and trading in, tokenized versions of conventional assets classes, including those subsumed under the term "securities".

One of the factors that would help make digital assets even more attractive is the possibility for their holders to use them as security in their borrowing operations. The use of digital assets as security for their holders' borrowing obligations would bring with it several benefits. To start with, it could allow digital asset holders to monetize their holdings without having to divest themselves thereof (thereby not foregoing the benefits of their future appreciation).[13] Moreover, it could help to ease liquidity conditions in the market by allowing market actors at both ends of a prospective lending agreement to tap into a substantial, but unutilized, depository of good collateral. That said, the use of digital assets as collateral would also come with certain challenges. As explained in this article, these would affect both the creation of security interests in digital assets and their realization. Until those challenges have been overcome, in some cases with the benefit of legislative intervention, digital assets are unlikely to represent a source of collateral that many debtors and creditors alike will be willing to draw on for their routine business dealings, not because digital assets lack value but, rather, because market actors lack the tools necessary to "unlock" that value by, inter alia, tendering and accepting them as collateral.[14]

[13] An investor's ability to monetize their digital asset investments, without having to divest themselves thereof, is also likely to provide an additional incentive for them to invest further, increasing demand for digital assets.

[14] At the time of writing, it was mostly specialized venues and platforms, such as SALT, Nexo, and Abra, that extended loans to borrowers against their cryptocurrency holdings (mostly, bitcoin and ethereum) as collateral (on condition that borrowers transfer their cryptocurrencies to custodian wallets). The interest rate on loans is calculated depending on the loan term (the shorter the loan, the lower the interest rate) and the loan-to-value ratio (the more the collateral, the lower the interest rate).

CENTRAL BANK DIGITAL CURRENCIES AND PAYMENTS: A REVIEW OF DOMESTIC AND INTERNATIONAL IMPLICATIONS

LILAS DEMMOU | Deputy Head of the Structural Policy Analysis Division at the OECD Economics Department, OECD

QUENTIN SAGOT | Junior Advisor, Centre for Tax Policy and Administration, OECD[1]

ABSTRACT

Recent technological developments linked to secure messaging and traceability present an opportunity to address certain challenges in international and domestic payment systems. From an international perspective, foreign exchange markets remain costly and relatively less efficient than domestic payment systems. From a domestic perspective, the decline in the relative importance of cash in most economies reflects changes in consumers' preferences, which questions the future of money and payment infrastructure. Against that background, private initiatives falling outside of current regulation, such as stablecoins and other virtual assets, are associated with several risks and opportunities and have fueled the debate on the merit for central banks to issue new form of digital public currency. This article reviews these different propositions and examines their implications for the international and domestic payment systems.

1. INTRODUCTION

The development of financial market infrastructure is inherently linked to technological innovation and has evolved in the second part of the 20th century in response to an increasing integration of actors across borders at an ever-lower cost. Electronic money gained momentum from the 1970s allowing vast amounts of money to be transferred first between financial institutions and then to a larger set of actors. Those developments have played a key role in supporting trade and economic activity. Yet, in the face of recent technological advances, the existing settlement system is still considered slow and costly and the demand for new kinds of medium of exchange, notably for digital currencies or tokens, has increased, reflecting the emergence of new needs. The growing digitalization of retail trade has fueled this demand even further. While recent private and public initiatives aim at responding to those new needs, new challenges emerge for policymakers.

This paper takes stock of these developments and puts forward some economic implications on payment markets. First, ten years on from the worldwide emergence of a new type of privately-owned and decentralized digital financial asset, of which bitcoin was the first and currently the most well-known example, their potential economic impact is hugely debated. In November 2021, the total market capitalization of cryptocurrencies amounted to U.S.$2,973 billion, from U.S.$140 billion in March 2020, when COVID-19 hit. Yet, this market remains largely volatile, and costs of production inherently limit their use as a medium of exchange and reserve of value [FSB (2018a)]. Crypto assets' characteristics

[1] The authors would like to thank for helpful comments, valuable discussions, and insightful suggestions Laurence Boone, Luiz De Mello, Alain de Serres, Dennis Dlugosch, Guido Franco, Filippo Gori, Giuseppe Nicoletti (all from the OECD Economics Department), Caroline Malcom, Robert Patalano, Sebastien Schich, Ania Thiemann (all from the OECD Directorate for Financial and Enterprise Affairs), and Benoit Coeuré (from the Bank of International Settlements).

indeed make them a weak substitute to fiat currencies, while the underlying technology of these assets may not be flexible enough to ensure an adjustment of money supply to economic conditions.

However, further innovations in the crypto-economic world present the potential to change this global picture, in particular the development of stablecoins, i.e., crypto assets featuring a stabilization mechanism allowing them to anchor their price to a basket of stable fiat currencies or assets. While rather small in terms of market capitalization (circa U.S.$180 billion),[2] scaling projects, notably Facebook's Diem, have the potential to disrupt the current monetary system based on national fiat currencies and pose several economic risks. Firstly, regulatory settings on crypto assets and stablecoins, established as speculation instruments, may not abide by payment service providers (PSP) standards and thus may not guarantee users similar operational security and system resilience. In particular, private actors present higher credit risk (or probability of default) than central agents [Sveriges Riksbank (2018)] and even if solvent, private entities face an inherent liquidity risk associated with their business cycles. Competition issues add to the problem, as tech giants could leverage their dominant positions on international commerce by concentrating the operations of the marketplace on their own platforms, from advertising, to payments, and potentially lending [OECD (2020a)]. Such concentration could also challenge the stability of the payment system as the more concentrated a payment market, the greater the risk of contagion in the system. These risks have fueled the public debate on the necessity to regulate private currencies[3] (an issue not discussed in the present paper, but which notably affected the launch of Facebook's Diem initiative, shut down by regulators) and on the opportunities for central banks to issue new forms of digital public currencies (CBDCs).[4] In 2019, 80 percent of world central banks, surveyed by the BIS, had declared pursuing work in the area of CBDCs, though only a few engaged in the active development of pilots [Rice et al. (2020)]. Against this background, this article takes an exploratory perspective to examine the potential impacts of different CBDC designs on three areas: i) cross-border and domestic payment systems, ii) the role of the banking system; and iii) the efficiency of monetary policy toolkits. Country-specific experiences are also reported given that the motivations for expanding CBDCs may vary across countries, as do pilots' implementation level.

2. THE OPPORTUNITY OF CBDCs TO ADDRESS INTERNATIONAL PAYMENT SYSTEM CHALLENGES

2.1 Brief overview of international payments: a costly and slow payment system that may act as a barrier to trade and growth

An efficient cross-border payment infrastructure, enabling fast and affordable payments, is paramount to support international trade. Indeed, transaction costs appear as an important cost component in international trade for goods and services, amounting to roughly a fifth of total costs [Rubínová and Sebti (2021)] (Figure 1). The recent worldwide surge in e-commerce, fostering business-to-person sales as well as the significant increase in the volume of remittances, exacerbate further the need for cost-efficient cross-border transactions. Against that background, the current cross-border payment system is deemed to be slow, costly, and opaque, when compared to domestic payment systems.

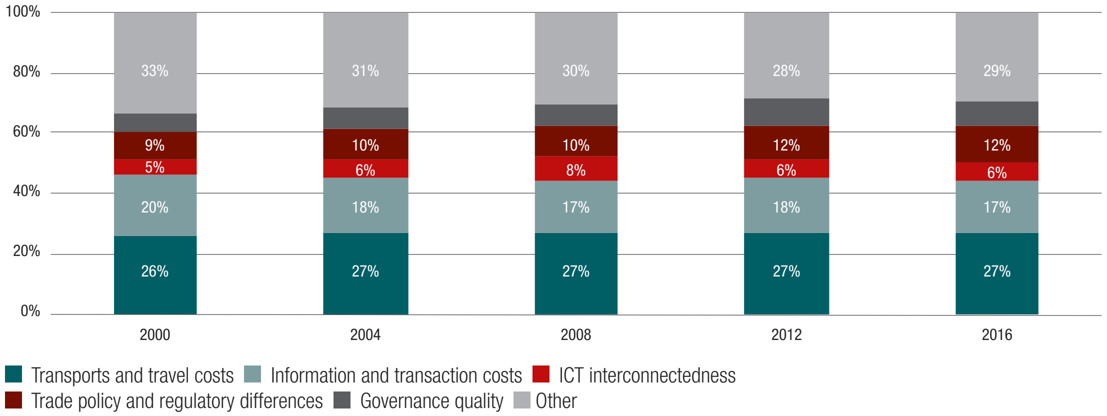
First, the international payment infrastructure is largely dominated by a few large players constituting the so-called correspondent banking system: payment service providers (PSP) and international banks having a presence in several countries – or correspondents – settle international claims on their own accounts across borders. Correspondents totalize roughly 90 percent of cross-border payment volumes, the remaining 10 percent being covered by the marginal presence of money transfer operators (MTOs – e.g., Western Union). Further, the FSB indicates that 45 percent of surveyed banks rely on two or fewer correspondents for more than 75 percent of the value of their wire transfer. This concentration around correspondent banks is even higher for small and medium banks. Such market power can have potential negative impacts on costs and efficiency, especially for smaller banks more vulnerable to abuses from dominant positions [FSB (2018b)].

[2] Figures extracted from the website: https://coinmarketcap.com/view/stablecoin.

[3] To this date, after the initial development of virtual assets outside of established regulatory framework, the G7 and the G20 have called upon a coordinated research and collective regulatory effort on these issues and their links with payments. The G7 has mandated the Financial Stability Board (FSB) to frame regulatory aspects of stablecoins. The G20 has mandated the BIS Committee on Payments and Market Infrastructure (CPMI) to investigate/identify policy options to address weaknesses in cross-border payments, considering CBDCs, among other options. The Financial Action Task Force focuses on regulating virtual asset service providers, in light of the standards of financial regulation regarding anti-money laundering and combatting the financing of terrorism (anti-money laundering/combatting the financing of terrorism – AML/CFT).

[4] Note that digital central bank money already exits under the form of commercial bank reserves deposited at the central bank.

Figure 1: Breakdown of international trade costs



Source: Rubínová and Sebti (2021)

Second, international payments remain costly compared to domestic payments: in particular, banks realize larger margins on international transactions, nearly 20 basis points (bp) against 2 bp for domestic transfers [McKinsey (2016)]. The impact is particularly large in "low and middle-income countries" (LMICs), where remittances have become the main source of external financing, surpassing FDI flows in 2018.[5] While the global goal for the cost of remittances has been established at 3 percent in the Sustainable Development Goals for 2030, the current global average stands at 7 percent.

Third, developments over the past few years exacerbated the risk of exclusion for LMICs from global markets: as reflected by the decline of correspondent relationships in many remote regions [Durner and Shetret (2015), Alwazir et al. (2017)]. A yearly analysis performed by the Committee on Payments and Market Infrastructures (CPMI) has shown that correspondent banking relationships have been in severe decline since 2012, as open correspondent corridors and active relationships have declined respectively by 10 percent and 20 percent from 2012 to 2018 (Figure 2) despite a relative surge in volume of cross border payments [Rice at al. (2020)]. Such reduction in service has been driven by several factors:

- The increase in the compliance burden has discouraged banks from managing less profitable correspondent relationships [BIS (2018)]. Specifically, the necessity to engage and manage AML/CFT measures, including the costly "know your customer" (KYC) procedure, has put pressure on the back-offices of correspondent banks, reducing the overall profitability of the relevant business line [Breslow et al. (2017)].
- Correspondent banks have generally adopted a lower risk profile to adapt to new post-GFC (global financial crisis) regulation (for which greater capital is required for riskier activities), notably discouraging them from engaging in jurisdictions where comprehensive due diligence of respondent banks (receiving the funds) could not be enforced [IMF (2017)].
- The degree of integration of information and communication technology fosters interconnections between international and local PSPs. Less integrated jurisdictions are, thus, suffering from reduced competitiveness, which often cause the reduction in active corridors [BIS (2020)].

Finally, the lack of interoperability between domestic payment systems makes cross-border payments slow and expensive compared to domestic payments: if not regulated under a harmonized payment area, such as the Single Euro Payment Area (SEPA), international payments rely on specific bilateral relationships, which are less efficient as they raise legal, regulatory, and technical issues [ITU (2016)]. The lack of interoperability is even more salient for "low and middle-income countries" (LMICs), which report

[5] Remittances flows in LMIC are evaluated at U.S.$462 bln, excluding China compared to U.S.$344 bln for FDI. Remittances are on track to become the largest source of external financing in developing countries [WBG (2019)].

interoperability of their automated teller machines (ATMs) and points of sale (POSs) between countries at roughly 50 percent, compared to 86 percent for high-income countries [WBG (2012)].

2.2 Recent public and private sectors initiatives have improved the current cross-border payment system

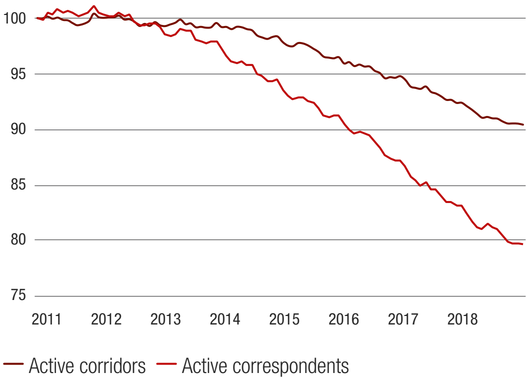
Recent innovations in exchange of information and digital repositories present the potential to raise the global efficiency of international payments by reducing the cost of cross-border transactions while increasing their speed of execution. A number of private and public initiatives have in particular emerged.

Firstly, correspondents are undertaking collective initiatives to lower the transaction costs of international trade: the pooling of customer regulatory information has been integrated into bank processes and should result in lower compliance costs (e.g., KYC depositories). Additionally, a sector-wide harmonization is being conducted by commercial banks and PSPs to establish global standards for payment messaging (ISO 20022 or SWIFT Global Payment Initiative), facilitating cross-border messaging while ensuring payment transparency.

Secondly, fintech is gaining momentum especially in Europe: where non-banks payment systems have experienced a rise of +529 percent in investments from 2013 investment levels [Bruno et al. (2019)]. They specifically tackle the retail segment by offering less costly and more rapid transaction services. This growth is mainly driven by two factors: the global expansion of online commerce and relative lower compliance costs, spurring from a more lenient regulatory regime, as most do not register as banks. Even if a complete substitution is not yet to be considered, this competitive pressure does, however, force prices down and foster operational innovation in the market.

Thirdly, central banks shoulder the responsibility for harmonization of cross-border payment infrastructures by improving the interoperability of national payments infrastructures: since 1999, the U.S. has been able to open its domestic system to cross-border payments by extending automated clearing house (ACH) services to foreign banks. These initiatives are, however, not widespread, notably because they feature lower margins for participating banks, making them less attractive overall. As an example, since it joined in 2003, Mexico has processed U.S.$2.6 billion worth of cross-border transactions [BIS (2018)]. Central banks are also increasingly researching "distributed ledger technology" (DLT)-interoperability to link real-time gross settlement (RTGS) systems across the globe (see Box 1).

Figure 2: Cumulative decline in correspondent banking



Source: New correspondent banking data – the decline continues [BIS (2018)]

Note: In the context of international money transfers, bilateral relationships exist between countries, in the form of active corridors, generally operated by SWIFT or between banks, forming active correspondents.

Finally, international standard-setting agencies are actively researching common measures to address the above-mentioned frictions: noted advances under the G20 mandate of the FSB to address the continuing decline of active corridors have focused on harmonizing regulations in national jurisdictions. Empirical research has recently shed light on the causes of de-risking, providing evidence that the loss of a corridor was related to Financial Action Task Force (FATF) country high-risk profiles, as well as their level of technological integration [Rice et al. (2020)].

2.3 Stablecoins could potentially increase efficiency of cross border payments, but their wide adoption would come with several risks

Created in 2014, stablecoins are crypto assets aimed at operating payments on distributed ledger technologies, allowing for peer-to-peer transactions. They are designed to address the most salient setback of crypto assets, namely their price volatility that prevents them from being used as a stable medium of exchange and unit of account, two fundamental

BOX 1. PUBLIC INITIATIVE OF DLT-BASED INTERNATIONAL SETTLEMENT SYSTEM

The Bank of Japan and the ECB's project ["Project Stella" – ECB, BoJ, (2019)] aims to leverage interoperability of "distributed ledger technology" (DLT) in different currencies. Such a system would rely on pre-funded deposit accounts, conditional payments, and guaranty lines. Just as for domestic payments, a central ledger would be operated based on these pre-funded accounts and exchanges would be performed, as well as recorded, irrevocably on the ledger. However, the project does not elaborate on the creation of a dedicated token.

The Monetary Authority of Singapore (MAS), jointly with the Bank of Canada and the Bank of England, has developed several models to establish a framework to use DLT in cross-border payments ["Project Ubin" – BoC, BoE, MAS (2018), Shabsigh et al. (2020)]. These models focus on the interoperability of decentralized ledgers to allow CBDCs to be used for cross-border payments:

- In the first model, central banks would issue CBDCs against their local currency on specific accounts opened by private entities – probably correspondent banks. The latter would hold accounts in multiple central banks to satisfy consumer demand in a varied range of currencies. This approach would provide a good technical solution to reduce both operating costs and settlement time as transactions would be performed within a single decentralized ledger. However, it would not significantly reduce the transaction costs associated with the system of correspondent banks, which is mainly driven by the regulatory reserves required when dealing with high-risk countries (as per Basel regulation).
- A second model explores potential agreements between central banks to operate CBDC accounts accessible to any participating banks. Practically, a ledger would exist for each currency in the monetary agreement and banks could directly access a foreign currency, without relying on any system but the DLT network, thus speeding up the process and potentially reducing fees to be paid to multiple actors. The foreign exchange rate would be determined by fractional reserve of the participating central bank's currency.
- Finally, the last model envisages the creation of a universal international currency, similar to the model of the stablecoin (reviving the idea of Keynes' bancor), against which all currencies would be quoted. Central banks and banks alike would open accounts on the DLT-operated networks and would trade the currency in line with their clients' needs. Exchange rates would be determined by fractional reserves.

characteristics of currencies. To do so, the price of the coin is anchored to a pool of assets. Stablecoins may use different mechanisms to stabilize prices: backing their value on assets or on algorithms controlling the supply of new stablecoins to preserve the value of existing coins [FSB (2020)].

Though their adoption as a new means of payment has been so far limited, their characteristics give them the potential for a more widespread use. Stablecoins could potentially represent an alternative means of payment for international settlement, bypassing the current correspondent banking system. Practically, one buyer would be able to purchase goods and pay in stablecoins, which could in turn be exchanged for an equivalent amount of fiat currency reflecting the price of the currency relative to the basket of currencies of the stablecoin. This process may increase the efficiency of cross-border payments by reducing transaction costs.

The use of a global medium of exchange is not a new phenomenon and has been undertaken by several national currencies. However, the specificity of crypto assets lies in the fact that they can be used at the same time as a means of payment, competing with national currencies [Benigno et al. (2019)]. Such a configuration would impose drastic changes on the existing financial system if largely used, the consequences of which remain to be formally assessed by regulators, in particular with respect to exchange rate, monetary, and competition policies.

2.3.1 EXCHANGE RATE POLICIES – A POTENTIALLY LESS EFFICIENT MARKET CLEARING

With a crypto asset used in parallel with national currencies, the fiat FX market would potentially clear less efficiently due to the lag induced by the currency being exchanged for stablecoins instead of another fiat currency. Market

clearing would be suborned to the stabilization mechanism of the stablecoin based on an algorithm. The efficiency of the latter remains to be proven effective under minimal market depth and low liquidity in asset reserve markets. Central banks may hence face difficulties in implementing their exchange rate policy.

Another concern relates to the capacity of a private actor to maintain the desired level of the peg, as claimed [Bullmann et al. (2019)]. Similarly, to maintain fiat currency pegs, stablecoins need to balance their collateral (foreign exchange reserves) on a continuous basis, to stabilize the coin value. Stablecoins algorithms have not yet proven capable of maintaining the peg value. Without the insurance that the pegs could be maintained by liquidity injections, stablecoins would require a lender of last resort to secure trust in the coin value, as is the case for any fiat currency [Schich (2019)]. Yet, such facility comes at the cost of heavier regulation and dependency on a central agent, which intrinsically opposes the initial motive for the development of stablecoins.

2.3.2 MONETARY POLICY: A DILUTION OF THE MONETARY POLICY CHANNEL

A widespread adoption of stablecoins would immunize the economy from central bank intervention. In particular, high-inflation currencies could see their citizens shifting towards the stablecoin to pay domestically, thus creating a type of dollarization of their economies. The reduction in banks' deposits, turned into stablecoins, would, therefore, render monetary policy, based on the two-tier system, less efficient in accommodating exogenous shocks through the interest rate channel [(Edwards and Igal Magendzo (2003)]. In addition, economies featuring a partial integration of stablecoins in their payment systems would also suffer from any appreciation of the external currencies, causing output to contract on accounts of higher stablecoin-denominated costs. Experience from dollarized economies has shown that an appreciation of the U.S. dollar may cause up to 1.5 percent reduction in emerging markets outputs [BoE (2017)].

In the current international monetary system, a trilemma prevents the simultaneous pursuit of three policy goals: financial integration, fixed exchange rate, and independence of monetary policy (conceptualized in the Mundell-Fleming framework). In the event of an economic downturn, central banks tend to conduct expansionary monetary policies to pull down the interbank interest rate and foster investments. Under the current system, the decline in the relative interest rate would trigger capital outflows to more generous jurisdictions, bringing down the exchange rate, boosting imports, and fostering additional growth (through the exchange rate channel).

The adoption of stablecoins would constrain monetary policy, leading to a potential dilemma, a situation where countries are forced to adopt synchronous monetary policy, even in the event of free capital flows and flexible exchange rate [Benigno et al. (2019)]. This comes from the fact that stablecoins, by acting as a global currency and at the same time as a means of payment, may be used as a substitute at the local level. The risk of portfolio shift between different currencies through the global money would increase and imply that currencies would compete indirectly with the global alternative. Exchange rates would then have to remain constant to avoid a flight towards the global money. Furthermore, if exchange rates remain constant, interest rate parity, which is required when capital movements are free, implies that nominal interest rates should be equalized, and hence monetary policy in the trading countries should be synchronized. Such a synchronization has been adopted in certain regions, with some benefits (e.g., eurozone), yet if stablecoins were to be prevalent at the global scale, countries may find themselves forced to such synchronization. In practice, the international role the U.S. dollar already plays prevents some jurisdictions from conducting an independent monetary policy.

2.3.3 COMPETITION POLICIES: THE RISK OF DOMINANT MARKET POSITION ABUSES[6]

The entry of tech giants in the payment services market may reduce its contestability. These firms gain dominant position in international commerce by concentrating the operations of the marketplace on their own platforms, from advertising, to payments, and potentially lending. Indeed, tech firms' business models are based on Data analytics, Network externalities, and interwoven Activities (DNA), which allow them to benefit from network effects. Simply put, adding additional users to the network increases the value to each user, notably through accessibility to a wider variety of individuals. These positive returns to scale usually create large barriers to entry and introduce a "winner takes all" risk. Furthermore, tech firms collect and manage users' data with more efficiency than banking actors, due to the inherent benefits for users to transmit data to the platform. Lastly, a decline in the use of cash might further reduce the market contestability of payments; in the event of no alternative public option, consumers could be subject to an oligopolistic behavior from payment infrastructure providers. Against this background, a first challenge relates to the protection of consumers' data, while a second critical issue relates to the need for new regulatory measures to reduce the risks of potential anti-competitive practices from dominant tech giants.

3. THE OPPORTUNITY OF CBDCs TO ADDRESS LOCAL PAYMENT SYSTEMS CHALLENGES

3.1 Brief overview of domestic wholesale payment infrastructure system – efficient but liquidity requirements remain high

While largely recognized as efficient, wholesale payments have been associated until recently with a trade-off between settlement risk reduction and up-front liquidity requirement for banks. National payment infrastructures are multi-layered and involve a multiplicity of actors in a two-tier model. Exchanging goods and services against cash or deposit claims electronically is made possible by a network of participants, operating transfers on a daily basis. Commercial banks operate large-value payments (LVP), as they deal with larger corporate and financial clients. These payments could generate settlement risk,[7] i.e., the risk that a counterparty

Table 1: RTGS systems opening hours

OPERATING HOURS (LOCAL TIME)	AUSTRALIA	COLUMBIA	EUROZONE	NORWAY	U.K.	U.S.
Opening time	07:30	07:30	07:00	06:40	06:00	21:00 (ET) the previous calendar day
Close for customer transfers	16:30	20:00	17:00	No standard cut-off times	16:00	18:00 (ET)
Final close	18:30	20:00	18:15	16:30	16:30	18:30 (ET)

Source: Allsopp et al. (2008)

[6] A wider analysis of the regulatory challenges and policy options on the topic have been explored by the OECD and its Delegates within the Competition Committee in June 2019 and the Committee on Financial Markets, which summarizes its conclusions in a recent paper [OECD (2020)].

does not receive its payment, while having disbursed the related securities. As wholesale payments became larger, coping with this settlement risk became paramount.

The move from deferred net settlement (DNS)[8] systems to RTGS wholesale systems in the 1990s and the progressive adoption of fast payment systems for retail infrastructure since the 2000s [BIS (2016)] have reduced substantially the settlement risk associated with payments. RTGS systems are dedicated platforms operated by central agents, allowing the immediate execution of wholesale payments in central bank money. Such systems, like the European TARGET2 or the U.S. Fedwire Funds Service, execute real-time settlements in central bank money. Settlement risk is then reduced as reserve pre-funding ensure the availability of funds, while dealing in central bank money protects the transaction from the default of the operator, given central banks are virtually immune from default in their own currency. By 2016, roughly 80 percent of the world central banks had implemented some form of RTGS.

Yet, the large adoption of RTGS systems, and the associated lower credit risk, has come at the cost of higher liquidity needs for commercial banks [Banque de France (2019)]. Indeed, RTGS systems require individual accounts to present the available funds to settle the transaction. If the funds are insufficient, the transfer is not performed or the payer needs to drawdown a credit line, often collateralized.

Furthermore, the system is only operational during central banks' opening hours, as outlined in Table 1 for six RTGS systems, which reintroduces settlement risk in the system. The collateralization of intra-day liquidity provided by central banks indeed causes mispricing as central counterparties shift their liquidity drawdown towards the end of the day, to save costs [Pfister (2018)]. To cope with this development, some central banks actively research full availability in their RTGS systems. For instance, the European Central Bank (ECB) TIPS operates pre-funded accounts, which can perform settlements on a 24/7 basis, but those accounts are funded only during the opening hours of the ECB and do not feature netting mechanisms.[9]

[7] BIS Glossary: settlement risk pertains to “the risk that settlement in a funds or securities transfer system will not take place as expected.”

[8] DNS aggregated daily transactions to net opposing positions and reduce liquidity intensiveness of wholesales payments yet building up credit risk as open positions increased.

[9] Netting is the process of offsetting the value of opposing positions or payments due to be exchanged between several parties. It generates credit and liquidity risks during the time the position remains open.

3.2 New technologies have the potential to increase further the efficiency of wholesale payment, though the overall gains appear limited

Wholesale central bank digital currencies (CBDCs) would represent a design increment to central bank money, which could present opportunities to reduce further intermediary costs and liquidity needs associated with the current RTGS systems.[10] This type of CBDC would exclusively aim at facilitating the exchange between the central bank and its designated central counterparties (systemically important commercial banks having access to the central bank's balance sheet through reserve deposits), within the interbank market. The main evolution from the existing system would be the migration from a gross system with partial availability to a netting system, featuring complete availability. Enhancement to the current system could include a reduction in settlement risk, liquidity needs [Garrat (2016)], and intermediary costs [Bech et al. (2017)], as well as ensuring complete availability of the payment system.

However, the DLT efficiency remains to be proven efficient and scalable. The execution speed of current DLT-systems would not support large-value payments (LVPs), notably due to lags in the validation process. Furthermore, no project large enough has been realized to test for the significance of cost-effectiveness of DLT, despite some interesting proofs-of-concept (see Box 2).

3.3 A universal CBDC could answer the decline in the demand for physical cash, yet with some profound economic implications

3.3.1 DEMAND FOR MEANS OF PAYMENT AND STORE OF VALUE PROVES INCREASINGLY DIGITAL

The decline of physical cash as a means of payment to the benefit of electronic money is noticeable in several developed countries. From 2006 to 2016, the share of transactions paid by cash declined: depending on the computational method, the yearly average reduction ranges from 1.3 percent to 2.2 percent across 11 countries and is forecasted to decline

BOX 2. THE CANADIAN CADCOIN PROOF-OF-CONCEPT AND POLICY RESEARCH

In 2016, the Bank of Canada (BoC), jointly with Canada Payments and the R3 Consortium, developed a pilot for its own CBDC: the CADcoin.[11] Their goal was to achieve operational efficiency through the creation of a wholesale currency, notably aimed at reducing back-office costs for users and the liquidity needs associated with RTGS systems. Indeed, the Canadian RTGS system mobilizes an increasing amount of liquidity, with roughly a tenth of the Canadian GDP (U.S.$175 billion) exchanged daily in central bank money.

The CADcoin is a DLT-operated central bank money based on digital depository receipts (DDR) that act as a pre-funded central bank zero-interest bond sent to the receiving counterparty. Transactions are netted and settled throughout the day until a "cashing-out" phase, which updates banks' positions in the central bank accounts. In essence, the BoC allows central counterparties to credit a segregated account on its books, in exchange for DDR to be spent during the day. Furthermore, because the money is deposited at the central bank, in its own currency, the credit risk would remain virtually nil. Liquidity needs are then reduced as DDR allow for an instantaneous netting of commercial banks' transactions, supporting higher volumes of transactions.

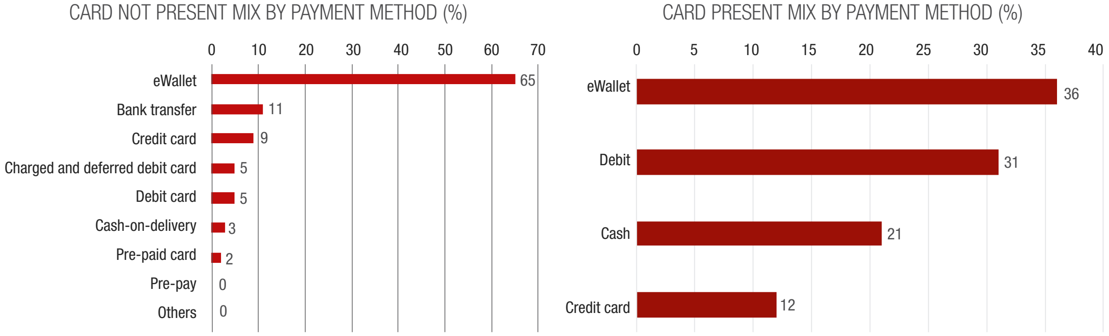
Overall, the project demonstrated successfully how DDR could be used to reflect existing securities markets on a digital ledger, featuring the issuance of securities from different actors and the existence of central bank issued cash to transact with. However, the BoC recognizes that the project's scope was not sufficiently large to detect any significant cost-saving opportunities related to the use of DLT.[12]

[10] Although not within the scope of this article, the authors recognize similar appetite for DLT existing in other capital markets, such as equity payments, to facilitate the settlement cycle or delivery versus payment [BIS (2020)].

[11] Several sources relay presentations and analysis of the project, notably Bank of Canada (2017), Chapman et al. (2017), and Bank of Canada (2018).

[12] Additionally, a legal framework is needed. In 2018, roughly 75 percent of central banks did not know or did not have the capacity to issue a new legal tender for a wholesale CBDC [Barontini and Holden (2019)].

Figure 3: Retail payment method mix in China (2017 figures)



Source: Worldpay (2018)

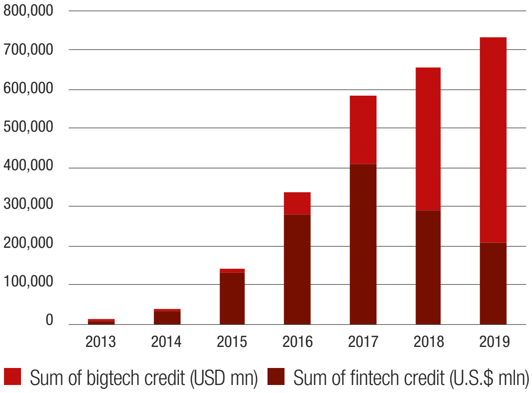
further at an annual average rate of 1.4 percent by 2026 [Khiaonarong and Humphrey (2019)]. Compositional changes in the population drive this trend, as younger adults use digital currencies (cards and mobile phones) for payments more often than physical ones.[13] Yet, cross-country differences in the use of cash remain large; the Germans pay for almost 70 percent of total transactions in cash, card, and e-money, compared to 10 percent for the Norwegians [Khiaonarong and Humphrey (2019)]. The general decline in the use of cash is associated with several opportunities and risks (Box 3).

3.3.2 AN INCREASING ROLE OF DIGITAL CASH PROVIDERS, ESPECIALLY IN EMERGING ECONOMIES

Accompanying the rise of online commerce, new payment systems have emerged and became widespread in some economies. These systems, more prevalent in developed economies, operate as overlay systems that relay the existing payment infrastructure (e.g., Paypal, Apple Pay). Alternatively, some platforms have developed a settlement system in-house, which features proprietary wallets [e.g., M-Pesa, AliPay, WeChat Pay – BIS (2019)]. While the former remains limited in use (Apple Pay in the U.S. only penetrates circa 7 percent of the population), presumably due to the established credit card infrastructure, the latter has experienced staggering growth in the past years. AliPay and Wechat Pay, respectively, account for 500 million (36 percent of the Chinese population) and 900 million users (65 percent), together realizing 94 percent of mobile payments in China. These new systems are now prevalent payments in China, with 36 percent of "card present" payments (based on the credit card infrastructure) and a staggering 65 percent of "card not-present" payments (Figure 3). Nonetheless, the People's Bank of China has been active in the development of a pilot CBDC, launching public digital wallets in four major cities to try to attract a share of the Chinese mobile payments.

These platforms generally operate as money market funds (MMFs), wherein they store and invest currency deposited in productive asset (generally repos or treasury bonds). Thus, they provide users with a store of value, alternative to banks' deposits. In China, these tech firms have grown to represent a significant part of traditional short-term funding, to such an extent that the Chinese government developed a dedicated clearing house to manage and secure these flows. A few of them actively engage in lending, however, this activity remains small relative to the global credit to private actors (less than 1 percent of total credit – see Figure 4 – [IMF (2019)].[14]

Figure 4: Global credit from tech firms (2013-2017)



Source: IMF (2019)

Note: 2019 fintech lending volume figures are estimated on AU, CN, EU, GB, NZ and US. 2 Data for 2019. 3 Domestic credit provided by the financial sector. Data for 2018.

[13] Here results for India could be counterintuitive as both birth and death rates stand high. According to "Beyond Cash" [USAID (2016)], these results might be due to the lack of penetration of digital infrastructure – "only 21% of these who earn digitally can save money in a bank account" – and the resulting low acceptance of digital means of payment by merchants.

[14] Big tech and the changing structure of payment services [BIS (2019)].

BOX 3. THE POTENTIAL RISKS AND OPPORTUNITIES ASSOCIATED WITH THE DECLINE OF CASH USE

Central banks bear the responsibility for maintaining the cash infrastructure of their given currency, which involves costs related to printing, designing, delivering, and replacing notes, among others. They earn, in turn, interest payments on the total of banknotes issued. Stronger efficiency gains related to the maintenance of a physical infrastructure are potentially associated with the digitalization of money. It is also associated with several potential gains related to a better traceability of payment, reducing the possibilities for tax evasion (in particular for VAT schemes) and other illicit financial flows.

One direct consequence of the decline in the use of cash is to lower seigniorage income (interest paid by banks in exchange for accessing central bank money) that can be quite substantial depending on the structure of the money demand. For instance, it ranged between U.S.$1-2 billion since the year 2000 at the Bank of Canada [Engert and Fung (2017)].

Furthermore, the decline of cash and a potential substitution towards crypto assets may be ultimately associated with financial risks. In the theoretical case of a cashless society, e-money and deposits would not be convertible into cash. The different forms of money would behave as financial assets, with their value against each other being continuously reassessed. The different forms of money would become an imperfect substitute and financial fragility could increase as the risk of runs from some forms of currencies emerges [Landau and Genais (2018)].

Finally, an effective decline in cash use would ultimately reduce the privacy of consumers' spending. As such, if effects of privacy on spending patterns remains debated [Acquisiti et al. (2017)], governments shall carefully define the means permitting the protection of consumer data.

Those risks related to the decline of cash are additional arguments feeding the debate about the opportunities to issue a CBDC in order to preserve demand for central bank liability and related seigniorage income.

3.3.3 A GENERALIZED ACCESS CBDC IS LIKELY TO DISRUPT THE FINANCIAL MARKETS, YET WITH POTENTIAL CONSIDERABLE BENEFITS IN THE CONDUCT OF MONETARY POLICY

3.3.3.1 Risks for financial stability in the deposit market and for economic growth

The provision of a risk-free option in the deposit markets is likely to increase the risk of bank runs from private actors unless protective dispositions are taken to counterbalance these effects. The threat of bank runs exists due to a lack of trust from consumers in a bank (or a group of banks) relative to their central liability (i.e., cash). By extending access to a risk-free bank liability (central bank money) through a CBDC, central banks would increase this threat. Different options would exist to dampen this risk, notably by designing restrictions or disincentives to portfolio shifts. First, promoting a financial safety net should preserve trust in the system in the event of a crisis. Among others, remaining lenders of last resort to immunize the economy from systemic risk losses, as well as protecting consumer accounts through deposit guaranty schemes, would be crucial for central banks. Second, central banks could also impose portfolio ceilings and dynamic transfer fees in order to curb portfolio movements, which could take the form of a volume fee, on the number of transactions, or a value fee, on the amount transferred [Mancini-Griffoli et al. (2018)][15].

The introduction of a public digital currency and the new deposit role of central banks would reduce financial market intermediation and potentially lower the profitability of the banking sector. The possibility for consumers to satisfy their demand for deposit via a risk-free asset is likely to reduce banks' main funding through deposits. In the current system, banks carry out transformations of short-term liquid deposits to long-term illiquid investments for individuals or firms. In addition, banks also have the capacity to create money, through seigniorage-financed lending[16] (even though this funding capacity is strictly regulated by central banks' reserve requirements, as a percentage of deposit held). With the creation of a public digital deposit, the central bank would reduce the amount of deposits available to banks and thus further deprive the banking sector of its primary funding mechanisms. Banks may then turn to commercial paper or equity for additional funding, yet these are likely to be more

[15] It should be noted that regulators should question the fairness of such a fee with regards to income inequality, not to disadvantage less endowed households.

[16] Commercial banks can create money through accounting by granting a loan and subsequently providing the funds in deposit accounts. Hence, the banks' balance sheet remains balanced and money, under the form of a deposit, can be expensed in the real economy. It is called seigniorage as in this case the banks' liabilities (deposit accounts) is used as currency.

costly and less stable, since the banks would retain the most junior share if a credit event occurs.

Introducing an interest-bearing CBDC may further deepen financial market disintermediation unless the supply of lending has the capacity to adapt. If an interest-bearing CBDC is introduced, the rate duly set by the central bank would constitute a floor to the market rate due to the risk-free characteristic of the central agent. This would influence the other actors in the deposit market; to remain competitive banks would need to increase their deposit rates vis-à-vis this risk-free option. This situation would shift up the supply curve faced by individual banks, as competition increases, and would bring about a subsequent reduction in banks' margins, especially if the price hike cannot be fully passed on to the lending rates [Chiu et al. (2019)]. Yet, if banks hold sufficient market power, it would be possible for them to pass on more of the additional costs to their lending rates, thus protecting their profit margins, and increasing their activities, by attracting more deposits [Mancini-Griffoli et al. (2018)]. However, as higher funding costs cascade into higher loan rates, potential adverse impact on economic growth may arise.

Depending on its design (see Box 4), account CBDC has hence the potential to weaken the overall prominence of commercial banks in the financial sector, to the benefit of

BOX 4: POSSIBLE DESIGNS FOR A CBDC[17]

The generalization of the access of a digital central bank liability to the wider population could rely on three distinct designs:

The first option would be to reduce the disruption to the current system by preserving the two-tier system, the banking business model and the existing form of cash: commercial banks could offer segregated accounts to consumers. Exchanges would then mimic current bank transfers and be operated by the existing organizational structures. Differences with the current system would lie in the legal arrangements pertaining to the rights of banks over this new form of money and the willingness of regulators to amend the current system. In this scenario, central banks offer an alternative public store of value, under the form of a protected account. Cash could, therefore, be preserved.

The second option would be to allow the public to hold accounts directly at the central bank, with potentially stronger effect to lower operating costs and settlement risks while still preserving cash: under this scenario, the central bank would provide a platform for exchange, immunizing the payment system from private actors' credit risk. Any payment performed on the platform would be irrevocable and guaranteed by the central bank. The need for intermediaries would then be reduced, as central banks would undertake a new role as payment system providers for individuals and non-financial firms. It would also need to manage individual deposit accounts in place of retail banks. In this case, overall operating costs could be reduced, as a unique central actor would perform all national transfers and thus would benefit from economies of scale. Importantly, central banks do not currently hold the adequate organizational setup to achieve these new functions, which constitute an important barrier for the adoption of this scenario. Cash could be preserved under this scenario, yet it would become less useful as most transactions could be performed under virtually frictionless platforms.

A token CBDC represents the furthest iteration to the current system, as it leverages DLT technology to substitute the existing payment infrastructure and the nature of cash: under this scenario, cash could be phased out completely and be replaced by a CBDC. Similar to wholesale CBDCs, all participants in the payment market (in this case, everyone) would hold a wallet from which exchanges would be performed. Each node would also participate in updating the version of the distributed ledger according to defined consensus mechanism. All tokens would then be created either by a transfer of cash or through the validation of this consensus mechanism. This scenario would thus preserve the peer-to-peer characteristic of cash, as DLT systems are based on the authentication and the validation of transactions by the decentralized network and do not require a central database gathering all information.

[17] The authors have selected only some of the design scenarios of a generalized CBDC. A more complete analysis can be found in Engert and Fung (2017).

central banks with two potential risks for economic growth, a reduction of the allocative efficiency of credit and a potentially negative impact on lending [Greenwood and Jovanovic (1990), Cetorelli and Gambera (1990)]. In order to reduce this risk of disintermediation, central banks could substitute retail deposits and lend directly to banks the money transferred to CBDC [Mancini-Griffoli et al. (2018)]. By disentangling the deposit and lending activities of banks, central banks would then secure the role of private actors to allocate credit while still reaping the benefits of a general-purpose CBDC. However, in such a scenario, central banks still need to devise a framework establishing rules of financing for banks, notably aimed at preserving central bank independence, which is crucial to guaranty the credibility of monetary policy [(Bordo and Siklos (2014)].

3.3.3.2 New tools for monetary policy and new risks for central banks

By controlling the rate of return on an interest-bearing CBDC, central banks could gain total control over the market rate, ultimately strengthening the monetary policy transmission channel. The difficulties met by central bankers to ensure the transmission of monetary policy to the real economy during the last (double dip) financial crisis has highlighted a weakness of our two-tier monetary order. Indeed, as the credit freeze occurred in Europe in 2010-2011, banks impeded the transmission mechanism of monetary policy through the interest rate channel and thus prevented the economy from adapting to the severe downturn. Those difficulties would be arguably stronger in a system where the share of privately operated money is larger. By contrast, an interest-bearing CBDC could bypass central counterparties and communicate rates to the market directly, thus allowing a complete transmission of monetary policy. Because central banks are the safest counterparty in their own currency, any rate they offer is virtually risk-free and thus constitutes the market floor. The rate would then be offered to all, and not limited to a single tier of central counterparties. In essence, CBDC holders would have an incentive to spend or to hoard depending on their expectations on the CBDC rate, thus smoothing potential output gaps.

An interest-bearing token CBDC could more specifically prevent economies from entering a “liquidity trap”, by alleviating the “zero-lower bound” (ZLB), which was hit by several advanced economies following the global financial crisis (GFC). This barrier to negative interest rates actually exists due to the presence of a zero-interest asset in the economy: cash.

Indeed, if the central bank can set negative rates, investors always have the option of holding cash, as a safe asset earning no interest. This possible "flight-to-safety" thus makes any increase in liquidity inefficient. The possibility of a flight-to-safety disappears if an interest-bearing CBDC supplants cash; central banks would gain immediate impact when applying a negative rate to boost currency circulation. Consequently, only an interest-bearing CBDC, with no remaining cash in the economy, would strengthen the efficiency of monetary policy. In contrast, implementing a non-interest-bearing CBDC (the closest to cash) would only have the effect of raising the lower bound from negative rates to zero [Sveriges Riksbank (2018)]. The current ZLB stands below zero (e.g., -0.4 for the eurozone) due to the relative burden of holding cash (e.g., cost of moving physical cash, insurance costs). In a digitized environment, there is no such physical slack. A negative policy rate would then always push investors towards holding CBDCs instead of central bank deposits, effectively raising the ZLB to zero.

Finally, an account-based,[18] interest-bearing, generalized CBDC would also provide a platform for Friedman's famous "helicopter money" [Engert and Fung (2017), Bordo and Levin (2017)]. As popularized by Bernanke (2002), this unorthodox monetary tool aims to combat risks of deflation in a low rate environment by increasing consumer demand, and thus welfare. This fiscal policy measure provides consumers with additional income, financed by newly printed money rather than by the monetization of existing assets, as traditionally undertaken in central bank operations. This emergency income handed over to consumers and businesses would be financed on the central bank balance sheets, rather than by national treasuries, through write-offs on the asset side or using the subsides of other monetary operations [Galì (2020)]. If, on the liability side, helicopter money is distributed under the form of a CBDC, central banks could then benefit from an additional option to overcome the ZLB and further strengthen monetary policy. Some argue, however, that this solution may prove less efficient than the current targeted monetization of government debt, with the latter remaining sovereign in the allocation of fiscal support [(Blanchard and Pisani-Ferry (2020)].

4. CONCLUSION

Global trends in international and domestic payments are driven by buoyant innovations that challenge established systems, both in the private and the public sphere. The digitalization of payment messaging and security has helped bring down some of the existing entry barriers, resulting in acknowledged portfolio shifts towards new virtual assets. We argue that these developments came about partly to cope with existing limitation in payments but also questioned policymakers on the collective actions needed and potential options to address such limitations. The international payment system features the most advanced proof-of-concept and focuses primarily on fostering the integration of emerging economies in global trade. On the domestic front, recent crises have shed light on opportunities to improve the conduct of monetary policy. Overall, central bank digital currencies remain a relatively new field in the economic and financial literature and many questions, notably on financial stability and privacy, remain. In this, it is likely that the numerous projects undertaken in central banks, intergovernmental organization, and academia will provide valuable insights in the years to come.

[18] Even though helicopter money could be programmable on a DLT, it may appear difficult to forecast its characteristic with any degree of precision, hence calling for a centralized provision of the CBDC to achieve this specific goal.

REFERENCES

Acquisiti, A., C. Taylor, and L. Wagman, 2017, "The economics of privacy," Federal Trade Commission, https://bit.ly/3svZz4V

Allsopp, P., B. Summers, and J. Vaele, 2008, "The future of real-time gross settlement: the role of the central bank," Bank of Canda, https://bit.ly/3ly5Rqt

Alwazir, J., F. Jamaludin, D. Lee, N. Sheridan, and P. Tumbarello, 2017, "Challenges in correspondent banking in the small states of the pacific," International Monetary Fund IMF working paper no. 2017-90, https://bit.ly/35DESeD

Andolfatto, D., 2018, "Assessing the impact of central bank digital on private banks," Federal Reserve Bank of Saint-Louis, https://bit.ly/3tn2vzY

Bank of Canada, 2017, "Project Jasper: a Canadian experiment with distributed ledger technology for domestic interbank payments settlement," https://bit.ly/3MbKZr4

Bank of Canada, 2018, "Project Jasper: phase 3 – securities settlement using distributed ledger technology," https://bit.ly/3hrHm27

Banque de France, 2019, "TARGET2, le RTGS de l'Eurosystème," https://bit.ly/3tiZCAv

Barontini, C., and H. Holden, 2019, "Proceeding with caution – a survey on central bank digital currency," Bank for International Settlements – Monetary and Economic Department BIS working paper no.101

Bech, M. L., and R. Garrat, 2017, "Central bank cryptocurrencies," BIS Quarterly Review, https://bit.ly/3lEXNV2

Bech, M. L., Y. Shimizu, and P. Wong, 2017, "The quest for speed in payments," Bank for International Settlements BIS Quarterly Review March 6, https://bit.ly/3Hrp3Vj

Benigno, P., L. Schilling, and H. Uhlig, 2019, "Cryptocurrencies, currency competition, and the impossible trinity," National Bureau of Economic Research NBER working paper no. 26214, https://bit.ly/3prwV3f

Bernanke, B., 2002, Speech before the National Economists Club, https://bit.ly/3Hz4eHH

BIS, 2016, "Fast payments – enhancing the speed and availability of retail payments," Committee on Payments and Market Infrastructures, CPMI Report November, https://bit.ly/3C0HZt0

BIS, 2018, "Cross border retail payments," Committee on Payments and Market Infrastructures February, https://bit.ly/3hv6CVa

BIS, 2019, "Big tech in finance: opportunities and risks," Bank for International Settlements, Section III/BIS Annual Economic Report 2019, https://bit.ly/3Hxl35J

BIS, 2020, "On the future of securities settlement," https://bit.ly/35Dr3ge

Blanchard, O., and J. Pisani-Ferry, 2020, "Monetisation: do not panic," CEPR, https://bit.ly/3MgsGBo

BoC, BoE, MAS, 2018, "Cross-Border interbank payments and settlements – emerging opportunities for digital transformation," https://bit.ly/3vpmiBH

BoE, 2017, "The global role of the US dollar and its consequences," Bank of England, December 15, https://bit.ly/3thrzZn

Bordo, M., and A. Levin, 2017, "Central bank digital currency and the future of monetary policy," National Bureau of Economic Research working paper no. 23711, https://bit.ly/3pocGTJ

Bordo, M., and P. Siklos, 2014, "Central bank credibility, reputation and inflation targeting in historical perspective," National Bureau of Economic Research working paper no. 20693, https://bit.ly/35d7TOs

Breslow S., M. Hagstroem, D. Mikkelsen, and K. Robu, 2017, "The new frontier in anti-money laundering," McKinsey & Co., https://mck.co/3K7Iy7i

Bruno, P., O. Denecker, and M. Niederkorn, 2019, "Global Payments Report 2019: amid sustained growth, accelerating challenges demand bold actions," McKinsey & Co., https://mck.co/3htlrYE

Bullmann, D., J. Klemm, and A. Pinna, 2019, "In search for stability in crypto-assets: are stablecoins the solution?", European Central Bank occasional paper series no. 230, August

Cetorelli, N., and M. Gambera, 1990, "Banking market structure, financial dependence and growth: international evidence from industry data," FRB of Chicago working paper no. 99-08, https://bit.ly/3pnY9aM

Chapman, J., R. Garratt, S. Hendry, A. McCormack, and W. McMahon, 2017, "Project Jasper: are distributed wholesale payment systems feasible yet?" Bank of Canada, Financial System Review, https://bit.ly/3Cby5oE

Chiu, J., M. Davoodalhosseini, J. H. Jiang, and Y. Zhu, 2019, "Central bank digital currency and banking," Bank of Canada staff working paper no. 2019-20, https://bit.ly/3lsgATf

Durner, T., and L. Shetret, 2015, "Understanding bank de-risking and its effect on financial inclusion – an exploratory study," Oxfam GB for Oxfam International, https://bit.ly/3ptXuod

ECB, 2013, "Annual Report 2012, https://bit.ly/3preuf1

ECB, BoJ, 2019, "Synchronised cross-border payments – Project STELLA," https://bit.ly/3K6VSbX

Edwards, S., and I. Igal Magendzo, 2003, "Dollarization and economic performance: what do we really know?" International Journal of Finance & Economics 8:4, 351-363

Engert, W., and B. Fung, 2017, "Central bank digital currency: motivations and implications," Bank of Canada staff discussion paper no. 2017-16, https://bit.ly/36TzUef

Frost, J., L. Gambacorta, Y. Huang, H. S. Shin, and P. Zbinden, 2019, "BigTech and the changing structure of financial intermediation," Bank of International Settlements – Monetary and Economic Department working paper no. 779, https://bit.ly/3vuuFfo

FSB, 2018a, "Crypto-asset markets: Potential channels for future financial stability implications," https://bit.ly/3lBw8nC

FSB, 2018b, "FSB correspondent banking data report – update," Financial Stability Board, https://bit.ly/3ssz3JL

FSB, 2020, "Regulation, supervision and oversight of "global stablecoin" arrangements," Financial Stability Board, https://bit.ly/3pu3xZU

Galì, J., 2020, "Helicopter money: The time is now," CEPR, https://bit.ly/3swnsJP

Garrat, R., 2016, "CADcoin vs. Fedcoin," R3 Consortium Report, https://bit.ly/3C1s3GS

Greenwood, J., and B. Jovanovic, 1990, "Financial development, growth, and the distribution of income," Journal of Political Economy 98:5, 1076-1107

He, D., R. B. Leckow, V. Haksar, T. M. Griffoli, N. Jenkinson, M. Kashima, T. Khiaonarong, C. Rochon, and H. Tourpe, 2017, "Fintech and financial services: initial considerations," International Monetary Fund IMF staff discussions note no. 17/05, https://bit.ly/3HKkbLr

IMF, 2017, "Recent trends in correspondent banking relationships – further considerations," IMF policy papers, https://bit.ly/3sulPfF

IMF, 2019, World Economic Outlook," International Monetary Fund, https://bit.ly/3pu1mpf

ITU, I., 2016, "Payment system interoperability and oversight: the international dimension," ITU Telecommunication Standardization Sector/Focus Group Digital Financial Services

Khiaonarong, T., and D. Humphrey, 2019, "Cash use across countries and the demand for central bank digital currency," IMF working paper no. 2019-46

Landau, J., and A. Genais, 2018, "Les crypto-monnaies: rapport au Ministre de l'Économie et des Finances," French Ministry of Economy and Finance, https://bit.ly/3C23l9x

Mancini-Griffoli, T. M., M. S. Martinez Peria, I. Agur, A. Ari, J. Kiff, A. Popescu, and C. Rochon, 2018, "Casting light on central bank digital currency," International Monetary Fund staff discussion note, https://bit.ly/3td2EpC

McKinsey, 2016, "Rethinking correspondent banking," https://mck.co/3the7Vf

O'Dwyer, K., and D. Malone, 2014, "Bitcoin and its energy footprint," National University of Ireland Maynooth, https://bit.ly/3vyj0Mq

OECD, 2020a, "Digital disruption in banking an its impact on competition," https://bit.ly/3K5lJB1

OECD, 2020b, "BigTech, financial intermediation and policy considerations," OECD

Pfister, C., 2018, "Real-time is money," Banque de France working papers no. WP 675, https://bit.ly/3HEvbtK

Rice, T., G. van Peter, and C. Boar, 2020, "On the global retreat of correspondent banks," BIS Quarterly Review, March, https://bit.ly/3lAkKsm

Rubínová, S., and M. Sebti, 2021, "The WTO trade cost index and its determinants," World Trade Organization staff working paper ERSD-2021-6, https://bit.ly/36ThjPu

Schich, S., 2019, "Is there a useful role for stable coins in a tokenised world," OECD DAF/CMF(2019)12

Shabsigh, G., T. Khiaonarong, and H. Leinonen, 2020, "Distributed ledger technology experiments in payments and settlements," Fintech note, IMF, https://bit.ly/3C7uGHs

Sveriges Riksbank, 2018, "The Riksbank's e-krona project – Report 2 – October," https://bit.ly/3HrXETb

USAID, 2016, "Beyond cash – why India loves cash and why that matters for financial inclusion," https://bit.ly/36LUflz

WBG, 2012, "World Bank's Global Payment Survey 2012," World Bank

WBG, 2019, "Record high remittances sent globally in 2018," World Bank Press Release 2019/148, https://bit.ly/35FzgQJ2018

Worldpay, 2018, "Global Payment Report – the art and science of global payments," https://bit.ly/3prLROu

WTO, 2018, "World Trade Report 2018: the future of world trade: how digital technologies are transforming global commerce," World Trade Organization, https://bit.ly/3HvlvS3

DECENTRALIZED FINANCE (DEFI) FROM THE USERS' PERSPECTIVE

UDO MILKAU | Digital Counselor[1]

ABSTRACT

Decentralized finance (DeFi) applications have been surging with incredible speed for about two years. Some DeFi enthusiasts aim to recreate financial services on the foundations of distributed ledger technology and smart contracts, i.e., computer scripts executed on a distributed runtime platform. This perspective has a clear focus on technology. To shift the debate, this paper examines DeFi from the perspective of users and their contractual relationship with DeFi. Given that DeFi removes traditional intermediaries, one needs to ask which entity becomes the counterparty? One fundamental element of contract law is the "meeting of the minds"; hence we need to determine who are the interacting minds in a DeFi agreement? A second fundamental question is about the beneficiary, or in other words: cui bono? Finally, it is important to determine whether DeFi in fact provides "financial services" or whether it is simply a gaming table, upon which different tokens move positions? The question of the applicable law has to be answered by regulators, nevertheless, the analysis in this paper reveals that DeFi exhibits a structure with "central" entities and a trend towards "gamification".

1. INTRODUCTION: FROM TECHNOLOGY TO HUMAN AGENTS

Decentralized finance (DeFi) is a new phenomenon that has grown rapidly since 2020. DeFi is also a new paradigm to reinvent financial services on the foundations of distributed ledger technology and smart contracts, i.e., computer scripts executed on distributed runtime platforms according to a code written by a programmer with an intention. Controversial opinions about the impact and the benefits of DeFi exist, which can be illustrated with three recent quotes:

- "DeFi offers exciting opportunities and has the potential to create a truly open, transparent, and immutable financial infrastructure" [Schär (2021)].
- "DeFi presents a panoply of opportunities. However, it also poses important risks and challenges for regulators, investors, and the financial markets" [Crenshaw (2021)].
- "There is a 'decentralization illusion' in DeFi since the need for governance makes some level of centralization inevitable and structural aspects of the system lead to a concentration of power. If DeFi were to become widespread, its vulnerabilities might undermine financial stability" [Aramonte et al (2021)].

Any analysis of DeFi that is based solely on technology is bound to be limited by implicit assumptions. Such a perspective puts technology and back-end automation at the center, assumes technological perfection, and ignores the imperfections of our real world, including errors, hacking, scams, and so-called software aging due to changing building block of technological stacks [Parnas (1994)]. The assumptions can be summarized as a mechanistic utopia.

Human agents are at the perimeter of this paradigm – especially as human beings are fallible. Nonetheless, we all know that everything we "write" – whether legal contracts or computer code – is always incomplete, as we have limited knowledge about the past and present and cannot forecast the future with all of its contingencies [see Coase (1937),

[1] I would like to thank Michael Jünemann for his valuable comments on the topic of this article.

Williamson (1979), Grossman and Hart (1986), and Aramonte et al. (2021)]. It is for this reason that this article will focus on the users' perspectives of DeFi, as opposed to its technological foundations.

2. APPROACHING DEFI: THREE EXAMPLES OF HOW TO ACCESS DEFI

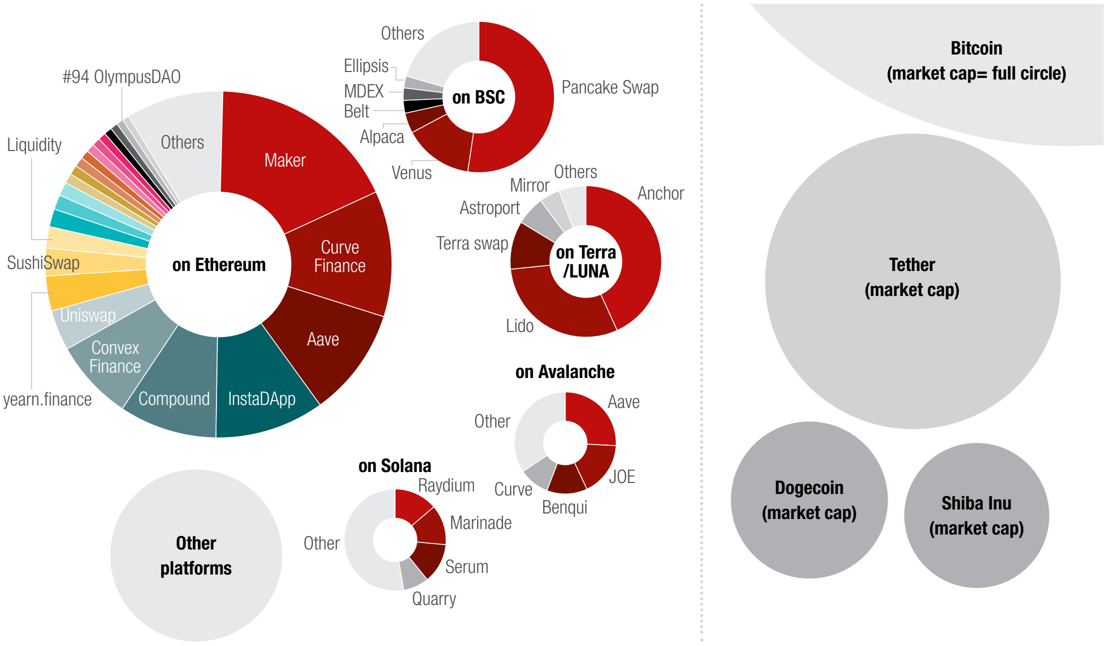
In this section, I will evaluate three examples of how DeFi could be accessed, namely Uniswap, Aave, and MakerDOA (Figure 1).

Let us follow a user who is looking for access to Uniswap (2021) and enters the internet page https://app.uniswap.org/ for the first time. The user is asked to "Connect a wallet – By connecting a wallet, you agree to Uniswap Labs' Terms of Service and acknowledge that you have read and understand the Uniswap Protocol Disclaimer" (page accessed December 8, 2021).

After acceptance by the user (and technical connection to a user's wallet), this **central app** provides all the necessary information and access to Uniswap as a "decentralized exchange". This type of access is characteristic of DeFi.

First, one can see a typical contract agreement with an "offer" made by Uniswap via the front-end app (as described in the "Terms of Service"), "acceptance" by the user (by clicking on a button), and an explanation about different "considerations" (the interface being free of charge, but there are fees such as "gas" in Ethereum for processing; see below), which makes this situation a distinctive "meeting of the minds" as stipulated by contract law. It should be noted that the "Terms of Service" contain a "Privacy" statement, which includes a **consent of the user** for the application of third-party data to "collect to detect, prevent, and mitigate financial crime and other illicit or harmful activities on the Interface."[2]

Figure 1: Snapshot of the DeFi universe



This figure compares total value locked (TVL) of different blockchain platforms with the market capitalization of major crypto coins (right side; all values assessed at December 23, 2021). The TVL, which is the sum of all assets deposited, is typically displayed in U.S.$., and the sum of all TVL was around U.S.$256bn, as compared to around U.S.$25bn at the beginning of 2021. TLV should be regarded as proxy, because data are taken from different sources and the deposited assets are tokens with high volatility. Few DeFi apps are implemented on various platforms (e.g., Aave and Curve).

Sources: defipulse.com; www.defistation.io; defillama.com; and coinmarketcap.com

[2] A detailed analysis of "know-your-customer" (KYC) and "anti-money-laundering" (AML) requirements, as well as data protection regulation and the issue of outsourcing, is beyond the scope of this paper.

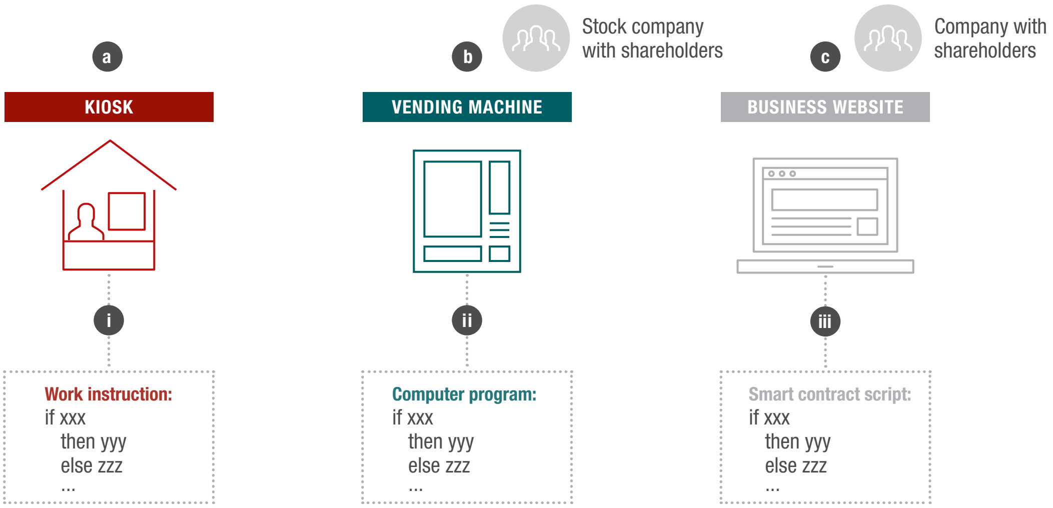
Second, there is a fragmentation of service provision across different layers:

- The user's wallet (on the user's device) is required to store the cryptographic keys, which enable access to the tokens recorded on the blockchain platform.
- The app, as general access to the service, was developed and is provided by Universal Navigation Inc. based in Delaware with the trademark "Uniswap Labs", and displays available combinations of tokens to be "swapped" (e.g., DAI versus ETH), prices for tokens, trading information, etc.
- The "Uniswap protocol" is the central algorithm of this "decentralized exchange". Technically, this is a software executed on a blockchain platform. According to the Uniswap disclaimer: "Although Universal Navigation Inc. d/b/a/ 'Uniswap Labs' ('Uniswap Labs') developed much of the initial code for the Uniswap protocol, it does not provide, own, or control the Uniswap protocol, which is run by smart contracts deployed on the Ethereum blockchain." The governance of Uniswap is based on holders of "governance tokens" (UNI Token) providing voting rights. As these UNI tokens can be traded on crypto exchanges, UNI represents a "share-like" concept with voting rights and participation in the value of the virtual company. While governance tokens of other DeFi applications, such as PancakeSwap's CAKE token, entitle the holder to earn a portion of the revenues, Uniswap distributes fees to so-called liquidity providers.
- The tokens/token pairs on Uniswap represent a huge token universe. Xia et al. (2021) state that they "identified over 10K scam tokens listed on Uniswap, which suggests that roughly 50 percent of the tokens listed on Uniswap are scam tokens. All the scam tokens and liquidity pools are created specialized for the "rug pull" scams, and some scam tokens have embedded tricks and backdoors in the smart contracts." In contrast to the most traded tokens, these scam tokens represent low/no liquidity tokens waiting for victims. The problem of scams will be discussed later in this paper.
- Liquidity providers can be compared to market makers in traditional security exchanges, as they provide tokens to pools for trading and receive rewards (0.3 percent of the value of trades). Whereas market makers typically offer a quote, Uniswap applies an "automated market making" based on a simple algorithm: a "constant product formula" $x * y = k$, where x and y are the amounts of tokens A and B in the pool [Aramonte et al., 2021]. Despite the differences in the mechanism of market making (quote versus algorithm), liquidity providers and market makers have comparable economic functions and incentives.
- The last layer is the actual processing, i.e., running the smart contract computer scripts on an execution platform. Any processing on the Ethereum platform requires fees to be paid (called "gas" and to be in the native ether "ETH" tokens), which are composed of a base fee and a "tip" (priority fee) since Aug. 2021. This is similar to the more proprietary Binance Smart Chain platform (BSC; linked to the unregulated Binance crypto exchange) or Terra/LUNA (see Figure 1). This can be compared with commercial cloud service providers offering "outsourcing" of processing.

The second example, Aave, has a similar structure. Aave is described on its website (aave.com) as a "liquidity protocol for earning interest on deposits and borrowing assets," which is an emulation of the core banking function to intermediate between savers and borrowers. The "general terms of use" [Aave (2020)] unveils five layers: users, the general interface, the decentralized protocol, the liquidity providers, and the Ethereum platform. Similar to Uniswap, the website is the access interface and is provided by a commercial company: "In these Aave General Terms of Use ('Terms'), 'Aave', 'we' and 'us' refers Aave SAGL and we own and operate the website https://aave.com/ ('the Site') which acts as a front-end to the decentralized Aave Protocol. ... As part of the Site, Aave provides access to a decentralized finance application ('Application') on the Ethereum blockchain, that allows lenders or borrowers of Ethereum assets ('Cryptocurrency assets') to transact using smart contracts ('Smart Contracts'). Using the Aave Protocol may require that you pay a fee, such as gas charges on the Ethereum network to perform a transaction." "Aave SAGL is a company incorporated in Switzerland, ...Privacy Policy ... gives you rights by operation of the EU GDPR. ... Your agreement with Aave's Terms of Service constitutes your consent to the collection and use of Personal Information as described in this Privacy Policy."

The AAVE token is used for a governance with voting rights and the possibility to "receive incentives" [Aave (2021)]. Furthermore, Aave was the first DeFi app that introduced so-called "flash loans", which are "repos without collateral" within

Figure 2: Three situations of "meeting of the minds" at a kiosk, a vending machine, and an online business website



The layer of interaction (i.e., the "contract") is shown in the upper row, while the lower row shows the "technical" processing in the background, which was ex-ante defined by a programmer.

one block in the blockchain. These transactions allow for a temporary creation of "value from nothing" due to bookkeeping on a blockchain in batches – against fees to be paid, because there is no free lunch, and every transaction in DeFi has a beneficiary.

The third example is MakerDAO, which has a longer history. MakerDAO was created by Rune Christensen in 2014, and the core product is DAI, a "decentralized stablecoin" based on overcollateralization with other crypto assets [MakerDAO (2020)]. This resembles a repo agreement in traditional money market operations with one significant difference, as tokens without intrinsic value are collateralized by other tokens without intrinsic value. The DAI token was launched at the end of 2017, a "Maker Foundation" took over control in 2018, and the governance was shifted in 2021 to the holders of the governance token "MKR".

The front-end was separated as Oasis app to: "Borrow Dai and Multiply your exposure to crypto Open a Maker Vault, deposit 25+ crypto collaterals. Either borrow Dai or buy additional collateral to increase your exposure. Connect a wallet to start." Oasis is now operated by a company incorporated and registered in England according to the "Terms of Service": "Please read these Terms of Service (this 'Agreement') carefully. Your use or access of the Site or the Services (as defined below) constitutes your consent to this Agreement." "This Agreement is between you (the 'User' and collectively with others using the Site, 'Users') and Oazo Apps Limited, a company incorporated and registered in England, United Kingdom ('Company' or 'we,' 'our' or 'us' and together with you, the 'Parties') concerning your use of (including any access to) Company's websites, currently located at oasis.app." "By clicking or tapping any button or box marked 'accept' or 'agree' (or a similar term) in connection with this Agreement, or by accessing or using the Site or the Services (as defined below), you agree to be bound by this Agreement."

This DeFi application can also be described as a sequence of a user/wallet, a front-end app that is operated by a commercial entity, a protocol with a token-holder governance (voting and incentives), and processing on a blockchain platform.

3. CONTRACTS: THE MEETING OF THE MINDS

These examples reveal that the user of DeFi is typically interacting with a front-end app provided by a commercial firm. The user enters into a contract by accepting an offer – typically with a click on a button and accepting the Terms of Services. This matches the traditional perspective of a "meeting of the minds" as basis for a contract. The technical computing "behind the curtain" – whether based on smart contract computer scripts or traditional programming – can provide documentation and processing, but the formal agreements are made between the user and a firm. This supports the notion of a "decentralization illusion" in DeFi [Aramonte et al. (2021)].

Some scholars argued that code itself could establishes a "lex cryptographica" and computer programs could be "self-contained and autonomous". According to De Filippi and Wright (2018), "blockchain technology facilitates the emergence of new self-contained and autonomous systems that rely on lex cryptographica. These systems enable people to communicate, organize, and exchange value on a peer-to-peer basis, with less of a need for intermediary operators."

One could ask, what "self-contained and autonomous" means in the context of technological systems, which, for the time being, have to be programmed ex-ante and have no free will to act independently from the original intention of the programmer. It helps to start with a comparison of three stylized situations when someone enters into a contract to buy a bottle of lemonade or a lottery ticket (also a "token").

Figure 2 illustrates three situations: 1. at a kiosk, 2. at a vending machine, and 3. via a website. The first situation demonstrates the principle that a contract (between the consumer and the kiosk merchant with a published price list) in a written format is not required. In addition, the "at a kiosk" example also highlights the fact that the "internal processing" in the kiosk prescribed in a "manual" is neither the offer nor the contract, but simply "execution", which could be "automatic" if the written manual is translated into a computer program and performed by a robot.

The next step, situation 2, tries to determine whether deals undertaken with a technical device, such as a vending machine, are considered legal contracts, from a contract law perspective, and consequently binding. To cut the long story of many decades short, a vending machine can display an offer as the first step of a legal contract (to be accepted by a consumer) but "on behalf" of a legal entity – i.e., natural person or legal company – which installs and operates the machine, including all internal programming. This legal entity is a beneficiary of the contractual agreement and would be liable for any breach of contract (keeping in mind that contracts have a remedial function in cases of dispute).

The last step is the "digitalization" of a physical vending machine as an online business. The actual difference is marginal, because the interaction takes place with a "technical" representation of a legal entity: either with buttons on the machine with advertised prices (on behalf of a company) or with buttons on a website with advertised prices (on behalf of a company).

This perspective was applied in 2012, when the German Federal Court of Justice (Bundesgerichtshof) ruled that the way an automated system is expected to process and execute a declaration of intent, which was made using electronic means of communication and via an automated booking or ordering system, does not determine the content of the contract. What matters is how the human addressee is allowed to understand the respective declaration in good faith and custom [BGH (2012)].

The displayed ("offer") and confirmed ("acceptance") content is binding like a contract, whereas the execution of bits and bytes within a computer system is technically relevant for the performance but can be seen as a black box behind the curtain (Figure 2 at the bottom).

Even if we make a Gedankenexperiment ("thought experiment") and assume that the potential buyer would operate directly on the blockchain, i.e., with the ability to parse published smart contract code, there has to be an offerer, who creates and publishes a smart contract like a declaration with a price list. The user could accept this offer by signing the smart contract (with a cryptographic key), and the offerer can confirm with a second signature. In the model of Figure 2, a smart contract is a declaration ("offer") that has to be agreed to ("acceptance").

The reality, however, diverges from this Gedankenexperiment, mainly because all examples of DeFi application implement the model with a front-end interface and a back-end processing, and every smart contract on the blockchain is published in so-called "bytecode", which resembles assembler computer language of yesteryears. Only experts, and not your average consumer, can read, translate, and understand such bytecode.

Of course, expert witnesses could be asked to "interpret" the original computer script in court – and even they could overlook errors or backdoors – but the consumer interacts with the "display".

Only the "meeting of the minds" establishes a contract and not any "mechanisms" inside the vending machine. The agreement to a contract does not usually need to be in a specific form, and "protocols" can range from a simple handshake (horse-trading) to signed paper-based documents to electronic messages with digital signatures. Taken together, DeFi can be structured in the interaction part ("meeting of the minds"), a part of the "written" documentation (with our limited strength of knowledge about the future) and the performance, which can be automated technically (with traditional computer programs or smart contract scripts).

4. AUTOMATION, GOVERNANCE, AND CUI BONO

The promise of DeFi is a new financial system that operates without the involvement of either centralized entities, such as central banks or exchanges, or intermediaries, like banks or asset managers. These financial institutions would be replaced with a system of automates, i.e., smart contract scripts on blockchains. However, in reality, users will enter into contractual relationship with some "frond-end" providers, which resemble traditional brokers or financial services providers.

These access providers develop and operate the interfaces and transmit the transactions to some DeFi protocols. These protocols perform similar functions to those performed by traditional exchanges, money market funds, or bank lenders, if we apply the hypothesis that the tokens are a type of financial instrument and not mere play money (see further discussion below).

However, we already have "algorithmically managed" financial services, with the best examples being exchange traded funds (ETFs), which "passively" track an external index, and algorithmic high-frequency trading with a "programmed" trading strategy. These cases demonstrate the essential difference between operational management (performing functions) and governance (making decisions).

There is also a difference between "automated" and "autonomous". While a formula, an algorithm, or a trading strategy can be written down – as a "manual" on paper or as a "programmed" software – and executed "automatically" without human intervention, we all have limited knowledge about the future. This incompleteness of knowledge, programs, and contracts require us to make decisions about changes to adapt to the future. With free will and without external force human beings can make these decisions "autonomously".[3]

While operational management can – under well-defined conditions – be executed automatically, all actual decision-making requires a human governance, be it by an individual owner, a member of a cooperative organization, by shareholders of a stock company, or by "governance token" holders of a DeFi business. This governance includes a defined process to exercise voting rights (due to the ownership share) and can include rights for dividends or other distributions.

The U.S. state of Wyoming amended its legal definition of a limited liability company in mid-2021 by issuing the "Wyoming decentralized autonomous organization supplement" [Wyoming (2021)]. This supplement allows a limited liability company to be "algorithmically managed" by a "smart contract on the blockchain" but will require a membership governance and a registered agent (as point of contact). Contrary to the semantics, such "decentralized autonomous organizations" (DOAs) will be neither "decentralized" nor "autonomous" because they are tangible companies (LLCs) and require a governance by owners/members/token-holders.

In recent years, there has been a shift from initial coin offerings (ICOs) to the current DeFi governance tokens. Independent of the terminology, the fundamental question is always who controls the business and receives the benefit? In other words: cui bono? There is always an initiator, an "ideator", or at least somebody who sells an idea to some investors. The SEC (2017) report about decentralized autonomous organizations is a comprehensive benchmark for economic purposes and attempts to obfuscate them. No "smart contract" is a divine contract ex machina – there has to be an economic agent from the start with commercial objectives and ongoing incentives.

[3] Just as a remark, and not relevant to this paper, there is a third "a", namely "autarchic", which means independent of environmental impact.

5. WHERE IS "THE" BLOCKCHAIN?

The remaining question about the technological foundations of DeFi is: where is the blockchain? This paper will not delve into the reasons why there is not "the" blockchain, but different platforms that are either updated by "minors" (proof-of-work consensus), which have a natural tendency to consolidate (due to the generic game-theoretical economic incentives), or "validators" (proof-of-stake consensus), who are a manageable group by definition. In reality, few miners/validators dominate blockchains, while "normal network nodes" provide connectivity and archiving only. DeFi running on these blockchains is processed by a redundant (aka distributed) network infrastructure, but the functionality is not "decentralized" but performed by an algorithm with a centralized governance.

This operating model is similar to "cloud" computing, with a runtime platform provided for some fees, except for the "readable" bytecode, i.e., the published code of all smart contracts. The blockchain platforms – whether Ethereum or some newer platform like Solana or Avalanche – have the generic disadvantage that every vulnerability is visible to every hacker, but not to the average users. According to reports [Elliptic (2021)], DeFi crime increased to $10.5 billion in 2021 (January to November 2021), up from U.S.$1.5 billion in 2020.

There is only one exception, and that is the original concept of the Bitcoin blockchain, in which all participants run network notes in a peer-to-peer network without any hierarchy or intermediaries. The Bitcoin blockchain with its proof-of-work consensus mechanism was a game-theoretical solution to "emulate" electronic cash in a repeated game, in which the rules and incentives reward cooperation and penalize manipulation [Auer et al. (2021)]. This concept has the implicit assumptions of a closed game and equal chances for everybody. The actual difference in expensive computer resources is a starting point for concentration, and a few so-called mining pools dominate the economy of Bitcoin today [Makarov and Schoar (2021)]. According to BTC (2021) and Etherscan (2021), by December 18, 2021, Bitcoin was effectively "controlled" by four mining pools, as counted by generated blocks, (60 percent in four pools: AntPool, Foundry USA, F2Pool, and an unknown address), and Ethereum by only three mining pools (57 percent in three pools: Ethermine, F2Pool, and Hiveon Pool).

This natural centralization allows sophisticated strategies to achieve "miner extractable value" (MEV) [Shin (2021)] due to the opportunity to include, exclude, or re-order transactions within the blocks by the mining pools. This ranges from priority fees (so-called "tips" in Ethereum for faster transaction processing) to the use of information asymmetry with own "front-running" token deals (by so-called "searchers"). An excellent summary of the economy of the Bitcoin blockchain was provided by Auer (2019), and an overview of the fee problem was presented by Kreitmar (2021).

6. GAMES – GAMBLING – SCAMS

Parallel to centralization and commercialization of blockchain platforms, one can observe another current trend, which is "gamification". A recent example is quite representative of the current environment but requires explanation.

According to the website olympusdao.finance, OlympusDAO (2021) claims to be "The Decentralized Reserve Currency – Olympus is building a community-owned decentralized financial infrastructure to bring more stability and transparency for the world."

Neither OlympusDAO nor the OHM token possess any of the features of money (medium of exchange, unit of account, store of value) and they are, of course, not currency according to the definition of a governmental framework for money. The website, which offers (sic!) trading of OHM tokens, has no imprint, company information, venue, etc., at all. Some media refer to a pseudonym "Zeus" as the founder, and a white paper on the website is signed by "nf.carlo.acutis" as a pseudonym that can be found as the copyright of another firm "Intrinsic Research Co.: Macro Fundamentals" on substack.com.

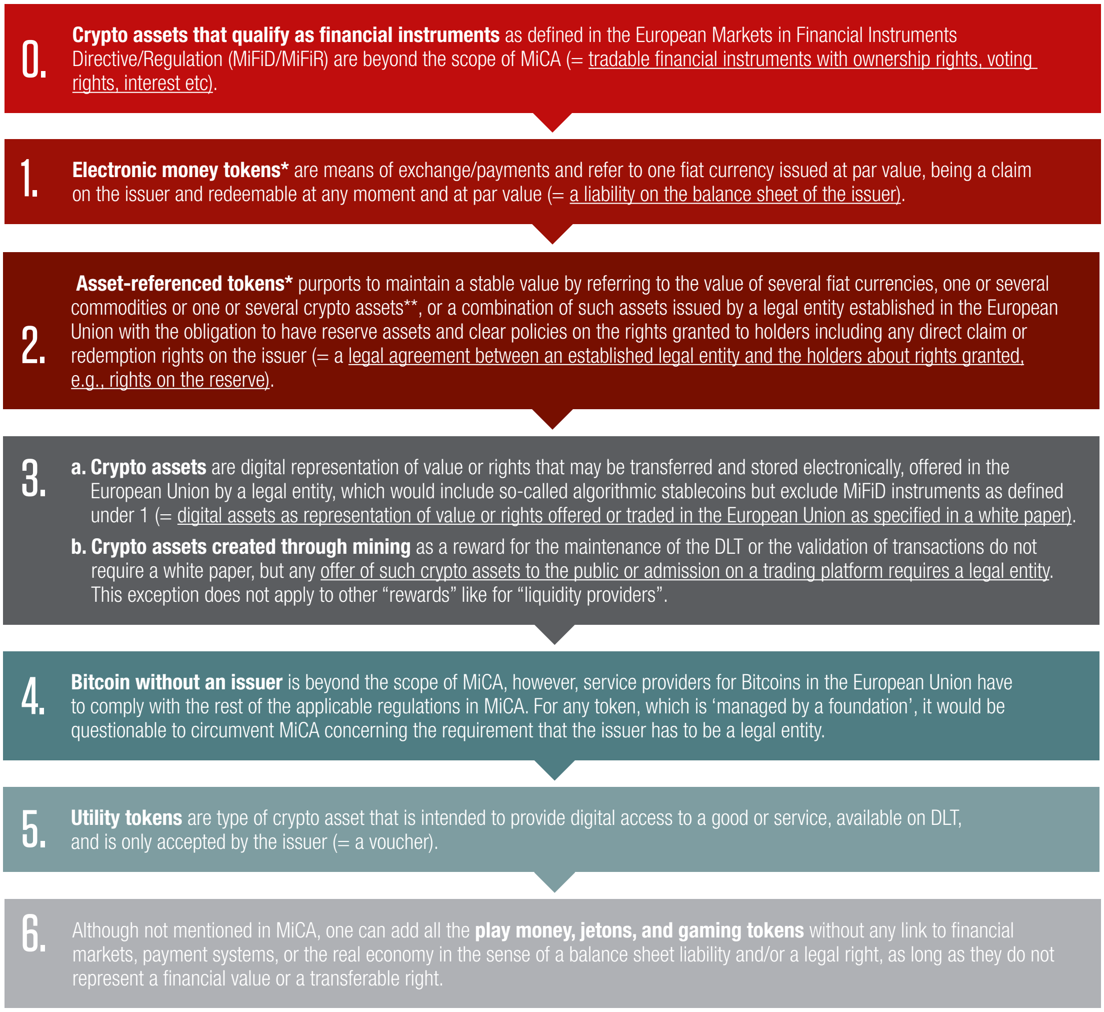
The way to get OHM could be as follows. First, one buys ether (highly volatile ETH tokens) for real money at a crypto exchange. Second, these ethers can be used to "borrow" DAI tokens (overcollateralized) at MakerDAO. Third, one can use the DAI to purchase ("to bond") OHM at OlympusDAO at a "market price" for this highly volatile and not very liquid token. As illustrated by Messari (2021), an investor can buy 1 OHM for a price of say 501 DAI and will receive 1 OHM for a small discount, while so-called "stakers" will be rewarded with 450 OHM, and 50 OHM will go to the OlympusDAO wallet. This adds to a generation of 501 OHM (for 501 DAO at 1:1), whereas the investor receives only 1 OHM at the market price. The rational for the investor could be either to sell this 1 OHM at a crypto exchange later (if there is a "greater fool"), or to "stake" (i.e., deposit) the OHM at OlympusDAO and participate in the specified incentive mechanism from following sales. In short, the idea is to invest DAI tokens in a "scheme" of OHM tokens to obtain a return from later investors' purchases.

An article on Coindesk.com, a news site for the crypto community, entitled "Olympus DAO might be the future of money (or it might be a Ponzi)," stated "Yes, it's a Ponzi scheme. But who cares? So are the dollars in your pocket" [Thurman (2021)].

Both statements are wrong. For one thing, money is a social agreement about future usage [Milkau and Bott (2018)] with "no questions asked" [Holström (2015)] and instances range from centrally issued banknotes to cigarette money, which existed in Germany after World War II. For another thing, a Ponzi scheme is a fraudulent investing scam, which disguises the fact that promised profits for earlier investors are taken from later investors. The mechanism of OlympusDAO may be near to a pyramid scheme, but the rules of the game are published in two white papers and the code can be inspected on the blockchain (by those who can read the bytecode).

However, the TVL of OlympusDAO collapsed on December 19, 2021, to U.S.$4.5 million after an all-time-high of nearly U.S.$500 million on November 24, 2021. Shortly before, a fork (i.e., a clone) of OlympusDAO, called AnubisDAO, surfing on the dogecoin hype, was launched on October 28, and attracted U.S.$60 million worth of ETH – and one day later all investments were drained from the protocol in a so-called "rug pull" scam (closing a DeFi app after redirection of the investments) [TheDefiant (2021)].

Figure 3: Schematical structure of MiCA regulating the authorization and supervision of crypto-asset service providers and issuers



0. **Crypto assets that qualify as financial instruments** as defined in the European Markets in Financial Instruments Directive/Regulation (MiFiD/MiFiR) are beyond the scope of MiCA (= tradable financial instruments with ownership rights, voting rights, interest etc).

1. **Electronic money tokens*** are means of exchange/payments and refer to one fiat currency issued at par value, being a claim on the issuer and redeemable at any moment and at par value (= a liability on the balance sheet of the issuer).

2. **Asset-referenced tokens*** purports to maintain a stable value by referring to the value of several fiat currencies, one or several commodities or one or several crypto assets**, or a combination of such assets issued by a legal entity established in the European Union with the obligation to have reserve assets and clear policies on the rights granted to holders including any direct claim or redemption rights on the issuer (= a legal agreement between an established legal entity and the holders about rights granted, e.g., rights on the reserve).

3. **a. Crypto assets** are digital representation of value or rights that may be transferred and stored electronically, offered in the European Union by a legal entity, which would include so-called algorithmic stablecoins but exclude MiFID instruments as defined under 1 (= digital assets as representation of value or rights offered or traded in the European Union as specified in a white paper).

   b. Crypto assets created through mining as a reward for the maintenance of the DLT or the validation of transactions do not require a white paper, but any offer of such crypto assets to the public or admission on a trading platform requires a legal entity. This exception does not apply to other "rewards" like for "liquidity providers".

4. **Bitcoin without an issuer** is beyond the scope of MiCA, however, service providers for Bitcoins in the European Union have to comply with the rest of the applicable regulations in MiCA. For any token, which is 'managed by a foundation', it would be questionable to circumvent MiCA concerning the requirement that the issuer has to be a legal entity.

5. **Utility tokens** are type of crypto asset that is intended to provide digital access to a good or service, available on DLT, and is only accepted by the issuer (= a voucher).

6. Although not mentioned in MiCA, one can add all the **play money, jetons, and gaming tokens** without any link to financial markets, payment systems, or the real economy in the sense of a balance sheet liability and/or a legal right, as long as they do not represent a financial value or a transferable right.

Notes: * The European Banking Authority (EBA) can classify tokens as significant tokens due to significance or the interconnectedness with the financial system with specific additional obligations for issuers. ** But not high-quality liquid assets such as treasuries or government bonds; however, reserve assets may be invested in highly liquid financial instruments. The requirement of a reserve excludes "algorithmic stablecoins", which would be "crypto assets".

It can be debated whether such DeFi is gaming (for fun) or gambling (for dollars), and the boundaries between speculation, gamification, and scams are fluid in DeFi:

- **Speculation**, in a completely value-free sense, is trading in financial markets with real-world assets by an individual economic agent in search of profit [as discussed in Théory de la Spéculation by Louis Bachelier in 1900].
- **Gamification**, from the perspective of this article, is a collective phenomenon of online communities created around common narratives ("us against them") to use venues for mutual actions like online games and usually coordinated via social media or messenger services. Such influencer-follower dynamics creates information asymmetry and makes it possible to game the system, which is prohibited within regulated financial markets.
- **Scam in the DeFi context** – independently of the technical details of how smart contracts could be exploited – has two sides, especially in new scam types like "rug pulls" [Xia et al. (2021)]: greedy fraudsters communicate their new "tokens" on social media like Telegram or Twitter, abandon it unexpectedly, and channel users' funds to their own accounts [Chainanalysis (2021)]. However, there has to be a greedy player, who sent value to opaque projects without reading the fine print in the technical documentations or smart contract bytecode scripts.

Although "gamification" seems rather modern, it has been around for centuries, with one the most well-known ones being the "tulipmania" in The Netherlands in 1634-1637. Contrary to the belief that tulipmania was a financial bubble, several researchers [Garber (1989), Day (2004), Thompson (2007)] have found that it was a phenomenon detached from the real economy. A closed community traded future contracts for tulips (that sellers never owned) against promises for money in the future (that buyers never had).

In the 21st century, this kind of collective hype developed into what the SEC (2021) called a "meme stock phenomenon". This correlates with the common narratives in gamification. Whether it is a story about a meme stock like Gamestop initiated by collective action of retail investors at the neo-broker Robinhood (described as "predatory trading" by Hasso et al. (2021), meme coins (like Dogecoin, Shiba Inu, or Floki Inu coins – driven by tweets or social media), or meme tokens (like many DeFi stories).

7. REGULATION AND CONSUMER PROTECTION

In a free economy, everything can be traded if one finds somebody else to pay the price you are willing to sell at, including used stamps, baseball cards, or meme coins, even though none has any intrinsic value. Likewise, every adult may play games or gamble for money at a licensed casino. But the "rules of the game" have to be transparent for the consumer. They need to know what liabilities or obligations a counterparty has. From the perspective of the user, consumer protection, transparent agreements, certainty about the legal liabilities of a counterparty, and the possibility of legal recourse are key.

An instructive example is the current proposal for the Markets in Crypto-Assets (MiCA) regulation in the European Union[4] [EC (2020)], which defines the obligations of issuers or service providers towards consumers [condensed in abbreviated form in Figure 3]. Independent of the hierarchical classification of MiCA, one general requirement has to be emphasized. As Article 53 of MiCA states: "Crypto-asset services shall only be provided by legal persons that have a registered office in a Member State of the Union and that have been authorized ..."

Only legal entities can perform regulated crypto-asset services in the European Union and have to act in the best interest of clients, especially if they exchange crypto-assets against fiat currency or against other crypto-assets (MiCA, Art. 68-70).

MiCA will regulate the authorization and supervision of crypto-asset service providers and issuers (or refers to tokens as financial instruments covered be the Markets in Financial Instruments Directive/Regulation (MiFiD/MiFiR)) and provides the basis for DeFi service providers or DeFi issuers active in the European Union. This approach avoids discussions about theoretical taxonomies independent of any enforceability. MiCA makes clear that any "service without entity" or "issuance without entity" would be not compliant.

As MiCA will not come into force until end of 2022, the impact of such a regulation on crypto-assets and DeFi remains to be seen. However, there is one case in the "crypto ecosystem" – although not DeFi, but CeFi (centralized finance) – which is emblematic of the problem with consumer protection: the stablecoin tether. Tether was launched in 2014 and is at present the "stablecoin" with the largest market capitalization (about 48 percent of all stablecoins as quoted on CoinMarketCap on December 20, 2021).[5]

[4] For a review of the regulatory environment in the U.S., see Gorton and Zhang (2021) and Allen (2021).
[5] For a general overview of stablecoins, please refer to Aramonte et al. (2021) and Waller (2021).

However, an important point was made by Jerome Powell and Jens Weidmann (2021) at the 2021 BIS Innovation Summit: all stablecoins need to "borrow" their stability from traditional currencies and are, subsequently, no rivals to the U.S. dollar, yen, or the euro. Aramonte et al. (2021) suggest that "Stablecoins are inherently fragile. ... The vulnerability is similar to that of traditional intermediaries, such as money market funds, whose investors expect to be able to redeem in cash at par." One can say that any constant net assets value (CNAV) model is based on investors' money, while a currency with silver or gold standard is based on reserved assets of the issuer.

The conduct of the conglomerate issuing tether was investigated by the N.Y. Attorney General (2021) and the U.S. Commodity Futures Trading Commission, and the latter issued warnings regarding tether and the associated Bitfinex crypto exchange. According to CFTC (2021): "However, the Tether order finds that from at least June 1, 2016 to February 25, 2019, Tether misrepresented to customers and the market that Tether maintained sufficient U.S. dollar reserves to back every USDT in circulation with the 'equivalent amount of corresponding fiat currency' held by Tether and 'safely deposited' in Tether's bank accounts. ... Tether held sufficient fiat reserves in its accounts to back USDT tether tokens in circulation for only 27.6 percent of the days in a 26-month sample time period from 2016 through 2018. ... and that Tether transferred Tether reserve funds to Bitfinex, including when Bitfinex needed help responding to a 'liquidity crisis'."

Subsequently, Tether started to publish "consolidated reserves reports" in 2021, and the most recent report [Tether, 2021] reveals that this "stablecoin" resembles a (CNAV?) money market fund with a reserve of 54 percent commercial papers, corporate bonds, and secured loans, but only 39 percent cash, bank deposits, and treasuries. Although tether is not a DeFi token but issued by a "central" legal entity, the development illustrates the difference between original claim and actual implementation, which is relevant for the user. Regulation and supervision with defined standards and obligations for issuers and service providers of CeFi and DeFi crypto assets achieves transparency and, as a result, customer protection.

8. CONCLUSION

There are increasing concerns among banking supervisors regarding DeFis. Agustín Carstens, General Manager of the Bank of International Settlements (BIS), recently highlighted this issue by stating that "Concerns also arise in the growing crypto universe of decentralized finance ... DeFi appears to be operating largely within its own ecosystem, with little in the way of financial intermediation services being provided to the real economy. ... But the potential for spillovers should not be underestimated, especially since the stablecoin arrangements themselves can create important links. As history confirms, anything that grows exponentially is unlikely to remain self-contained and thus merits the closest attention" [Carstens (2021)].

Contrary to the discussions regarding technology or semantics (like "decentralized" for an app on a "distributed" ledger technology), the users tend to predominantly focus on, and are most concerned with, legal, economic, and sociological issues. They want to know whether what they are considering is speculation, gaming, or a scam? Is there adequate consumer protection? Cui bono?

Within the DeFi universe, there are single entities providing the interface to the "meeting of the minds" and economic beneficiaries with voting rights and/or receiving rewards – independent of whether something is called "decentralized finance" in 2022 or "Société Anonyme" as in the French Code de Commerce of 1807.

Taking the users' perspective as the guiding principal, legislation, regulation, and supervision have to clearly state whether a business is a regulated legal entity with sufficient consumer protection and mandatory "shock absorbers", whether it is company with a casino license for gambling (with jetons to be exchanged for money), a platform provider of a collective online game (potentially with fee-based add-ons), or whether it is an activity beyond the remit of regulations.

In my opinion, DeFi is neither "decentralized" nor provides financial services with the necessary consumer protection in place. It can best be portrayed as a "game of tokens" with substantial information asymmetries. However, as Carstens (2021) states, its spillovers could jeopardize the real economy and create new risks – for financial services and consumers alike.

REFERENCES

Aave, 2020, “Aave general terms of use,” https://bit.ly/3n0FjFq

Aave, 2021, “Avenomics,” https://bit.ly/34sC7Mv

Allen, H. J., 2021, “Stablecoins: how do they work, how are they used, and what are their risks?”, Prepared statement made at the hearing held at the U.S. Senate Committee on Banking, Housing, and Urban Affairs, December 14

Aramonte, S., W. Huang, and A. Schrimpf, 2021, “DeFi risks and the decentralisation illusion,” BIS Quarterly Review, December, 21-36, https://bit.ly/3t2n2eE

Auer, R., 2019, “Beyond the doomsday economics of “proof-of-work” in cryptocurrencies,” BIS Working Papers no. 765, https://bit.ly/3eXzPXy

Auer, R., C. Monnet, and H. S. Shin, 2021, “Distributed ledger and the governance of money,” BIS Working Papers no. 924, https://bit.ly/3pWGYxQ

Bachelier, M. L., 1900, “Théorie de la spéculation,” Annales scientifiques de l'É.N.S., 3e série, tome 17 (1900), 21-86, https://bit.ly/3t6zkmm

BGH, 2012, “Urteil des X. Zivilsenats vom 16.10.2012 – X ZR 37/12,2 Bundesgerichtshofs, https://bit.ly/3q0VQeG

BTC, 2021, https://bit.ly/3JHBNd3

Carstens, A., 2021, “Non-bank financial sector: systemic regulation needed,” Foreword, BIS Quarterly Review, https://bit.ly/3eT36Tm

CFTC, 2021, “CFTC orders Tether and Bitfinex to pay fines totaling $42.5 million,” U.S. Commodity Futures Trading Commission, https://bit.ly/3ztr0OK

Chainanalysis, 2021, “The biggest threat to trust in cryptocurrency: Rug Pulls put 2021 scam revenue close to all-time highs,” Chainanalysis, https://bit.ly/3G2RoBo

Coase, R., 1937, “The nature of the firm,” Economica 4:16, 386–405

Crenshaw, C. A., 2021, “Statement on DeFi risks, regulations, and opportunities,” SEC Statement, published in The International Journal of Blockchain Law, Vol. 1, https://bit.ly/32Ntdc7

Day, Ch. C., 2004, “Is there a tulip in your future? Ruminations on tulip mania and the innovative Dutch futures markets,” Journal des Economistes et des Etudes Humaines 14:2, 151-170

De Filippi, P., and A. Wright, 2018, Blockchain and the law: the rule of code, Harvard University Press

EC, 2020, “Proposal for a regulation of the European Parliament and of the Council on Markets in crypto-assets (aka MiCA),” European Commission, https://bit.ly/338qCZL

Elliptic, 2021, “DeFi: risk, regulation, and the rise of DeCrime,” Elliptic Report, https://bit.ly/339vo9m

Etherscan, 2021, https://bit.ly/3t0eXXU

Garber, P. M., 1989, “Tulipmania,” Journal of Political Economy 97:3, 535-560

Gorton, G. B., and J. Zhang, 2021, “Taming wildcat stablecoins,” https://bit.ly/3pYRZib

Grossman, S. J., and O. D. Hart, 1986, “The costs and benefits of ownership: a theory of vertical and lateral integration,” Journal of Political Economy 94:4, 691–719

Hasso, T., D. Müller, M. Pelster, and S. Warkulat, 2021, “Who participated in the GameStop frenzy? Evidence from brokerage accounts,” Finance Research Letters, https://bit.ly/3qUKBDv

Holmström, B., 2015, “Understanding the role of debt in the financial system,” BIS Working Paper no. 479, https://bit.ly/3HCq7q0

Kreitmar, M., 2021, “Tangany: innovation killer: network fees. And solutions,” Crypto Asset Conference, Frankfurt School Blockchain Center, https://bit.ly/3eTnN1B

MakerDAO, 2020, “The maker protocol: MakerDAO's multi-collateral Dai (MCD) system,” https://bit.ly/32LSkvX

Makarov, I., and A. Schoar, 2021, “Blockchain analysis of the bitcoin market,” NBER Working Paper no. 29396, https://bit.ly/3G5Kxr1

Messari, 2021, “How sales and profits works in OlympusDAO,” Twitter, https://bit.ly/3HEWDrp

Milkau, U., and J. Bott, 2018, “Digital currencies and the concept of money as a social agreement,” Journal of Payments Strategy & Systems 12:3, 213-231

N.Y. Attorney General, 2021, “Attorney General James ends virtual currency trading platform Bitfinex's illegal activities in New York,” Press Release, https://on.ny.gov/3qRsG0K ; https://on.ny.gov/3HDrCEn

Oasis, 2021, “Terms of service,” https://bit.ly/32R9BDG

OlympusDAO, 2021, https://bit.ly/32TRNrl

Parnas, D. L., 1994, “Software aging,” ICSE ‘94, Proceedings of the 16th international conference on Software engineering, IEEE, 279-287, https://bit.ly/3JONhew

Powell, J., and J. Weidmann, 2021, “How can central banks innovate in the digital age?” BIS Innovation Summit 2021, https://bit.ly/3HVyqxt

Schär, F., 2021, “Decentralized finance: on blockchain- and smart contract-based financial markets,” Federal Reserve Bank of St. Louis Review 103:2, 153-174

SEC, 2017, “Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934: The DAO,” Securities and Exchange Commission, Release No. 81207, https://bit.ly/3pZBm5T

SEC, 2021, “Staff report on equity and options market structure conditions in early 2021,” U.S. Securities and Exchange Commission, https://bit.ly/3t33KWS

Shin, H. S., 2021, “Decentralisation in digital finance: possibilities and limits,” presentation London Business School, https://bit.ly/3HlF61t

Tether, 2021, “Consolidated reserves report,” Tether Holdings Limited, https://bit.ly/3zsO8wV

TheDefiant, 2021, “Anubis DAO descends into the underworld after $60M exploit,” https://bit.ly/3EZksZp

Thompson, E. A., 2007, “The tulipmania: fact or artifact?” Public Choice 130:1–2, 99–114

Thurman, A., 2021, Coindesk, https://bit.ly/3J0ajm3

Uniswap, 2021, “Uniswap Labs terms of service,” https://bit.ly/32LGUs1

Waller, Ch. J., 2021, “Reflections on stablecoins and payments innovations,” speech at “Planning for surprises, learning from crises” 2021 Financial Stability Conference, https://bit.ly/3Hlybpo

Williamson, O. E., 1979, “Transaction-cost economics: the governance of contractual relations,” Journal of Law and Economics 22:2, 233-261

Wyoming, 2021, “Wyoming decentralized autonomous organization supplement,” State of Wyoming, File No. SF0038, https://bit.ly/3mXBVve

CENTRAL BANK DIGITAL CURRENCIES: MUCH ADO ABOUT NOTHING?

JAY CULLEN | Professor of Financial Regulation and Head of Law, Criminology and Policing, Edge Hill University; Research Professor in Law, University of Oslo

ABSTRACT

This article examines retail central bank digital currencies (CBDCs), a proposed financial technology that central banks around the world are considering implementing. Proponents of such payment instruments argue that they will produce considerable benefits for adopting countries, principally in the fields of competition in payments markets, financial inclusion, and macroeconomic stability. This article critically evaluates these purported benefits and finds that many of the claims made in their support do not stand up to scrutiny and could, in fact, be realized without the introduction of a central bank retail payment instrument. More significantly, the benefits cited by proponents of such instruments may produce considerable negative externalities in other domains, particularly in relation to financial stability.

1. INTRODUCTION

The emergence of cryptocurrencies and alternative currency systems such as stablecoins holds promise for widening access to finance and addressing certain socio-economic issues. However, these new currencies also present considerable challenges in relation to consumer protection, the dangers of data mismanagement, preservation of privacy, and the mitigation of cybersecurity risks. Cryptocurrencies are, for example, prone to security issues, susceptible to theft, and extremely energy intensive. Such factors have led some jurisdictions to ban them altogether [BBC (2021), Fabrichnaya and Marrow (2022)]. Stablecoins such as Facebook's Diem[1] offer more stability than cryptocurrencies, but if widely adopted they also threaten to concentrate power further in the hands of large multinational corporations. Moreover, the financial stability dangers of privately-created "monies", designed to operate like regulated money but in largely unregulated spaces, are well documented.

Given these trends, a much-heralded recent development has been the mooted introduction of retail central bank digital currencies (CBDCs). This financial instrument would be provisioned to households and businesses alike: individuals and firms would be granted the option to have an account at the central bank holding fiat digital money that would provide, inter alia, convenient real-time payments, cash-like peer-to-peer functionality, and, where appropriate, anonymity between users. By nature, the CBDC would also provide a safe asset for holders, as the digital money held would be subject to zero default risk [Auer and Boehme (2020)].

Many leading central banks are now researching and assessing the feasibility and desirability of introducing such payment instruments, including the European Central Bank [ECB (2020)], the Bank of England [Bank of England (2021)], and the U.S. Federal Reserve [Federal Reserve System (2022)]. The People's Bank of China has already taken the decision to implement a CBDC [People's Bank of China (2021)] whilst the Bank of Japan is piloting CBDC in experimental settings with a view to potential introduction at a later date [Bank of Japan (2020)].

In this article, the claims of CBDC supporters are assessed to determine whether, on balance, the introduction of a retail CBDC would result in a net positive outcome for those jurisdictions with the capacity to do so. It considers whether the purported benefits of such instruments outweigh the

[1] Diem is the latest name of the currency proposed by the Libra Association in its White Paper v2.0.

potential risks attached to their introduction, particularly in the domains of competition in the payments market and financial stability. The article concludes that the answer to whether retail CBDCs would, on balance, be a progressive development in the monetary space, is no. Instead, regulation should be used to improve access to existing payments infrastructure and improve efficiencies in the consumer-finance interface.[2]

2. THE CONTEMPORARY PAYMENTS SYSTEM

At root, a payments system is the system through which units in an economy – governments, households, and businesses – move money between one another. There are two dominant payment media in modern economies: electronic money (often bank deposits) and physical currency. Electronic payments systems are used in lieu of tendering physical currency in transactions and in most countries continue to comprise by far the largest payment instrument by volume, a trend which was exacerbated by the COVID-19 pandemic. Cash substitutes including debit cards, credit cards, direct debits, and e-commerce payment service providers continue to expand, as demonstrated in the data from the U.K., where use of cash has fallen from over 50 percent of all payments in 2010 to around 17 percent in 2020 [U.K. Finance (2021)].[3]

Table 1: Total payment volumes in the U.K. 2020 (excluding CHAPS)

	£ MLN	%
Debit card	15,812	44.43
Cash	6,075	17.07
Direct debit	4,507	12.66
Faster payments (including other remote banking)	2,952	8.29
Credit/charge/purchasing card (of which 1,216 was contactless)	2,827	7.94
Bacs direct credit	1,945	5.46
Other	732	2.06
Standing order	557	1.56
Cheque	185	0.52
Total	35,592	100

Source U.K. Finance (2021)

But how are such electronic payments ultimately executed? In most jurisdictions, only a limited number of commercial (“settlement” or “clearing”) banks hold accounts at their national central bank (so-called reserve accounts) and engage in direct participation with the central bank’s payment infrastructure. When payments are made between accounts at these banks, the central bank moves reserves (central bank money) between the reserve accounts corresponding to the amounts paid. The outstanding bilateral “netted” balance is transferred in reserves each day. In turn, other financial institutions, which are not part of the clearing system (so-called indirect participants), hold accounts at commercial banks.

When a payment is made between these financial institutions, instructions are sent to debit or credit the correspondent accounts at the clearing banks, and reserves will be transferred at the central bank level to settle the payment. This means that at present a payment made through a payments service provider that does not have a reserve account at the central bank is still transacted via reserve accounts held at the central bank by the clearing banks.

Access to the central bank’s balance sheet for a narrow set of financial institutions, referred to as a “tiered participation arrangement” (TPA), is therefore a feature of today’s payments market infrastructure in many countries. These TPAs allow many participants to access the central payments system, but they must do so indirectly, building upon the settlement and clearing services provided by those institutions with access to the central bank’s reserve systems. This means that most payment institutions have no access – direct or indirect – to central bank money. Network effects, in combination with economies of scale and regulatory access restrictions, mean that establishing competing networks is economically unviable. Instead, the only option for rival suppliers wishing to compete in the market is to gain access to an existing installed infrastructure base [Cullen (2021)].

In recognition of these obstacles, in some jurisdictions, attempts have been made to expand payments market access. For example, the E.U. has introduced the second Payments Services Directive (PSD2) [E.U. (2015)], which enables retail and business bank customers to use third-party providers (TPPs) to manage their finances and initiate

[2] The article cannot evaluate all public digital currency initiatives in all jurisdictions. It, therefore, confines its analysis to the purported generalized benefits of CBDC in principle, which may vary in some states. For a discussion of the potential forms of new digital money, see Grey (2019).

[3] In fact, cash payments decreased by 35 percent to 6.1 billion between 2019 and 2020.

electronic payments on their behalf, removing the need for banks to actively participate in a payments service. To achieve this, PSD2 requires firms that hold individuals' payment accounts to provide TPPs with access to bank's customer data and payment functionality of users' online payment accounts. Indeed, some countries, including the U.K. and Lithuania, have gone further and access to the central bank balance sheet has been expanded to include non-banks, although this arrangement remains uncommon globally.

3. ENTER CBDCS

The Bank for International Settlements (BIS) notes that central bank digital currencies are "envisioned by most to be a new form of central bank money. That is, a central bank liability, denominated in an existing unit of account, which serve both as a medium of exchange and a store of value" [BIS (2018)]. Although central bank money already exists, most proposed iterations of a central bank digital currency would expand user eligibility to encompass retail consumers. In legal terms central bank digital currencies would, like paper banknotes and coins, be fiat money: a liability of the government. As such, it would provide a digital counterpart to physical cash and should, therefore, share the features of cash, which make it attractive as a payment medium. Such features include trust in the issuing entity, guaranteed real-time finality and settlement, widespread acceptance, ease-of-use, unfettered access to the medium, and legal tender status.

Proponents claim that the benefits – both direct and indirect – of a retail central bank digital currency would be substantial. These benefits are normally grouped into three broad categories:

1. Financial market competitiveness: because central bank digital currency users would be granted direct access to central bank money, existing payments markets would be liberalized and the tiered participation arrangement model would become defunct. Consumers and less-established financial institutions in many jurisdictions remain reliant upon access to the rails of a few large providers of settlement systems. Whilst the provisions of legislation such as PSD2 mandate that payments providers must be granted access to the data held by settlement banks, the network effects of holding consumer bank information mean that banks operate at a competitive advantage in relation to these payment providers. Because banks may offer bundled products alongside payment services, they can cross-subsidize their payments services and infrastructure costs; there are well-established findings that banks and other financial institutions with direct access to central bank settlement systems enjoy competitive rents from these privileges [Ferreira (2013)].

2. Financial inclusion: a retail central bank digital currency with universal coverage would ensure access for all citizens to a simple method of payment and store of value, particularly in circumstances where alternative payments providers have been unable to offer transaction accounts to target populations. A central bank digital currency might be highly beneficial for low-income households, which tend to rely heavily on cash and whose access to bank accounts may be limited. Introducing retail central bank digital currencies might, therefore, promote financial inclusion amongst economically vulnerable households. A central bank digital currency might also enhance commerce. Small businesses, which are often charged large account and transaction fees, and must contend with additional charges for accepting debit and credit card payments, might benefit from the introduction of a central bank digital currency; research suggests that removal of existing payment transaction fees has the capacity to raise GDP by as much as 3 percent [Barrdear and Kumhof (2016)].

3. Financial stability: in a financial world in which institutions rely upon the production of a constant flow of safe assets to act as repositories for capital and for funding purposes, central bank digital currencies provide a new asset class of secure central bank instruments, no different in credit or liquidity terms than bank reserves. Large institutional cash pools held by money managers cannot be deployed in meaningful volumes into bank deposits thanks to deposit insurance caps that limit their utility as stores of value. This, in turn, reduces the supply of safe assets to the financial system and has contributed to the growth of shadow banking which, at its core, is a system designed to cater to the institutional need for private forms of money. History has demonstrated on numerous occasions that runs on forms of such private money-substitutes present systemic threats to the wider economy [Ricks et al. (2021)]. Central bank digital currencies would also conceivably make monetary policy more effective. On the assumption that central bank digital currencies pay a rate of interest, they could increase the responsiveness of an economy to changes in the policy rate. If any entity in the economy can earn the central bank rate, then there would be no incentive to place their funds on deposit or make loans for lower than the rate they could earn, risk-free, from the central bank. In the event of recession or other form of economic crisis, central bank digital currencies would facilitate provision of fiscal stimulus to citizens, thereby avoiding some of the blockages that undermine rescue and recovery efforts during times of stress.

4. CBDC: MUCH ADO ABOUT NOTHING?

Notwithstanding the putative benefits of central bank digital currencies, it remains unclear whether they would solve any of the supposed market failures they are designed to address or, even if those problems might be addressed, why the private sector cannot do the same at less cost and less disruption. While it is important to note that the claims made in favor of central bank digital currencies enjoy different degrees of salience dependent upon the jurisdiction in question, when weighed against the potential damage they could inflict upon the financial system, it becomes clear that many of the claimed benefits of central bank digital currencies also involve considerable negative externalities.

4.1 Claim 1: CBDCs will break oligopoly in payments markets

In the context of payments systems competition, claims are made frequently that incumbent private payment service providers, such as Visa and Mastercard in the West, and Alipay and Tencent in Asia, enjoy oligopolistic powers.[4] As such, the introduction of a central bank digital currency would arguably serve to widen access to payments markets infrastructure by permitting access to the central bank balance sheet to non-bank competitors. This would have the potential to significantly disrupt the payments markets through offering routes to circumvent the hold that existing payment providers enjoy over the payments market through their relationships with large commercial banks.

Yet, there are at least two principal objections to this claim. The first is that it is far from clear that introducing a central bank digital currency would improve competition in the payments market. Indeed, central banks' cost efficiencies and potential dominance in such markets – and the fact that they are rule-setters for market participants – might stifle competition and dissuade potential alternative payment infrastructure development. Public funds would necessarily have to be employed to administer such accounts, which is, in and of itself, an allocation decision that ought to be subject to democratic, not technocratic, scrutiny. Even if a retail central bank digital currency system was approved by legislators, the central bank – indeed any government entity – can provide such services at or below cost, which is a substantial competitive advantage in a market with such volume and scale. Whilst private sector providers might have to increase prices elsewhere to subsidize the costs of maintaining payment systems, a central bank would be under no such pressure. Moreover, as central banks progressively ate into the payments market space, it would likely require private sector entities to increase, rather than decrease, prices in other business lines in order to maintain margins. Given that it is not envisaged that central banks would offer products beyond basic payment services, the knock-on effects of central bank digital currency introduction might actually force some providers out of the market, reducing competition in the process, while simultaneously making financial products in other areas less affordable.

In the case of a central bank digital currency, payment markets – which are often not the preserve of central bank oversight alone – would be drawn into central banks' direct regulatory purview. The central bank would, by implication, be required to act as regulator and competitor in the payments market. This is, by any standard, unusual in markets. Accordingly, even if a potential market failure is identified – which arguably exists in relation to markets for payment media technologies in some jurisdictions – remedying it through the introduction of a government instrumentality might be considered excessive intrusion.

4.2 Claim 2: CBDCs will improve financial inclusion

This is, on the face of it, a compelling claim. In Eastern Europe, for example, large proportions of citizens remain unbanked.[5] Similar trends exist in the U.S., where in 2021 the Federal Reserve estimated that 5 percent of U.S. citizens were unbanked, with a further 13 percent "underbanked" [Federal Reserve System (2021)]. Where someone lacks access to bank account services, they will often suffer significant financial detriment; for example, they may be charged higher fees for making payments, those payments may take longer to clear, and they are subject to increased risk of fraud or theft because

[4] There are many different payment markets and technologies used, so direct comparisons are difficult to make. However, card payments and mobile payments are two of the largest payment markets. In the debit and credit card payment market, Visa and Mastercard between them controlled approximately 90 percent of the U.S. market in 2020. Alipay and TenCent controlled approximately 95 percent of the mobile payments market in China in 2020, where mobile payments comprised over 85 percent of all payments made.

[5] According to recent research by the World Savings Bank Institute, in the European context, more than 37 million adult E.U. citizens (8.6 percent of Europe's adult population) lack access to formal financial services. The numbers in Eastern Europe (including euro area countries) are noteworthy. In Romania, almost 40 percent of the population is unbanked, in Bulgaria it is 37 percent, in Hungary it is 27.7 percent, in Slovakia it is 22.8 percent, and in Lithuania and Poland it is 22.1 percent each. Even countries with relatively advanced financial systems have a large proportion of unbanked adults, including Italy (12.7 percent), Portugal (12.6 percent), and Greece (12.5 percent) [WSBI (2016)].

they usually transact in cash. Such citizens also find it more difficult to access credit. A retail central bank digital currency would potentially reduce the number of citizens locked out of basic financial services, which would not only improve quality of life but would also boost economic performance, as the citizens in turn could divert their resources away from the time-consuming and expensive tasks of meeting basic banking needs.

On the other hand, it is also true that not accessing basic financial services is often a choice for people, rather than an imposition. Although it is correct to note that many citizens in some countries do not hold bank accounts, their motivations for not doing so are plural and – in many cases – entirely unrelated to a lack of access to available services. For example, in the U.S., of the 5 percent of adults without a bank account, the most common reason (29 percent of respondents) provided for why this was the case was because they did not have enough money to open one. Now, this might be because the associated bank account fees are too high, but this is not conclusive. Moreover, if this was indeed the case, as will be explained shortly, there are better solutions to remedying such a problem than by deconstructing the incumbent payments industry.

In the same survey, just over 16 percent of respondents stated that the main reason they did not hold bank accounts was that they did not trust banks. While this may be understandable given recent banking history, it is certainly not out of the question that those people would be equally distrustful of a government-sponsored instrumentality. Other reasons for not accessing regulated financial services included that respondents did not want to compromise their privacy (7.1 percent), which provides another potential stumbling block for the adoption of a government-issued digital currency: after all, is it likely that consumer would be more willing to hold an instrument that might be monitored by government than one issued by a private sector entity? The only motivations to the question of why respondents lacked access to basic financial services that a central bank digital currency might conclusively address were that they had poor credit histories and were, therefore, ineligible (8 percent), or that account fees were too high (7.3 percent).

Even if one agrees that a lack of access to basic financial services amongst certain segments of the population is a problem that must be tackled, it is not clear that the provision of a system of government bank accounts, serviced by a central bank digital currency, is a prerequisite to achieving that objective; indeed, in many cases providing a central bank digital currency might not improve financial inclusion levels amongst citizens. First, the introduction of digital technology does not always guarantee greater access to services if it is done in a way that does not preserve older, more established technologies. In the U.K. context, several studies have found that increasing digitalization in banking is likely to reduce access to finance for the most vulnerable [House of Lords Liaison Committee (2019)]. Second, although some are excluded from the credit system or are charged more for financial services due to their lack of stable financial histories, a central bank digital currency would likely not help with this. As noted by Barry Eichengreen, the unbanked pay more for services because credit providers treat possession of a bank account as a signal of financial responsibility and reliability, yet a central bank digital currency "available to everyone unconditionally would not signal anything" [House of Lords Liaison Committee (2019)].

For these reasons, it could be argued that a better solution to addressing the problem of affording basic financial services to individuals who are involuntarily unbanked is to require private sector organizations with expertise in delivering such services to provide them directly. Evidence from the U.K. suggests strongly that laws requiring private sector financial firms to offer basic bank accounts – accounts with feature such as direct debit facilities, debit cards, cash machine access, and no fees – can work exceptionally well. In December 2016, the U.K. Treasury reported that 4 million such accounts had been opened in the U.K. since the 1990s [H.M. Treasury (2016)]. In other jurisdictions, such accounts could also be provided by existing government instrumentalities, such as postal service organizations or national savings banks. In short, while there may be an issue of financial inclusion in some jurisdictions, it is not settled as to whether a retail central bank digital currency is the solution.

4.3 Claim 3: CBDCs could be used to improve macroeconomic outcomes

As noted in the opening section, there are many convincing arguments that the introduction of central bank digital currencies would enhance macroeconomic outcomes: in particular, they would improve the stability of the financial system by restricting the universe of "shadow" monies, and they would also make monetary policy more efficacious. Whilst perhaps seemingly distinct, these themes are closely linked.

Taking the financial stability point first; no doubt that there are good arguments for restricting the creation of private monetary instruments, which are often at the root of financial crises. Yet, introducing a retail central bank digital currency could actually impact financial stability in a most perilous fashion. If retail bank deposits were made exchangeable at par for central bank money, and non-banks and individuals were permitted to hold central bank accounts, a significant proportion of bank deposits may flow into central bank digital currencies. This would lead to the – potentially fatal – loss of low-cost and stable funding for the commercial banking system. Banks could attempt to address any deposit outflows by raising deposit rates or seeking other funding sources, such as wholesale or bond financing, but such funding sources are more expensive and, in the case of wholesale funding, much less stable. Such funding structures would also be penalized by liquidity regulations under the Basel Accords, placing cost pressures on bank balance sheets and forcing them to shrink. This may be a desirable outcome from a public policy perspective, but it is unclear why introducing a central bank digital currency ought to be the gateway to this shift.

These dangers would only be exacerbated during financial crises: there are strong incentives for bank depositors to "run" from bank deposits into central bank digital currencies during periods of banking system stress. As noted earlier, private monetary liabilities, including bank deposits, are subject to credit risk, whereas claims on the central bank are not. At present, during systemic financial distress depositors may shift their deposits to alternative financial institutions, into financial assets such as government securities, or withdraw their deposits in cash. The widespread availability of a safe central bank asset would give them the option to instead move their deposits into central bank money and give rise to the potential of a "digital run" even on the strongest financial institutions, leading to contagion and wider financial system instability. This was witnessed during the great financial crisis (GFC), as governments in many countries were forced to guarantee the entire bank deposit base in order to forestall a widespread run on national banking systems.

These issues are amplified when central bank digital currencies are invoked as a potential monetary policy tool. As stated by many proponents of central bank digital currencies, monetary policy transmission could be optimized using such balances by the central bank paying interest on accounts at approximately the same rate it pays currently on bank reserves. According to Bordo and Leven (2019), for example: "Consumers and businesses would be able to receive essentially the same interest on checkable deposits and other current accounts that commercial banks receive on reserves held at the central banks, that is, the interest rate on reserves (IOR) less a very small margin to cover operating costs."

> *"There are extremely serious consequences for the banking system that might flow from the introduction of central bank digital currencies."*

The consequence of such a development is likely to be that prompted by the safety that central bank digital currencies provide, depositors will transfer their funds from the banking system (where deposits are normally not remunerated or pay very little interest) to central bank digital currencies, where holders will be guaranteed somewhat close to the interest rate on reserves paid by the central bank. In addition, because the interest rate on reserves exceeds the return on other safe liquid assets, such as Treasury bonds, the likelihood that other near-monies would also come under pressure cannot be discounted. Accordingly, there are tremendous incentives for bank depositors to switch into central bank digital currencies with the likely result that bank liquidity will dry up.

Now, at first brush this may seem reasonable: after all, why should commercial banks profit from interest on their assets at the central bank, which other entities cannot hold, particularly as the spread between deposit rates at banks and the interest rates they charge are often large? To this, one must consider the costs that banks must contend with, above and beyond the costs that maintaining a payment system entails. Private intermediaries must cover their non-interest expenses, in particular the costs associated with maintaining physical infrastructure and IT systems, guarding against fraud, engaging in compliance, and assessing borrower creditworthiness. In addition, banks must absorb bad debts, which, if not accounted for fully though interest rate offsets elsewhere, must be written off against capital.

There are, therefore, extremely serious consequences for the banking system that might flow from the introduction of central bank digital currencies. Although there may be ways to mitigate the risk of deposit outflows into central bank digital currencies – for example, by capping the interest paid on them or, alternatively, to limit the balances that may be held in such accounts – none would be immune from potential political interference because the fundamental inequity between the interest rate on reserves paid to banks and the interest paid to retail central bank digital currencies holders would subsist [Selgin (2021)].

Moreover, assuming that a central bank digital currency would crowd out some private sector intermediaries, who will then provide credit? While banks are not the only credit providers, they remain the principal class of lenders that provide credit to the private (and indeed, public) sectors. Assuming that the introduction of central bank digital currencies impacts the banking system to the extent that many banks fail, it is likely that (i) competition in the banking system is reduced and only those institutions that are large and strong enough survive; and (ii) sources of credit in the economy for retail borrowers are narrowed. Neither of these outcomes would be net positive for the majority of consumers. To the extent that the central bank becomes pressured itself to issue credit to fill the gap left by those that have exited the market, the consequences might be disastrous.

5. A MIDDLE WAY: THE BANK OF ENGLAND

There are design options available to central banks that would diminish the impact that central bank digital currencies would exert on the financial system while allowing citizens more choice and flexibility in their selection of financial services and accelerating financial inclusion. If such designs were adopted, a central bank digital currency, rather than threatening the financial system might be financial stability enhancing.

The most convincing, from the perspective of promoting equity between financial institutions and promoting consumer choice, is for central banks to open their settlement systems to non-bank payment service providers. By doing so, such firms would be able to access the same payment systems as traditional commercial banks. This avoids the obstacles of PSD2 discussed earlier, the provisions of which are predicated on a special category of financial institution opening their payment rails to rival firms. Instead, under a broadened access plan, wholesale access to digital currency is provided to non-bank payment firms outright. In fact, this has been the approach taken by the Bank of England since 2017, when it allowed fintech firms to open settlement accounts with it [Bank of England (2017)]. Under this arrangement, the Bank of England permits non-banks to hold reserve accounts directly, although importantly, not their customers.

This initiative could be operationalized by other means, by allowing financial institutions to hold what is referred to as "synthetic CBDC" (sCBDC) [Adrian and Mancini-Griffoli (2019)]. In this case, the accounts would contain central bank digital currencies rather than central bank reserves. Private tech firms would then issue their own e-money which would be backed by the synthetic central bank digital currencies. The central bank would thereby merely offer settlement services to e-money providers. While full access to central bank reserves would not be a part of this plan, it would be expected to increase payments market efficiency by carving out a payments infrastructure with access to the central bank's balance sheet that is not routed through incumbent banks. In this way, competitor institutions would have the ability to tap into the central bank framework and diminish the market power of existing large financial institutions.

In combination, these factors might also improve financial stability by broadening the landscape of institutions able to offer settlement services in central bank money. It would also arguably reduce the potential for a further financial stability risk from crystallizing: the risk that rival unregulated currencies and means of payment might emerge to rival fiat money and undermine regulatory capacities. Planned forms of such alternative currencies anchor themselves to fiat money in order to gain broad acceptance and trustworthiness. By offering a standardized and non-proprietary interoperable payments infrastructure, this might also ensure that large tech firms could not come to dominate payments markets; in effect avoiding the replacement of one set of dominant institutions by another.

6. CONCLUSION

In most jurisdictions, there is very little that a central bank digital currency might achieve that current public and private sector solutions cannot, provided that certain access rules to payments infrastructure are modified. In relation to financial stability and competition in the payments landscape, a retail central bank digital currency has the potential to upend the traditional banking and payments systems. Whether this would be a welcome turn depends upon a number of judgments but most pertinently: whether one believes that retail central bank digital currencies would offer services that properly regulated private sector intermediaries could not; that the introduction of a potential monopoly power in the payments space is desirable; and that the potential destruction of the predominant source of credit in the economy is warranted. If there is insufficient evidence for these propositions, as this article suggests, regulators in most jurisdictions should remain circumspect about retail central bank digital currencies.

REFERENCES

Adrian, T., and T. Mancini-Griffoli, 2019, "The rise of digital money," IMF FinTech Note no. 19/01

Auer, R., and R. Boehme, 2020, "The technology of retail central bank digital currency," BIS Quarterly Review, March 1, https://bit.ly/3rSVGqF

Bank of England, 2017, "Bank of England extends direct access to RTGS accounts to non-bank payment service providers," July 19, https://bit.ly/3sl6FlP

Bank of England, 2021, "Statement on central bank digital currency next steps," November 9, https://bit.ly/3gJPN8R

Bank of Japan, 2020, "The Bank of Japan's approach to central bank digital currency," October, https://bit.ly/3rR4aP2

Barrdear, J., and M. Kumhof, 2016, "The macroeconomics of central bank issued digital currencies," Bank of England Staff Working Paper no. 605

BBC, 2021, "China declares all crypto-currency transactions illegal," September 24, https://bbc.in/3HM6E6J

BIS, 2018, "Central bank digital currencies," Bank for International Settlements, Committee on Payments and Market Infrastructures Markets Committee Paper no. 174, https://bit.ly/3LDxW1y

Bordo, M. D., and A. T. Levin, 2019, "Digital cash: principles and practical steps," NBER working paper no. 25455, https://bit.ly/3LAiqmS

Cullen, J., 2021, "'Economically inefficient and legally untenable': constitutional limitations on the introduction of central bank digital currencies in the EU," Journal of Banking Regulation, Special Issue, https://bit.ly/3rQiQOg

ECB, 2020, "Report on a digital euro," European Central Bank, October, https://bit.ly/3GQDWAn

E.U., 2015, "European Parliament and Council Directive (EU) 2015/2366 of November 25, 2015, on payment services in the internal market," OJ L 337, https://bit.ly/351xeud

Fabrichnaya, E., and A. Marrow, 2022, "Russia proposes ban on use and mining of cryptocurrencies," Reuters, January 21, https://reut.rs/3GTnEqp

Federal Reserve System, 2021, "Report on the economic well-being of U.S. households in 2020 – May 2021," May, https://bit.ly/3Lyq29F

Federal Reserve System, 2022, "Money and payments: the U.S. dollar in the age of digital transformation," January, https://bit.ly/3JszPfn

Ferreira, C., 2013, "Bank market concentration and bank efficiency in the European Union: a panel Granger causality approach," International Economics and Economic Policy 10:3, 365-391

Grey, R., 2019, "Banking in a digital fiat currency regime," in Hacker, P., I. Lianos, G. Dimitropoulos, and S. Eich (eds.), Regulating blockchain: techno-social and legal challenges, Oxford University Press

H.M. Treasury, 2016, "Basic bank accounts: January to June 2016," December, https://bit.ly/3HVJqLE

House of Lords Liaison Committee, 2019, "Tackling financial exclusion: A country that works for everyone? Follow-up report," 10th Report, Session 2019-21, HL Paper 267, https://bit.ly/3LBFiT5

People's Bank of China, 2021, "Working group on E-CNY Research and Development of the People's Bank of China," July, https://bit.ly/3Ju9bD3

Ricks, M., J. Crawford, and L. Menand, 2021, "FedAccounts: digital dollars," 89 George Washington Law Review 113

Selgin, G., 2021, "Central bank digital currency as a potential source of financial instability," Cato Journal 41, 333-341

U.K. Finance, 2021, "UK payment markets 2021," June, https://bit.ly/33rFh30

WSBI, 2016, "Close to 40 million EU citizens outside banking mainstream," World Savings Banks Institute, Latest News, April 5, https://bit.ly/3HUHq67

BITCOIN'S IMPACTS ON CLIMATE AND THE ENVIRONMENT: THE CRYPTOCURRENCY'S HIGH VALUE COMES AT A HIGH COST TO THE PLANET[1]

RENEE CHO | Staff Writer, Columbia Climate School, Columbia University

ABSTRACT

Bitcoin, with a market cap of U.S.$727 billion, is the largest cryptocurrency in the world. It can be used to buy cars, furnishings, vacations, and much more. In 2011, one bitcoin was worth U.S.$1; at the time of writing this article each bitcoin is worth approximately U.S.$38,000. Because some bitcoin investors have become millionaires overnight, more and more people are intrigued by the possibility of striking it rich through investing in cryptocurrencies like bitcoin. But bitcoin's rising popularity may make it impossible for the world to stave off the worst impacts of climate change, because the energy consumption of this cryptocurrency is enormous and its environmental implications are far-reaching.

1. INTRODUCTION

In April of 2011, the price of one bitcoin was U.S.$1; last November it reached an all-time high of almost U.S.$68,000, and when this article was written, each one was worth approximately U.S.$38,000. Because some bitcoin investors have become millionaires overnight, more and more people are intrigued by the possibility of striking it rich through investing in cryptocurrencies like bitcoin. But bitcoin's rising popularity may make it impossible for the world to stave off the worst impacts of climate change, because the energy consumption of this cryptocurrency is enormous and its environmental implications are far-reaching.

To understand bitcoin's environmental impacts, we first need to know what it is and how it works.

2. WHAT IS BITCOIN?

A cryptocurrency is a virtual medium of exchange that exists only electronically; it has no physical counterpart such as a coin or dollar bill, and no money has been staked to start it. R. A. Farrokhnia, Columbia Business School Professor and Executive Director of the Columbia Fintech Initiative,[2] said, "It's a marketplace and as long as people are willing to assign value to it, then that's it." Bitcoin, the largest cryptocurrency in the world, accounting for almost half of all cryptocurrencies,[3] can be used to buy cars, furnishings, vacations, and much more. On February 21, 2022, bitcoin's market cap was U.S.$727.05 billion.[4]

Cryptocurrencies are decentralized, meaning that there is no central authority like a bank or government to regulate them.

[1] Originally published by The Columbia Climate School's State of the Planet in September, 2021.
[2] https://bit.ly/33HVUri.
[3] https://bit.ly/3l9vQEz.
[4] https://bit.ly/36xLtaT.

The advantage of this is that there are no transaction fees, anyone can use it, and it makes transactions like sending money across national borders simpler. While transactions are tracked, the people making them remain anonymous. This anonymity and lack of centralized regulation, however, means that tax evaders, criminals, and terrorists can also potentially use cryptocurrencies for nefarious purposes.

Without physical money or a central authority, cryptocurrencies had to find a way to ensure that transactions were secure and that their tokens could not be spent more than once. Bitcoin was born in 2008 when a mysterious person (or persons) named Satoshi Nakamoto (whose true identity remains unknown), found a solution to these issues. Nakamoto's answer was a digital ledger system with trust in the system achieved through mathematics and cryptography, and with transactions recorded in blockchain. Blockchain is a transparent database that is shared across a network with all transactions recorded in blocks linked together. Nodes – powerful computers connected to the other computers in the network – run the Bitcoin software and validate transactions and blocks. Each node has a copy of the entire blockchain with a history of every transaction that has been executed on it.

Nakamoto capped the number of bitcoins that could be created at 21 million. While there is speculation about the math theories that led to the choice of that number, no one really knows the reason behind it. When this article was written, an estimated 19 million bitcoins were in circulation;[5] it is expected that all remaining bitcoins will be released by 2140.

3. HOW DO BITCOINS ENTER CIRCULATION?

New bitcoins are released through mining, which is actually the process of validating and recording new transactions in the blockchain. The miner who achieves this first is rewarded with new bitcoin.

Miners must verify the validity of a number of bitcoin transactions, which are bundled into a block. This involves checking 20-30 different variables, such as address, name, timestamp, making sure senders have enough value in their accounts and that they have not already spent it, etc. Miners then compete to be the first to have their validation accepted by solving a puzzle of sorts. The puzzle involves coming up with a number – called the nonce, for "number used once" – that when combined with the data in the block and run through a specific algorithm generates a random 64-digit string of numbers and letters. This random number must be less than or equal to the 64-digit target set by the system, known as the target hash. Once the nonce that generates the target hash is found, the winning miner's new block is linked to the previous block so that all blocks are chained together. This makes the network tamper-proof because changing one block would change all subsequent blocks. The result is broadcast to the rest of the blockchain network and all nodes then update their copies of the blockchain. This validation process, or consensus mechanism, is known as proof of work. The winning miner receives newly minted bitcoin as well as the transaction fees paid by the sender.

The higher the price of bitcoin, the more miners are competing, and the harder the puzzles get. The Bitcoin protocol aims to have blocks of transactions mined every ten minutes, so if there are more miners on the network with more computing power, the probability of finding the nonce in less than ten minutes increases. The system then makes the target hash more difficult to find by adding more zeroes to the front of it; the more zeros at the front of the target hash, the lower that number is, and the harder it is to generate a random number below it. If there is less computing power operating, the system makes the puzzle easier by removing zeroes. The Bitcoin network adjusts the difficulty of mining about every two weeks to keep block production to ten minutes.

Every 210,000 blocks, the bitcoin reward for miners is halved. According to Investopedia,[6] when bitcoin was first mined in 2009, mining one block would earn 50 bitcoins. By November of 2020, the reward was 6.25 bitcoins, but as of March, 2022, the price was about U.S.$43,000 per bitcoin, so a miner would earn about U.S.$270,000 (6.25 x 43,000) for completing a block.

It is estimated that there are one million bitcoin miners operating and competing, though it is impossible to be sure because miners with less computing power of their own can join mining pools, which need not report how many active miners they have.

[5] https://bit.ly/35iIX8H.
[6] https://bit.ly/3HcMlsu.

"I have a suspicion that Nakamoto had the notion that everyone could be a miner – that you could mine with nothing more than your laptop," said Farrokhnia. "But as Bitcoin became more popular and more people got on the system and the rewards were actually worth money, you began to see the advent of these mining pools which significantly increased the difficulty level. This turned into a vicious cycle – an arms race – to have the most powerful computers, but then the more powerful hardware miners have, the more difficult it is to find the nonce."

This intense competition is where the environmental impacts of bitcoin come in.

4. BITCOIN'S ENVIRONMENTAL IMPACTS

4.1 Energy consumption and greenhouse gas emissions

The process of trying to come up with the right nonce that will generate the target hash is basically trial and error – in the manner of a thief trying random passwords to hack yours – and can take trillions of tries. With hundreds of thousands, if not more, computers churning out guesses, Bitcoin is thought to consume 707 kwH per transaction. In addition, the computers consume additional energy because they generate heat and need to be kept cool. And while it is impossible to know exactly how much electricity Bitcoin uses because different computers and cooling systems have varying levels of energy efficiency, a University of Cambridge analysis estimated that bitcoin mining consumes 121.36 terawatt hours a year.[7] This is more than what Argentina consumes, or more than the consumption of Google, Apple, Meta, and Microsoft combined.

And it is only getting worse, because miners must continually increase their computing power to compete with other miners. Moreover, because rewards are continually cut in half, to make mining financially worthwhile miners have to process more transactions or reduce the amount of electricity they use. As a result, miners need to seek out the cheapest electricity and upgrade to faster, more energy-intensive computers. Between 2015 and March of 2021, Bitcoin energy consumption increased almost 62-fold. According to Cambridge University, only 39 percent of this energy comes from renewable sources,[8] and that is mostly from hydropower, which can have harmful impacts on ecosystems and biodiversity.

In 2020, China controlled over 65 percent of the global processing power that runs the Bitcoin network; miners took advantage of its cheap electricity from hydropower and dirty coal power plants. In 2021, however, China cracked down on the cryptocurrency market and mining out of concerns about their financial risks and enormous energy consumption, which works against China's goal to be carbon neutral by 2060. As a result, many Chinese bitcoin miners are trying to move operations to other countries, like Kazakhstan, which relies mainly on fossil fuels for electricity, and the U.S. A number of U.S. states, such as Texas and Georgia, are eager to attract Chinese miners to boost their own economies. In addition, U.S. miners themselves are raising hundreds of millions of dollars to invest in bitcoin mining and converting abandoned factories and power plants into large bitcoin mining facilities.

One example of this is Greenidge Generation, a former coal power plant in Dresden, New York, which converted to natural gas and began bitcoin mining. When it became one of the largest cryptocurrency mines in the U.S., its greenhouse gas emissions increased almost ten-fold between 2019 and 2020. Greenidge plans to quadruple its mining capacity by 2022 and wants to convert more power plants to mining by 2025. While Greenidge pledged to become carbon neutral through purchasing carbon offsets,[9] the fact remains that without bitcoin mining, the plant would probably not be running at all. Other polluting peaker plants – power plants that usually only run during peak demand for a few hours a month – are being taken over for crypto mining to run 24/7.

Earth Justice and the Sierra Club sent a letter to New York State's Department of Environmental Conservation urging it to reject the renewal of Greenidge's permit that would allow it to increase its greenhouse gas emissions.[10] They also warned that there are almost 30 power plants in upstate New York that could potentially be converted to bitcoin mining operations; if this occurred, it could stymie New York State's efforts to eliminate virtually all greenhouse gas emissions by 2050.

Globally, Bitcoin's power consumption has dire implications for climate change and achieving the goals of the Paris Accord because it translates into an estimated 22 to 22.9 million metric tons of CO_2 emissions each year – equivalent to the CO_2 emissions from the energy use of 2.6 to 2.7 billion homes for one year.[11] If bitcoin grows in value – some analysts

[7] https://bbc.in/3K1gbrt.
[8] https://bit.ly/34YWaTI.
[9] https://bit.ly/3s7D6uK.
[10] https://bit.ly/3LU7bG3.

believe its price could hit U.S.$100,000 in 2022 – mining could increase, resulting in even more emissions, unless more renewable energy is used.

4.2 Water issues and e-waste

Power plants such as Greenidge also consume large amounts of water. Greenidge draws up to 139 million gallons of fresh water out of Seneca Lake each day to cool the plant and discharges it some 30° to 50°F hotter than the lake's average temperature, endangering the lake's wildlife and ecology. Its large intake pipes also suck in and kill larvae, fish, and other wildlife.

Even if it one day becomes possible to run all bitcoin mining on renewable energy, its e-waste problem remains.[12] To be competitive, miners want the most efficient hardware, capable of processing the most computations per unit of energy. This specialized hardware becomes obsolete every 1.5 years and cannot be reprogrammed to do anything else. It is estimated that the Bitcoin network generates 11.5 kilotons of e-waste each year, adding to our already huge e-waste problem.[13]

5. NFTS

A new phenomenon – NFTs – has added to the environmental concerns about cryptocurrencies. These are non-fungible tokens – digital files of photos, music, videos, or other kinds of artwork stamped with unique strings of code. People can view or copy NFTs, but there is only one unique NFT that belongs to the buyer and is stored on the blockchain and secured with the same energy-intensive proof of work process. NFTs are selling for hundreds of thousands of dollars; Beeple, a digital artist, sold one NFT for more than U.S.$69 million.

Ethereum, the second most popular cryptocurrency after bitcoin, creates the NFTs. The average NFT generates 440 pounds of carbon – the equivalent of driving 500 miles in a gas-powered car – producing emissions 10 times higher than the average Ethereum transaction. One digital artist estimated that the carbon footprint of an average NFT is equivalent to more than an E.U. resident's electricity consumption for a month. Some artists, concerned about NFTs' environmental impacts, are trying to raise awareness and look for more sustainable ways of creating them.

6. HOW CAN CRYPTOCURRENCIES BE MORE SUSTAINABLE?

Because the entire Bitcoin network has invested millions of dollars in hardware and infrastructure, it would be difficult for it to transition to a more energy efficient system, especially since there is no central oversight body. However, there are a number of projects seeking to reduce the carbon footprint of Bitcoin, and cryptocurrencies in general. Tesla CEO, Elon Musk, met with the CEOs of top North American crypto mining companies about their energy use. The upshot was the creation of a new Bitcoin Mining Council to promote energy transparency.[14]

The Crypto Currency Accord is another initiative,[15] with over 250 supporters, whose goal is making blockchains run on 100 percent renewable energy by 2025 and having the entire cryptocurrency industry achieve net zero emissions by 2040. It aims to decarbonize blockchains through using more energy efficient validation methods, pushing for proof of work systems to be situated in areas where excess renewable energy can be tapped, and encouraging the purchase of certificates to support renewable energy generators, much like carbon offsets support green projects.

Ethereum is aiming to reduce its energy use by 99.95 percent by 2022 through transitioning to an alternative validation system called proof of stake, as a few smaller cryptocurrencies have done. Proof of stake does not require computational power to solve puzzles for the right to verify transactions. Rather, it works like a lottery. To be considered, potential validators stake their ethereum coins (ETH); the more they stake, the greater their chances of being selected randomly by the system to be the validator. Participants will have to stake 32 ETH (each was worth about U.S.$2600 when this article was written) per validator opportunity, with multiples of 32 ETH for more chances. After a new block is accepted as accurate, validators will be rewarded with coins and keep the coins they staked. The system ensures security because if validators cheat or accept false transactions in the block, they lose their stake and are banned from the network. When the price of ETH rises, stakes become more valuable, and thus network security increases, but the energy demands remain constant. Some worry, however, that proof of stake could give people with the most ETH more power, leading to a less decentralized system.

[11] https://bit.ly/3s86xgj.
[12] https://bit.ly/3v72xOZ.
[13] https://bit.ly/34W1sPv.
[14] https://bit.ly/3sUdctT.
[15] https://bit.ly/3sY5r67.

"Blockchain is a highly customizable and flexible technology," said Farrokhnia. "You could design it in any shape or form that meets your objective. So, for example, another proof of consensus mechanism is called proof of reputation: the more reputable you are, the more votes you have in validating things." The proof of authority system relies on reputation and trustworthiness; blocks and transactions are verified by pre-approved participants who must reveal their true identities. A few cryptocurrencies use proof of coverage that requires miners to provide a service – for example, hosting a router in their home to expand the network.

Other ideas for greening cryptocurrencies involve moving bitcoin operations next to oil fields, where they tap waste methane gas that is usually flared, pipe it to generators, and use the power for bitcoin mining. Some bitcoin mining is moving to west Texas where wind power is abundant. Because there is sometimes more wind power than transmission lines can handle, bitcoin mining situated near wind farms can use their excess energy.

Farrokhnia said that while these ideas are theoretically possible, they may not be pragmatic. "Each of these ideas requires very high upfront capital expenditures," he said. "And we know that interest in mining is predicated on the price of bitcoin itself, so you could have all sorts of truly expensive solutions that would aim to be more energy efficient, but as soon as the price of bitcoin were to drop below a certain threshold, all these projects would be [canceled] because they're just not financially feasible. Who in reality would make those investments given the volatility in price of bitcoin and the uncertainty about the future of it?"

Farrokhnia's hope for greener cryptocurrency lies in its evolution. He believes that cryptocurrencies cannot ignore environmental considerations if they want to gain wider adoption, and that newer and greener cryptocurrencies will eventually eclipse bitcoin.

"There's a new generation of crypto coming on board," Farrokhnia said.[16] "They are going to move away from proof of work for a number of reasons, one of which is the environmental impact, because most of these are being created by young programmers. They're certainly more environmentally conscious, and hopefully, they understand the impact of the work beyond whatever they're building and will take into account the complexity of today's world."

7. CONCLUSION

It is essential that climate and environmental implications are considered and managed as cryptocurrencies gain in usage around the world. Today, we are on track for global temperatures to rise between 2.5°C and 4.5°C,[17] which could result in catastrophic impacts. It is challenging enough to reduce the greenhouse gas emissions we currently generate – the world cannot afford to add to them.

[16] https://bit.ly/3H8Tsrs.

[17] https://go.nasa.gov/3sbGFAr.

THE EVILS OF CRYPTOCURRENCIES

JACK CLARK FRANCIS | Professor of Economics and Finance, Bernard Baruch College
JOEL RENTZLER | Professor of Economics and Finance, Bernard Baruch College

ABSTRACT

In 2021 Gary Gensler, Chairman of the Securities and Exchange Commission, called cryptocurrency markets the "Wild West" and said they are rife with "fraud, scams, and abuse" [Talley and Volz (2021), Kiernan (2021), CBS News (2021)]. One of the main reasons they cause so many problems is that the U.S. has no laws governing cryptocurrencies. Since cryptocurrencies do not conform to the legal definition of securities, the existing U.S. securities laws do not apply to them. As a result, a complicated multi-billion dollar lawless industry has sprung up in the U.S. in recent decades.

1. INTRODUCTION

Berensten and Schar (2018a) prepared an introduction to cryptocurrencies, which are often called cryptos or digital currencies. Over 10,000 cryptocurrencies were listed on the CoinMarketCap.com website in 2022, and that number keeps increasing. Speculators, investors, and criminals that seek to operate confidentially are attracted to cryptocurrencies.[1] Cryptocurrencies are also of interest to central bankers because some people use them as if they were money. In 2021, Gary Gensler, Chairman of the SEC, called cryptocurrency markets the "Wild West" and said they are rife with "fraud, scams, and abuse" [Talley and Volz (2021), Kiernan (2021), CBS News (2021)]. Popper (2019) reports other cryptocurrency problems. One concern with cryptocurrencies that continues to arise is: are cryptocurrencies money?

1.1 What is money?

Money is:

1) a unit of account
2) a medium of exchange, and
3) widely acceptable as a form of payment.

To be useful, money should also be fungible, difficult to counterfeit, and easily transportable. Cryptocurrencies fail to meet most of these criteria. For example, cryptocurrencies are not a good medium of exchange because commercial banks in the U.S. and China are not allowed to accept them as deposits or execute transactions that involve cryptocurrencies. U.S. banks can allow their depositors to store cryptocurrencies in the bank's safe deposit box for rental fees, the same way a bank might allow depositors to store gold or silver in its safe deposit box. The Federal Reserve treats silver, gold, and cryptocurrencies as illiquid commodities, not as money [Vigna 2019b)].

[1] For more about the criminal activities see Lahart and Demos (2021), Osipovich (2021), Vigna (2019b), Popper (2019), Vigna and Jeonmg (2019), Hirtenstein (2021), Talley and Volz (2021), Yaffe-Bellany (2022b), and others.

The U.S. dollar, British pound, the euro, and Japanese yen are established fiat currencies that are not backed by any collateral. Fiat money has value because a sovereign government declares it to be legal tender that can be used to make full and final payment of legal debts. The only currency the U.S. government has designated to be legal tender is the U.S. dollar. Cryptocurrencies are not qualified to be used as a fiat currency in the U.S. and, thus, no cryptocurrency should be called U.S. money. This does not mean that other nation's governments, such as El Salvador in 2021, cannot designate cryptocurrencies to be legal tender in their country.

1.2 Virtual currencies

In 2012, the European Central Bank defined a virtual currency to be an unregulated, digital money that is issued and controlled by its developers. In other words, virtual currencies are used and accepted among the members of a virtual community. In 2013, the U.S. Treasury Department stipulated that a virtual currency is something that operates like a currency in some environments but does not have all the attributes of a real currency. These definitions both describe cryptocurrencies. Some writers have discouraged the U.S. from using cryptocurrencies [Hirtenstein (2021)]. Nevertheless, many private individuals invest in cryptocurrencies in the U.S.

2. BITCOINS

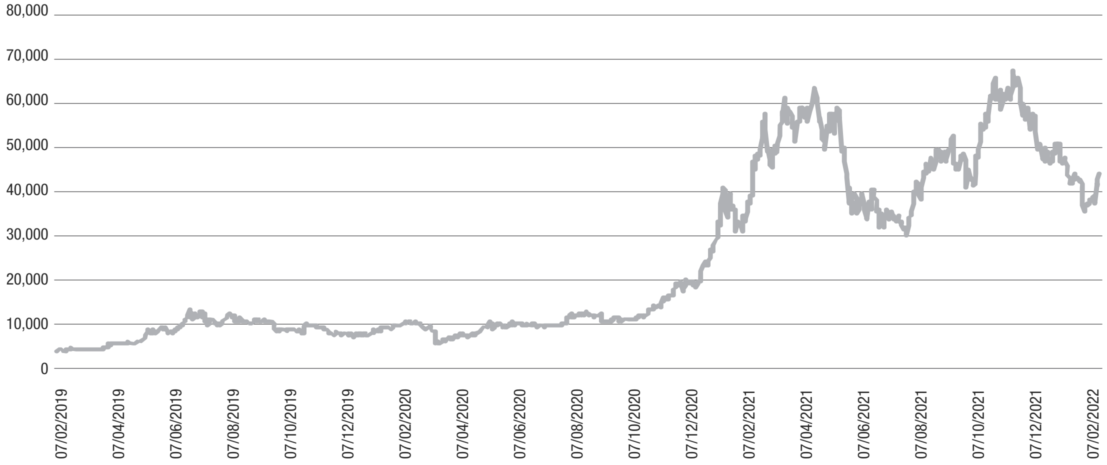
Satoshi Nakamoto introduced the first cryptocurrency in the world, bitcoin, in 2009.[2]

2.1 Characteristics of bitcoins

Brokers, traders, and exchanges wishing to trade bitcoins will find them listed and traded at approximately 300 cryptocurrency dealers in the U.S. In addition, the Chicago Mercantile Exchange (CME) and the Chicago Board of Options Exchange (CBOE) started selling bitcoin futures and options in 2017. Furthermore, a number of financial executives want to create and distribute exchange traded funds (ETFs) to introduce amateur investors to crypto assets. Financial regulators are reluctant to legitimize cryptocurrencies because most of them have serious issues [as discussed by Osipovich (2019) and Vigna (2019a, b), for example].

Several financial economists have noted that a major complaint against cryptocurrencies is that their market prices fluctuate randomly and sometimes excessively [Jain et al. (2021), Hu et al. (2019)]. For instance, the real sector of the U.S. economy experienced a two-month mini-recession spanning February and March of 2020. Figure 1 shows that the prices of bitcoins acknowledged this mini-recession by experiencing only one single once-and-for-all price drop from March 11th to March 12th. Furthermore, from February 2021

Figure 1: Daily bitcoin prices (U.S.$)



Source: Federal Reserve Economic Database

[2] Nakamoto (2009) introduced bitcoins and the blockchain database.

through February 2022 bitcoin prices rose to peak prices and then fell drastically twice for no apparent reason. As various researchers have suggested, it appears that bitcoin prices fluctuate randomly rather than fluctuate closely with (are highly positively correlated with) business activity in the U.S. In addition, Griffin and Shams (2020), Kroeger and Sarkar (2017), and Makorov and Schoar (2021), also reported that bitcoin prices sometimes generate profitable arbitrage opportunities by simultaneously trading at different prices in different markets. Thus, the markets in which cryptocurrencies are traded are not highly efficient.

Howell et al. (2019) investigated another common problem that plagues cryptocurrency buyers; their new investment becomes illiquid not long after it is introduced at its initial coin offering (ICO). Numerous issues of cryptocurrencies appreciate in the first few days after their ICO, but after a few months buyers can no longer be found. Since most cryptocurrencies have no assets, income, or collateral, rational buyers lose interest quickly.

2.2. The Bitcoin blockchain

Bitcoin is an international decentralized digital virtual currency that works without a central bank, financial intermediary, or other third party to handle its transactions. Every transaction is verified in an electronic network of nodes using cryptographic records that are maintained in a publicly distributed electronic ledger called the Bitcoin blockchain. The Bitcoin blockchain is shared, replicated, and re-finalized every time a bitcoin transaction occurs; this process results in what bitcoin users like to call a "continuous consensus" among the blockchain users. This "continuous consensus" does not prevent millions of other people, the U.S. government, and the Chinese federal government from disapproving of the use of bitcoins and other cryptocurrencies.

Bitcoin computer technicians called miners compete to validate every bitcoin transaction. Miners are paid 6.25 new bitcoins for the proof of work they provide by verifying transactions in the Bitcoin blockchain. In addition to 6.25 bitcoins for their proof of work, the miners can also receive a negotiated "transaction fee" from the bitcoin buyer [it should be noted that the blockchain itself has value apart from bitcoin, see Francis (2019)].

A problem with the Bitcoin blockchain is that it can handle a maximum of only about seven transactions per second. No cryptocurrency comes close to the 50,000 transactions per second that VISA credit card handles routinely. This scaling problem poses one of the fundamental limitations on the growth of cryptocurrencies [Andolfano (2018), Vigna (2018a)].

2.3 Halving

When bitcoins were created in 2009, Satoshi Nakamoto stipulated that a total of no more than 21 million bitcoins could ever be issued [Nakamoto (2009)]. Also in 2009, Satoshi Nakamoto declared that each time another 210,000 blocks of bitcoin were mined, the block reward given to bitcoin miners for validating transactions would be cut in half. These halvings took place in 2013, 2017, and 2021 – about every four years. Each halving is significant because it marks a significant drop in the total remaining supply of bitcoins. On May 11, 2021, after halving reduced the block reward to 6.25 bitcoins, approximately 18,715,050 million bitcoins had been released into circulation. Thus, on May 11, 2020, only about 2,284,950 million bitcoins remained to become future mining rewards.

When the number of unissued bitcoins falls to zero, if all the owners of bitcoins in existence can agree, it is theoretically possible to renegotiate a new bitcoin mining protocol. But, perhaps, it will be quicker, cheaper, and easier for bitcoin owners to abandon bitcoin and shift their activities to a more user-friendly investment. This potential internal problem is evidence that the oldest and most popular cryptocurrency in the world is built on a shaky foundation.

2.4 Competing cryptocurrencies

Every cryptocurrency is a decentralized autonomous organization (DAO). Each cryptocurrency's DAO operates according to the rules written into the computer program that handles its transactions.[3] The different DAOs are designed to compete against each other in order to maintain and augment their customer list. About half the DAOs traded in the U.S. are bitcoins; no other cryptocurrency is as popular as bitcoin.

A cryptocurrency named Ethereum has a DAO that handles "smart contracts". The smart contract permits transactions to advance in Ethereum only after certain conditions are fulfilled. Decisions made by Ethereum buyers can generate different

[3] Decentralized finance (DeFi) applications permit users to lend, borrow, earn interest, trade assets, and perform other transactions with various decentralized autonomous organizations (DAOs). See Vigna (2021a) and Hirtenstein (2021). DeFi is a newer, more complex system than cryptocurrencies [Economist (2021c)].

price paths for the cryptocurrency. Ethereum was launched in 2015; it is a system that resembles Bitcoin in two respects. First, Ethereum has its own unique blockchain. Second, it has miners that create Ethereum's cryptocurrency, which is called ether. Ethereum's blockchain miners are paid in ethers.

Ethereum is more complicated than bitcoin. For example, when a shipment of widgets is delivered, the recipient can be notified. This notification of arrival can activate the recipient's (widget buyer's) computer to send the appropriate payment to the widget seller. Most older computer systems were less flexible, they could only communicate simpler messages between the members of their uniquely established list of counterparties. In contrast, Ethereum allows new and different counterparties in its blockchain to transact. Furthermore, Ethereum permits unrelated parties to interact. Ethereum can also transfer money between wallets after a specific event is completed. This additional flexibility can result in new business transactions. The market capitalization of Bitcoin greatly exceeds that of Ethereum, but at this time the volume of transactions at Ethereum is increasing more rapidly than the growth in Bitcoin. Ethereum has become central in the budding field of decentralized finance (DeFi), where smart contracts make sophisticated decisions, such as whether or not to make a loan without the aid of a human decision-maker [Economist (2021a)].

Ripple XRP is a blockchain-based digital payment network that has its own cryptocurrency, named XRP. Instead of using blockchain mining, Ripple uses a consensus-gaining mechanism installed in a group of bank-owned servers to confirm transactions between the bank's clients. Ripple provides a system for making direct transfers of financial assets.

Binance Coin, Cardano, Dogecoin, Litecoin, Tron, Monero, NEO, and IOTA are other cryptocurrencies that are currently traded actively. As mentioned above, all are significantly less popular than bitcoin and ether.

3. CRYPTOCURRENCY MARKETS

The price of a bitcoin went from a penny for a single coin in 2009 to over U.S.$68,000 per coin in November 2021. Gains like these stimulate interest from buyers who have FOMO (fear of missing out). But bitcoin's price does not always rise smoothly. Unfortunately, bitcoin's price fell rapidly from U.S.$68,000 in November 2021 to U.S.$35,000 in January 2022. During that period the number of bitcoin trades per day ranged from a high of 56 million per day when bitcoin's price was near U.S.$68,000 to a low of around 21 million per day when bitcoin's price was down to U.S.$35,000 in January 2022. In spite of these wild gyrations, the market for bitcoins is flourishing [Easley et al. (2019)]. Bitcoin is the oldest and by far the most frequently traded cryptocurrency in the U.S., it is listed and traded at many cryptocurrency exchanges. Some of the most noteworthy cryptocurrency exchanges are discussed below.

3.1 Cryptocurrency exchanges

Some of the most noteworthy cryptocurrency exchanges are:

Binance: the largest cryptocurrency exchange in the world is named Binance. The founder of Binance, Changpeng Zhao, says this international company has no headquarters office because, in his opinion, having a corporate headquarters is "an antiquated concept." Unfortunately, for those who have a problem with Binance, this ambiguity could complicate settling their claim [Ostroff et al. (2022), Kowsmann and Ostroff (2021)].

Gemini: as discussed later, the Gemini Trust is a candidate to be the most ethical cryptocurrency exchange in the world. Gemini is a small operation in New York City that was founded and managed by identical twins named Cameron and Tyler Winklevoss. The brothers sued Mark Zuckerberg, founder of Facebook, in 2004, claiming he stole their ConnectU idea to create the popular social networking site Facebook.

Coinbase: on April 14, 2021, the Coinbase Exchange became the first cryptocurrency exchange in the U.S. to have its shares listed on NASDAQ (ticker: COIN). Shares in Coinbase began trading at U.S.$328 per share, but that price was reduced to U.S.$274 by May 4, 2021, as investors investigated Coinbase. First, the founder of Coinbase and one of his executives were inappropriately assigned to be on the corporation's internal Audit Committee [Eaglesham (2021)]. In other words, the managers were auditing themselves. Second, some Coinbase investors reported unfair losses and at least one filed fraud charges against the exchange [Browning (2021)].

The Coinbase Exchange typically charges its retail traders several transactions fees. If you buy U.S.$1,000 worth of a cryptocurrency in the U.S., for instance, you will pay a flat fee of 2.99 percent, or $29.90. In addition, Coinbase adds a fee of one-half of one percent to the transaction to bring the sub-total to 3.49 percent, or U.S.$34.90. Furthermore, if you pay with a credit card a one percent fee, or U.S.$10 will be added to bring the total fee to 4.49 percent, or U.S.$44.90. If you both buy and sell a cryptocurrency, Coinbase will double the one-way fee to U.S.$89.80 for your "round-trip" transaction. These transaction fees are much higher than the NYSE's fees for trading shares of stock, but lower than the small, privately owned cryptocurrency dealers would charge.

Other centralized cryptocurrency exchanges: BlockFi, Crypto.com, eToro, Kraken, and Robinhood Crypto are the names of other large cryptocurrency exchanges in the U.S. In 2020, these centralized cryptocurrency exchanges began facing new competition from the first decentralized cryptocurrency exchange.

DEXs emerge: decentralized cryptocurrency exchanges are often called DEXs. Decentralized exchanges enable users to buy and sell cryptocurrencies without the aid of a commission-hungry broker acting as their middleman. The crypto buyer and seller begin to transact by simply connecting their crypto wallets to a DEX, which temporarily acts as their middleman. Then the traders select the crypto they want to trade and enter the amount they wish to transact. The DEX endeavors to bring together two like-minded traders to consummate the transaction.

Uniswap: the first DEX is named Uniswap, it was created in 2020 by 27-year-old Hayden Adams while he was between jobs [Osipovich (2021a)]. Uniswap provides a way for computers to talk to each other. No central bank or other third party decides who will be allowed to trade or what tokens may be traded. Furthermore, DEXs do not require their traders to give their digital tokens to the DEX before they can trade. Traders find this last feature appealing because they worry about losing their digital tokens when they trade through the system of loosely managed private cryptocurrency exchanges used by many traders. Coinbase sometimes uses a DEX to act as an interface with public cryptocurrency traders.

Coinbase provides user-friendly websites but, behind the scenes, Uniswap sometimes performs the trading for some cryptocurrency dealers [Vuillemey (2020)].

3.2 Large central banks

The Federal Reserve, or, the Fed, is the monetary authority for the U.S. It controls the U.S. money supply, interest rates, inflation, and the credit markets. The Fed also routinely clears thousands of checks per day from banks around the world. These checks are all cleared through the Federal Reserve Bank of Cleveland.

The Fed does not allow any U.S. banks to accept deposits or execute transactions involving silver, gold, cryptocurrencies, or any other commodities. As explained above, the Federal Reserve treats silver, gold, and cryptocurrencies as illiquid commodities, not as a type of money. In October 2021, the Fed did permit MasterCard credit card company and a Georgia-based digital wallet company named Bakkt Holdings Inc. to join forces to create a cryptocurrency credit card in the U.S. [Andriotis (2021)].

The Peoples Bank of China (PBOC) is the Chinese government's monetary authority. In 2021, the PBOC outlawed all cryptocurrencies and any activities related to cryptocurrencies. Most businesses in the U.S. and China will not accept cryptocurrencies as payment. The Russian government has opposed cryptocurrencies for years, saying it could be used in money laundering or to finance terrorism. The Russian government gave cryptocurrencies legal status in 2020 but banned their use for payments. It is difficult to see how cryptocurrencies will ever become as liquid as some suggest as long as large industrialized nations like the U.S., China, and Russia continue to place crippling restrictions on their transactions.

4. THE HISTORY OF STABLECOINS

Stablecoins are cryptocurrencies with prices that are pegged to a cryptocurrency, fiat money, or to an exchange traded commodity like silver or gold. The advantage of these asset-backed cryptocurrencies is that the stablecoin prices are stabilized by their connections to assets that have prices which fluctuate outside of the cryptocurrency space. These uncorrelated connections are supposed to reduce the financial risk of the stablecoins. The disadvantage of stablecoins is that their underlying collateral is typically inadequate or simply nonexistent.

The first stablecoin in the U.S., nubits, was introduced in 2014. Initially, Nubits (ticker: NBT) was considered safe and its prices hardly fluctuated. People thought every nubits was safe because stories and advertisements said every nubits was fully collateralized by one U.S. dollar. Unfortunately, a reputable audit never verified that nubits had any collateral. As a result, investors' trust faded away and a nubits with a face value of one U.S. dollar was selling for 32 cents in early 2022.

If a cryptocurrency has a U.S. dollar backing that can be authenticated by a respectable auditor, that cryptocurrency will have a market price that fluctuates in a narrow range around the value of its collateral. But, if the existence of a cryptocurrency's collateral cannot be confirmed, then the collateral backing is questionable, and the cryptocurrency's market price will fluctuate towards zero. Almost all cryptocurrencies that were ever introduced in the U.S. and are not still actively traded today have market values of zero. This worthless and inactive segment of the population of cryptocurrencies equals the vast majority of all U.S. cryptocurrencies.

Not all collateralized cryptocurrencies are backed by U.S. dollars. Some claim to be backed by other fiat monies, actively traded commodities like gold or silver, cryptocurrencies, or ingenious computer software that is claimed to promote the price of the stablecoin.

4.1 Tether

Tether (ticker: USDT) is the most popular blockchain-based stablecoin. Tether's initial coin offering (ICO) was a crowdfunding issue of tokens that was managed by its issuer, Tether Limited. Tether Limited claimed that every tether token was backed by one U.S. dollar [Ackermann et al. (2020)]. At the same time, Tether Limited announced that tether buyers had no contractual rights to their underlying collateral of U.S. dollars. Surprisingly, this latter declaration did not cause the initial market price of tethers to plunge below U.S.$1. The market price of tethers remained very close to U.S.$1 for about a year after they were issued. But, in 2016 the market price of the tether began to wobble. More recently tether's market price fluctuated between U.S.$1.06 and 92 cents [Lyons and Viswanath-Natraj (2020)]. Tether Limited could have probably kept the market price of the stablecoin closer to U.S.$1 if it had opened its books to a public audit that verified Tether was fully collateralized by the promised number of U.S. dollars. Tether Limited provided no such public assurance. To make the situation more tenuous, in March 2019 Tether Limited announced that it was expanding tether's collateral beyond the U.S. dollar to also include loans to affiliate companies, a much riskier type of collateral than the U.S. dollar. Furthermore, on April 30, 2019, the cryptocurrency market was disappointed when Tether Limited announced that each tether was backed by only 74 cents in cash and cash equivalents, less than was promised at Tether's ICO. In other words, if Tether went bankrupt when it was backed by only 74 cents worth of U.S. dollars, the tether owner could expect to receive less than 74 cents for each tether.[4] Investing in tether is obviously much riskier than investing in U.S. dollars.

Law-abiding, risk-averse businesspeople have little incentive to buy tethers or other stablecoins. The features that motivate people to buy stablecoins are the absence of binding laws, the absence of records, and the complete lack of transparency that exists throughout the stablecoin market. As a result of these features, the so-called "privacy" that exists throughout the cryptocurrency markets makes them a particularly attractive place for swindlers and criminals to conduct business. In an early attempt toward regulation, in October 2021, the U.S. Commodity Futures Trading Commission (CFTC) fined Tether U.S.$41 million for misrepresenting itself to be fully backed by assets during 2016 and 2019.[5] Since SEC Chairman Gary Gensler stated publicly in 2021 that the U.S. cryptocurrency markets are rife with "fraud, scams, and abuse" [CBS News (2021), Talley and Volz (2021), Kiernan (2021)], the likelihood that the SEC will issue similar fines in the future increased.

4.2 Problems with Tether

An additional problem for Tether Limited occurred in 2018-2019. Two different cryptocurrency exchanges named Bitfinex and Tether Limited were both owned and operated by iFinex Inc. Although tether was issued in the U.S., iFinex is headquartered in Hong Kong and registered in the British Virgin Islands. Many people were confused by these complicated international arrangements. In 2019, the New York Attorney General's office alleged that in mid-2018 Bitfinex lost U.S.$850 million and secretly used funds taken from Tether to cover the resulting shortfall. This well-documented allegation further tainted Tether's reputation [Griffin and Shams (2019) Vigna (2019a), Ostroff (2021), Michaels (2021)]. Customers' money has been stolen or lost in several incidents and, as a result, like the other uncollateralized cryptocurrencies, Bitfinex, Tether, and iFinex have never been permitted to transact with any U.S. commercial bank.

While Tether Limited's reputation is not spotless, tethers have much better collateral than uncollateralized cryptocurrencies like bitcoin and ether. In spite of anything they might advertise, most, or all, of the uncollateralized cryptocurrencies in the U.S. have zero collateral. It appears that the founders of thousands

[4] In February 2021, Tether was fined U.S.$18.5 million by New York State for overstating the size of its of collateral backing of U.S. dollars [Economist (2021b)].

[5] Unfortunately, the U.S. has not yet enacted any laws regulating cryptocurrencies to furnish a legal basis for this fine.

of uncollateralized cryptocurrencies disappeared with the cash proceeds from their initial coin offering (ICO) soon after the ICO was completed. Such frauds occur repetitively in the U.S. because the Securities and Exchange Commission (SEC), the Federal Reserve, the Comptroller of the Currency, the U.S. Treasury, the Federal Bureau of Investigation (FBI), or some other arm of the U.S. federal government have never officially recognized that cryptocurrencies are securities. As mentioned above, the Federal Reserve treats cryptocurrencies as if they were commodities, not legal tender. Since cryptocurrencies are not legally defined to be securities, they cannot be regulated under the existing U.S. securities laws [Macintosh (2021), Smialek (2021)].

4.3 Gemini Trust

An entertaining and informative 2010 movie named The Social Network reported the interesting 2004 campus activities of several Harvard undergraduates. The focus of the movie is Mark Zuckerberg's formation of Facebook and his interactions with the identical twins named Cameron and Tyler Winklevoss (portrayed by actors) when they were all Harvard students. The twins are of interest here because in 2015 they founded and still run a New York City cryptocurrency broker-dealer firm named Gemini Trust.

Gemini Trust brokers and deals in selected cryptocurrencies. In addition to making markets in various cryptocurrencies, the Gemini Trust also issues a stablecoin of its own called the Gemini dollar. The Gemini Trust seems to operate at a higher level of security and professionalism than most other cryptocurrency exchanges. Gemini complies with both the New York state and the U.S. digital asset regulations and consumer protection laws. As a result of these legal operating standards, Gemini Trust was able to obtain Federal Deposit Insurance Corporation (FDIC) insurance for the U.S. dollar (but not Gemini dollar) accounts of its clients.[6]

The collateral for the Gemini dollar is kept with a highly reputable third party, the State Street Trust in Chicago. Unlike Tether, both Gemini Trust and State Street Trust have audited financial statements to confirm that Gemini's collateral actually exists. The Gemini dollar runs on an Ethereum-based blockchain system. Unlike many other cryptocurrency operations, the Gemini Trust appears to manage the Gemini dollar ethically [Jain et al. (2019)]. As a result, the market price of the Gemini dollar floats within the narrow range between U.S.$1 and $1.06. No U.S. commercial bank is allowed to accept Gemini dollar deposits. Thus, Gemini dollars are no more liquid than bitcoins or ethers.

Although tether is financially riskier than the fully collateralized Gemini dollar, tether is much more well-known within the cryptocurrency markets. One reason for this popularity is because tether enjoys a first-mover's advantage. Tether is an older stablecoin and its average daily trading volume of several billion exceeds the volume of the Gemini dollar. As a result, more cryptocurrency buyers are accustomed to dealing with the tether than with the Gemini dollar. In addition, many tether users are probably unaware of Tether Limited's unsavory past.

4.4. Three types of stablecoins

In this section, we describe three categories of stablecoins, each of which define their collateral differently.

4.4.1 STABLECOINS COLLATERALIZED BY FIAT MONEY

Many stablecoins claim to be backed by a fiat currency. The safest fiat-backed cryptocurrencies are collateralized by the U.S. dollar. Some of the most popular stablecoins that claim to be collateralized by U.S. dollars are tether (USDT), U.S. Dollar Coin (USDC), TrueUSD (TUSD), StableUSD, Dai (DAI), and the Gemini dollar (GUSD). Unfortunately, the Gemini dollar (GUSD) seems to be the only stablecoin that has verified the existence of some appropriate amount of collateral by allowing itself to be audited by a respectable auditor. Similar claims by competing stablecoins are unaudited and, as a result, are highly dubious.

4.4.2 CRYPTOCURRENCIES COLLATERALIZED BY OTHER CRYPTOCURRENCIES

Each BitShares coin, issued by BitUSD, claims to be worth one U.S. dollar. In fact, the BitUSD has at least 100 percent of its own outstanding cryptocurrency backed by BitShares core currency, BTS. This circular relationship provides no valuable collateral.

Havven issues nomin, which is a stablecoin backed by a portfolio of havvens. The value of havvens comes from transaction fees generated from nomin transactions that are paid into the portfolio of havvens. The value of nomins is supposed to be kept stable by the havven owners, who are supposed to be incentivized to manage the supply of

[6] The FDIC is an independent federal agency of the U.S. government that insures U.S. dollar deposits in U.S. banks and thrift institutions if the bank fails. FDIC insurance covers checking and savings accounts, CDs, money market accounts, IRAs, trust accounts, and employee benefit plans up to a maximum of U.S.$250,000 per client. But the FDIC does not insure any cryptocurrency deposits.

havvens propitiously [Brooks et al. (2018)]. These claims have not been verified by a reputable auditor and are, therefore, suspect. Essentially, all cryptocurrencies collateralized by cryptocurrencies that have no fundamental value have no fundamental value either.

4.4.3 STABLECOINS COLLATERALIZED BY SOFTWARE

Instead of being backed by monetary collateral, some uncollateralized stablecoins are backed by a computer algorithm that makes dubious claims that it can execute transactions that will stabilize the stablecoins price fluctuations.

- **Basis:** after a short run, Basis shut down in December 2018. The market price of basis was supposed to be kept stable by a trading algorithm. When demand was rising, the blockchain was supposed to create more basis. This expanded supply was supposed to meet the rising demand and reduce the rising price. When demand for basis was falling, the blockchain was supposed to buy basis. The resulting contraction in supply was supposed to bid up the market price of basis [Reuters (2018)]. These claims were never demonstrated.
- **Carbon:** Carbon is supposed to operate like Basis. Carbon uses an algorithm named Hedera Hashgraph, which is significantly faster than the system used by Basis. The passage of time should reveal if the cryptocurrency market finds any value in Carbon.
- **USDVault:** the USDVault stablecoin is pegged one-to-one to the U.S. dollar. It is supposed to be backed by either gold bullion stored in Swiss vaults or U.S. dollars. USDVault takes a novel approach to maintaining stability. This stablecoin is supposed to stay gold-price neutral while maintaining a one-to-one peg to the U.S. dollar through a sophisticated gold hedging process that is administered by fiduciaries and financial partners. Since the ambitious claims supporting the USDVault stablecoin have not been supported by a respectable auditor they are considered to be doubtful.

The founders of many stablecoins and almost all cryptocurrencies do not provide any clear, audited proof that their collateral exists. Examination of whatever plans may be provided reveals they are exaggerated and unclear.

Various writers have evaluated the characteristics of cryptocurrencies [Cheun and Guo (2018)]. One favorable feature that cryptocurrencies provide is a structure that facilitates the creation of non-fungible tokens (NFTs) [Ostroff (2021a)]. However, few people have any desire to invest in this tiny new market segment. A second favorable feature is wealth creation. Some speculators have quickly accumulated significant wealth from a cryptocurrency and they like to brag about that. Numerous losers that do not brag about their outcomes also exist.

5. EVIL ASPECTS OF CRYPTOCURRENCIES

In its original form, the bitcoin is an ingenious concept. But some unethical developers have reconstituted cryptocurrencies in ways that are harmful. Unfortunately, the U.S. government has been painfully slow in regulating the development of cryptocurrencies.[7]

5.1 Inadequate collateral

The Gemini dollar appears more likely than the average stablecoin to actually be worth its face value. Tethers appear to have some value, but their collateral is likely to be worth significantly less than their face value. The majority of other stablecoins are worth significantly less than they claim too. And some stablecoins are worthless.

All cryptocurrencies that are not stablecoins are totally void of collateral. Popular cryptocurrencies like bitcoin achieve and maintain their positive market prices because many investors have FOMO (fear of missing out) and, as explained in the next paragraph, some bitcoin traders are unethical.

Schoar and Makarov (2021) have recently mapped and analyzed every transaction in bitcoin's 13 years of transactions and reported that only 0.01 percent of the bitcoin owners own 27 percent of the outstanding bitcoins [Vigna (2021b)]. Concentrated holdings like this might be called cornering the market, which is illegal in regulated commodity and security markets in the U.S. However, since the U.S. securities laws have not yet been applied to the cryptocurrencies, these unseemly behaviors in the crypto markets continue to go unpunished. Furthermore, Schoar and Makarov (2021) report that about 90 percent of all bitcoin transactions have no actual economic function that can be determined from the publicly available bitcoin transactions records. These research findings are troublesome.

5.2 A fundamentally flawed governance plan

A fundamental problem underlying stablecoins involves bad governance. Consider the fact that if a private party issues stablecoins and is responsible for providing collateral for these stablecoins, that manager has continual economic incentives to under-collateralize the stablecoins. In addition, the stablecoin issuer has an incentive to invest the collateral in risky assets that have higher expected returns. Thus, stablecoins are a fundamentally flawed, unstable arrangements that encourage some bad management practices.

5.3 "Privacy" attracts criminals

Cryptocurrencies offer "privacy" that criminals find essential for survival. For example, if a kidnapper, computer ransomware seeker, tax cheat, divorce settlement cheater, bank robber, or other criminal wants to hide U.S. dollars obtained illegally, the criminal's "dirty money" could easily be hidden by investing it in a cryptocurrency that keeps their identity private [for criminal aspects of cryptocurrency trading, see Lahart and Demos (2021), Osipovich (2021b), Vigna (2019b), Popper (2019), Vigna and Jeonmg (2019), Hirtenstein (2021), Yaffe-Bellany (2022b), Talley and Volz (2021)]. Income from cryptocurrencies is taxable in the U.S. But the U.S. government cannot collect taxes on "private" transactions if it cannot discover them.

5.4 Money records provide a valuable memory

The privacy of cryptocurrency transactions conceals much valuable information that should be made legally available to beneficiaries, creditors and other interested third parties.[8] Law-abiding citizens prefer to transact in U.S. dollars because the U.S. check clearing system and related paper trails provide valuable information for police, regulators, and other interested third parties. In addition, a frequently cited research by Narayana Kocherlakota, former President of the Federal Reserve Bank of Minneapolis and a former Stanford University professor, argues that using U.S. dollars creates an audit trail [Kocherlakota (1998)]. This paper trail of money transactions provides a valuable chronological history of accessible records that can be used to determine causes and effects if a legal dispute or criminal investigation arises.

5.5 Significant environmental damage

As explained above, bitcoins, ethers, and many other cryptocurrencies are based on blockchain technologies that employ miners to verify every transaction. The mining process that accompanies most cryptocurrency transactions uses a massive amount of electricity to power the computers that verify every transaction. Electric generators that burn oxygen and create carbon dioxide are used to generate much of the electricity. The Cambridge Center for Alternative Finance (CCAF), for example, estimates the amount of electricity used by cryptocurrency miners to process only the world's bitcoin transactions has a market value approximately equal to the "energy draw of small countries like Malaysia or Sweden." [Carter (2021), Ostroff and Yu (2021)] When the aggregate electrical cost of simply verifying and reverifying the world's

[7] F. A. Hayek, a Nobel Prize winning economist, argued in favor of competing national currencies. He did not address most of the problems discussed herein.

[8] Regulators have succeeded in uncovering crooked cryptocurrency transactions. For information about a recent U.S.$2.3 million illegal bitcoin transaction that was uncovered and corrected see Volz et al. (2021). But most crooked transactions are not identified.

bitcoin transactions is considered, the negative impact that bitcoin mining has on the world's climate is troubling [Huang et al. (2021)].

5.6 Centralized mining

Bitcoins, ethers, and many other cryptocurrencies are based on a blockchain system that requires verification and reverification of every transaction by a computer called a “miner” each time another cryptocurrency transaction occurs. Thousands of specialized computers (miners) compete to finish each verification process first and win the reward of 6.25 bitcoins, which has a current market value of roughly U.S.$250,000 at current market prices. Over the years, this mining competition has evolved to the point where a few big “pools” of computers do most of the mining. The costs of this mining have become so high that only a small group of large firms can afford to do it [Economist (2022)]. These mining operations tend to centralize their locations in a few spots around the world where large amounts of electricity can be purchased cheaply. In 2022, these mining operations became such a problem in Russia that the government passed a law making cryptocurrency mining in Russia illegal. Some of the Russian bitcoin miners are now in the process of moving to Rockdale, Texas (population 5,600), where electricity is cheap and plentiful and the mayor welcomes new cryptocurrency businesses.

Two bitcoin mining firms named Bitdeer and Riot Blockchain are currently Rockdale’s only miners, and they are both growing as fast as they can. While additional new cryptocurrency miners relocate to Rockdale, Riot Blockchain currently claims to be the largest so far with 100,000 computers on site. The electric grid for the state of Texas is deregulated and has ample power to sell, which makes electricity inexpensive in Texas. The state of Texas is a strong candidate to becoming the new central headquarters of the world’s cryptocurrency mining industry.

5.7 Facilitating criminal activity

If a criminal has a large amount of cash or cryptocurrency in their electronic wallet, they can transfer this “dirty money” long distances to a recipient electronic wallet about as quickly and confidentially as the Federal Reserve could conduct a similar wire transfer between the bank accounts of non-criminals. The existence of cryptocurrencies and these developments in the cryptocurrency industry engenders criminal activity by facilitating electronic transfers of “dirty money.”

5.8 Law-abiding investors are scared away

Satoshi Nakamoto, who created the bitcoin in 2009, said that bitcoin was created to allow anyone to open a digital bank account and hold digital money in a way that no government could regulate [Nakamoto (2009)]. This innocent sounding goal overlooks some inconvenient realities. Actually, the privacy surrounding cryptocurrencies may entice criminals and scare away law-abiding investors who would prefer to have transparent transactions that generate paper trails that can be audited and policed.

Most cryptocurrency exchanges are only modest websites that sprung up in someone’s home during 2016-2017. Some of these cryptocurrency exchanges have lost millions of dollars of their clients’ money. For example, the following losses have been reported by cryptocurrency exchanges:

- Youbit lost U.S.$35 million in 2017
- DAO lost U.S.$55 million in 2016
- Bitfinex lost U.S.$77 million in 2017
- BitGrail lost U.S.$170 million in 2018
- Mt. Gox lost U.S.$450 million in 2014
- Coincheck lost U.S.$534 million in 2018.

More recently, in February 2022, the U.S. government seized U.S.$3.6 billion of cryptocurrency linked to the U.S.$4.5 billion 2016 hack of the cryptocurrency exchange, Bitfinex.[9] A married couple was arrested in this huge financial seizure. This arrest suggests law enforcement officers are sometimes able to recover stolen cryptocurrency. A Deputy Attorney General said the authorities captured the married couple by following the stolen funds as they were deposited and withdrawn in rapid succession while jumping between multiple forms of virtual currency exchanges and dark markets. When Satoshi Nakamoto was designing bitcoins in 2009 terrible events like these were probably never imagined.

5.9 Misleading transaction prices

Through no fault of the researchers, some of the cryptocurrency trades that researchers study and report do not always involve actual trades. For example, consider a hypothetical small sample of empirical data that has a mean daily return that is calculated from a sample of 132 large, frequently traded cryptocurrencies. If this sample of 132 selected observations

[9] As discussed above in Section 4.2, in 2018-2019 two different cryptocurrency exchanges named Bitfinex and Tether Limited were both owned and operated by iFinex Inc. Tether (ticker: USDT) is a stablecoin that has become popular in spite of being inadequately collateralized.

happens to equal a small percent of the total population of cryptocurrencies, the sample mean statistic is very likely to be an unrepresentative estimate of the underlying population parameter. Fake transactions have also been reported [Vigna (2019b)].

In the more fool-proof research methodology mentioned above, Schoar and Makarov (2021) analyzed every transaction in Bitcoin's 13 years of transactions and reported that only 0.01 percent of the bitcoin owners own 27 percent of the outstanding bitcoins [Makarov and Schoar (2021), Vigna (2019b)]. Schoar and Makarov (2021) also report that about 90 percent of all bitcoin transactions have no actual economic function that can be determined from the publicly available bitcoin transactions records. These facts suggest that a few large bitcoin traders could be in a position that would make it possible for them to profit from manipulating bitcoin prices.

5.10 Essential governmental functions

Milton Friedman (1960) made a statement that is still relevant today. "Something like a moderately stable monetary framework seems an essential prerequisite for the effective operation of a private market economy. It is dubious that the market can by itself provide such a framework. Hence, the function of providing one is an essential governmental function on a par with provision of a stable legal framework."

6. CONCLUSION

Cryptocurrencies, as explained in Section 1, have only a small resemblance to the popular fiat currencies. Since they do not attempt to duplicate any of the popular fiat currencies, cryptocurrencies cannot be called counterfeit currencies. The two aspects of cryptocurrencies that make them unique are, first, they are a medium of exchange that operates through a computer network and second, that they are not reliant on a central authority. These are the qualities that were innocently stressed by Satoshi Nakamoto when the bitcoin was introduced in 2009.

Since 2009, unethically greedy people and criminals have, unfortunately, reconstituted and adapted bitcoins in ways Satoshi Nakamoto might not appreciate. Although bitcoins are still the predominant cryptocurrency, over 10,000 other cryptocurrencies have been developed in the U.S. The developers of over 90 percent of these newer cryptocurrencies took the proceeds from their initial coin offering (ICO) and disappeared. Several get-rich-quick stories are told and retold while numerous losers remain silent.[10]

[10] For one of the few accounts of losses from cryptocurrency trading, see Yaffe-Bellany (2022a). The last half of the article discusses various problems cryptocurrency traders encounter.

REFERENCES

Ackermann, E., C. Bock, and R. Bürger, 2020, "Democratising entrepreneurial finance: the impact of crowdfunding and initial coin offerings (ICOs)," in Moritz A., J. Block, S. Golla, and A. Werner (eds.), Contemporary developments in entrepreneurial finance. FGF studies in small business and entrepreneurship, Springer

Alderman, A., 2018, "Sweden's push to get rid of cash has some saying, 'not so fast'," New York Times, November 21, https://nyti.ms/3LXXDdj

Alderman, L., 2019, "Despite bitcoin's dive, a former Soviet Republic is still betting big on it," New York Times, January 22, https://nyti.ms/3hbpatF

Andolfano, D., 2018, "Blockchain: What it is, what it does and why you probably don't need one," Review, Federal Reserve Bank of St. Louis 100, 2, 87-95

Andriotis, A. M., 2021, "Crypto payment coming to credit cards," Wall Street Journal, October 26, p. B1

Berensten, A., and F. Schar, 2018a, "A short introduction to the world of cryptocurrencies," Review, Federal Reserve Bank of St. Louis 100:1, 1-16

Berensten, A., and F. Schar, 2018b, "The case for central bank electronic money and the non-case for central bank cryptocurrencies," Review, Federal Reserve Bank of St. Louis 100:2, 97-106

Brooks, S., A. Jurisevic, M. Spain, and K. Warwick, 2018, Havven: a decentralised payment network and stablecoin, Version 0.8

Browning, K., 2021, "Coinbase users got hacked. Then they got stonewalled," New York Times, March 27, p. B7

Carter, N., 2021, "How much energy does bitcoin annually consume?" Harvard Business Review, May 5, pp. 1- 6

CBS News, 2021, "Crypto market "rife with fraud, scams and abuse," SEC chief says," August 3, https://cbsn.ws/3LXqdLL

Cheun, D. K. K., and L. Guo, 2018, "A new investment opportunity?" Journal of Alternative Investments, 20:3, 16-40

Eaglesham, J., 2021, "Coinbase insiders get roles guarding investors, Wall Street Journal, April 6, 2021, page B1

Easley, D., M. O'Hara, and S. Basu, 2019, "From mining to markets: the evolution of bitcoin transaction fees," Journal of Financial Economics 134:1, 91-109

Economist, 2021a, "The future of Banking," Special report, May 8, https://econ.st/36sFUKC

Economist, 2021b, "Beating bitcoin: cryptocoins are proliferating wildly. What are they all for?" June 12, https://econ.st/3LWKUYg

Economist, 2021c, "Decentralized finance," September 18-24, pp. 17-20

Economist, 2022, "Build block better: is a greener, faster and more decentralised alternative to Bitcoin possible?" January 1, https://econ.st/3pe7TEu

Francis, J. C., 2019, "Bitcoin, blockchain and cryptocurrencies," Journal of Financial Transformation, 49, 8-21

Friedman, M., 1960, A program for monetary stability, Fordham University Press

Griffin, J. M., and A. Shams, 2019, "Is bitcoin really un-tethered?," Journal of Finance 75:4, 1913-1964

Hayek, F. A., 1976, Denationalism of money, The Institute of Economic Affairs

Hirtenstein, A., 2021, "Hackers steal and return crypto," Wall Street Journal, August 12, pp. B1 and B4

Howell, S., M. Niessner, and D. Yermack, 2019, "Initial coin offerings: financing growth with cryptocurrency sales," Review of Financial Studies 33, 3925-3974

Hu, A. S., C. A. Parlour, and U. Rajan, 2019, "Cryptocurrencies: stylized facts on a new investible instrument," Financial Management 48:4, 1049-1068

Huang, J., C. O'Neill and H. Tabuchi, 2021, "Bitcoin uses more electricity than many countries. How is that possible?" New York Times, September 3, https://nyti.ms/3hgoYJs

Jain, P. K., T. H. McInish, and J. L. Miller, 2019, "Insights From bitcoin trading," Financial Management 48:4, 1049-1068

Kiernan, P., 2021, "Crypto executives defend industry as Congress considers oversight," Wall Street Journal, December 8, https://on.wsj.com/3Ilt5Qn

Kocherlakota, N., 1998, "Money is memory," Journal of Economic Theory 81:2, 232-251

Kowsmann, P., and C. Ostroff, 2021, "Traders want bitcoin losses back," Wall Street Journal, July 12, pp. 1-2

Kroeger, and A. Sarkar, 2017, "The law of one price?" Federal Reserve Bank of Philadelphia, https://bit.ly/3vEVVHX

Kuhn, T. S., 1962, The structure of scientific revolutions, University of Chicago Press

Lahart, J., and T. Demos, 2021, "How crime is haunting bitcoin, Wall Street Journal, June 19-20, p. B14

Lyons, R. K., and G. Viswanath-Natraj, 2020, "What keeps stablecoins stable?" NBER working paper no. 27136

Macintosh, J., 2021, "Stable coins hark back to wild west of finance," Wall Street Journal, 28 May, pp. B1 and B11

Makarov, I., and A. Schoar, 2021, "Blockchain analysis of the bitcoin market," NBER working paper no. 29396

Michaels, D., 2021, "Digital-coin exchange penalized $100 million," Wall Street Journal, August 11, pp. B10-11

Nakamoto, S., 2009, "Bitcoin: a peer-to-peer electronic cash system," Nakamoto introduced bitcoins and the blockchain database, https://bit.ly/33MtTPg

Osipovich, A., 2021a, "Peer trading rises in crypto sector," Wall Street Journal, May 25, p. B11

Osipovich, A., 2021b, "Crypto scams target newbies, market pros," Wall Street Journal, June 8, pp. B1 & B11

Ostroff, C., 2021a, "The cats that created the NFT explosion," Wall Street Journal, May 8-9, p. B8

Ostroff, C., 2021b, "Tether lays out details about assets," Wall Street Journal, August 11, pp. B10-B11

Ostroff, C., and E. Yu, 2021, "Cryptocurrency miners move out of China," Wall Street Journal, August 23, pp. B3-B4

Ostroff, C., P. Kowsmann, and D. Michaels, 2022, "SEC probes Binance's US Arm, affiliates of crypto exchange," Wall Street Journal, February 16, pp. A1-A2

Popper, N., 2019, "Terrorists are turning to bitcoin to raise funds discretely," New York Times, August 19, pp. B1 and B2.

Popper, N., and C. Li, 2021 "China is charging ahead with its digital currency," New York Times, March 2, pp. B1 and B5

Rappeport, A., and J. Smialek, 2020, "Treasury to introduce rules on cryptocurrencies," New York Times, February 13, p. B6

Reuters, 2018, "Cryptocurrency project Basis to shut down and return funding to investors," December 13, https://reut.rs/3pfL7w6

Schoar, A., and I. Makarov, 2021, "Bitcoin's ultra-elite dominate its wealth," Wall Street Journal, December 21, pp. B1 and B4

Smialek, J., 2021, "In a rush to rein in stablecoins," New York Times, September 18, pp. B1 and B4

Talley, I., and D. Volz, 2021, "U.S. to target crypto use in random cyberattacks," Wall Street Journal, September 18, pp. A1 and A4

Vigna, P., 2019a, "Two groups likely stole a billion dollars of crypto," Wall Street Journal, January 29, pp. B1-B2

Vigna, P., 2019b, "Most bitcoin trading is faked, study finds," Wall Street Journal, March 23-24, p. B13

Vigna, P., and E.-Y. Jeonmg, 2019, "Cryptocurrency scams took in $4 billion in 2019," Wall Street Journal, February 10, p. B4

Vigna, P., 2021a, "DeFi adds to crypto boom and volatility, Wall Street Journal, June 4, pp. B1 and B2

Vigna, P., 2021b, "Bitcoin's 'one percent' controls lion's share of the cryptocurrency's wealth," Wall Street Journal, December 20, https://on.wsj.com/3BLxpWG

Volz, D., S. Gurman, and D. Uberti, 2021, "Pipeline ransom money seized by US," Wall Street Journal, June 8, p. A4

Vuillemey, G., 2020, "The value of central clearing," Journal of Finance 75:4, 2021-2053

Yaffe-Bellany, D., 2022a, "Coin boom is drawing new traders and tokens," New York Times, February 7, pp. B1 and B3

Yaffe-Bellany, D., 2022b, "No real names in crypto, please," New York Times, March 3, pp. B1 and B6

AT LAST A REALLY SOCIALLY USEFUL STABLECOIN: SNUT (THE SPECIALIZED NATIONAL UTILITY TOKEN)

STEPHEN CASTELL | Founder and CEO, Castell Consulting[1]

ABSTRACT

The market price of a cryptocurrency – which, as a medium of financial exchange, generally has scarcity built into it, but little, if any, demonstrable economic utility – is driven and influenced principally by what its buyers and sellers believe its market price should, or will, be, i.e., by speculation. This article introduces the QE2-Coin, a U.K. central bank digital currency (CBDC), originally proposed in 2017, that is, first, inherently designed not to be driven by speculative pressures, i.e., is a stablecoin, and, secondly, is specifically engineered to have utility as a SNUT, a "specialized national utility token" – deliberately architected to be exchangeable for products, services, goods, and assets in the real world, in particular, in the affordable homes housebuilding sector. Throughout the post-WW2 decades, despite many political manifesto pledges for reform and repeated central government attempts at encouragement of the home construction industry, there has in reality been a constant and growing new affordable U.K. homes blight, characterized by woefully under-target new housebuilding and poorly executed government stimuli. Without new ideas, innovation, and a powerful vision, it seems unlikely that any U.K. government policy will evolve to rectify this situation and be able to narrow the growing gap between U.K. housing supply and housing need. The really socially useful and valuable stablecoin, the QE2-Coin, to be spent in the U.K. housebuilding sector economy, and not converted into any inert non-economically productive asset or instrument, will positively address these homes availability issues, fueling economic activity in the U.K. housebuilding sector specifically focused on providing affordable homes. Uniquely, the QE2-Coin is a "limited life utility token", meaning that it will have a smart contract baked into it, with functionality coded to "dematerialize" any QE2-Coin token instance, taking it out of existence if it does not get spent and used socially usefully within a defined time period. A QE2-Coin maquette has been created as the basic Ethereum crypto-token **QE2**. The U.K. Prime Minister's response to the SNUT proposal is awaited. However, it is undeniable that there is a severe affordable starter and rental homes shortfall in the U.K. and, whether or not the U.K. government decides to engage with SNUT, there is no reason why the QE2-Coin initiative should not proceed. Visionaries and entrepreneurs in fintech, the crypto community, the investment world, and the property sector are welcomed to join in developing the QE2 SNUT plan for fixing this starter and rental homes shortage.

1. INTRODUCTION: VACUITY VERSUS UTILITY

The dynamic factors affecting the price of any cryptocurrency, in particular bitcoin, have long been the subject of ongoing explanation and discussion on websites and blogs, and across social media platforms [Bloomenthal (2022), Castell (2021a), Haar (2021), Pierce (2022)]. A number of potential factors has been suggested, however, most analysts agree that the clear, constant, and overriding factor that is the common driver of the trading of any cryptocurrency is speculation. That is, the market price of a cryptocurrency – which, as a medium of financial exchange, generally has scarcity built into it, but little,

[1] Dr Stephen Castell, stephen@castellconsulting.com. I would like to thank Fernando Martinho, cybersecurity and cryptocurrency specialist, founder of https://www.unicorn.win/, for his assistance in the creation of the **QE2** ECR-20 Ethereum token maquette.

if any, demonstrable economic utility – is driven and influenced principally by what its buyers and sellers believe – hope – its market price should or will be (many analysts would add "and by little, if anything, else").

Castell (2021b) suggests that bitcoin, in particular, inherently exhibits this sad lack of economic utility. It exists only to exist: it has a fundamental existential vacuity, giving rise to the heartfelt lament that "Bitcoin itself" expressed in that article: "Can you find a way to save me and my cryptocurrency species, vacuously and pointlessly existing, commercially and legally dangerous, operating outside the Rule of Law? Is there not someone, who, understanding me and the miraculousness which I symbolize, can establish, on a firm commercial and regulatory footing, with rigorous operational and legal reliability, with solid trust and transparency, a truly viable and robust new crypto-economy – one of course driven by the Invisible Hand! Alas, I suspect not; I, Bitcoin, bereft of practical utility, may one day no longer exist – no longer even simply to exist. I will become just another abandoned human artefact, like the spinning jenny, the bearer share, the analogue television, the telex, the video recorder, the junk bond, the fax machine, the non-digital mobile phone, the non-electric, non-autonomous road vehicle..."

Take away this speculative cryptocurrency characteristic, this existential vacuity, and what of real social, human, economic value is left? Nothing minus nothing equals not much! Conversely, however, create a cryptocurrency that is, first, inherently designed not to be driven by speculative pressures, i.e., is a stablecoin, and is, second, specifically engineered to have utility – deliberately architected to be exchangeable for products, services, goods, and assets in the real world – et voilà, the crypto vacuum that nature abhors is productively filled.

In this article I introduce and describe the QE2-Coin, a cryptocurrency that is not only a stablecoin but is intrinsically configured to have practicality as a SNUT: a "specialized national utility token".

2. STABLECOINS

There is much available information and widespread discussion about stablecoins. Lipton et al. (2020), for example, state that: "What first started as a niche phenomenon within the cryptocurrency community has now reached the realms of multinational conglomerates, policy makers, and central banks. From JP Morgan's Jamie Dimon to Facebook's Mark Zuckerberg, stablecoins have made their way onto the agenda of today's top CEOs. As projects like Libra have enjoyed broad media coverage they are also increasingly scrutinized by regulatory authorities. And as the term "stablecoin" spread, its meaning started to blur. This is problematic. An unclear definition may make us susceptible to deceptive innovation, that is, reintroducing existing services but in a different appearance."

Fry (2021), examining "Why stablecoins are not just important for the crypto market," noted that: "To date, much of the focus and use cases for stablecoins has been to view them as a payment mechanism, as a way to … avoid the high transaction fees often associated with international remittances"; and ended by posing the question: "are stablecoins creating much of the infrastructure and processes and use cases which will make the adoption of Central Bank Digital Currencies (CBDC) all the smoother and faster?"

In January 2022, a report from the U.S. Board of Governors of the Federal Reserve on the benefits and risks presented by a potential U.S. central bank digital currency (CBDC) gave a fairly clear "No" answer to that latter question, as far as the U.S. is concerned [Fed (2022)]. It seems that the U.S. Federal Reserve is far from launching a CBDC at all, let alone "smoother and faster", and it has issued a discussion paper that examines the pros and cons of a potential U.S. CBDC.

A CBDC is generally defined as a digital liability of a central bank that is widely available to the general public. In the U.S., Federal Reserve notes (physical U.S. dollar) are currently the only type of central bank money available to the general public. Like existing forms of such money, a CBDC would enable the general public to make digital payments. A CBDC would be the safest digital asset available to the general public, with no associated credit or liquidity risk.

Despite, so far, the lack of reception for CBDCs in the U.S., I firmly believe that they have useful capabilities, if structured correctly, to benefit society at large. An example is discussed below.

3. FIXING THE U.K. HOUSING SHORTAGE: THE QE2-COIN U.K. CBDC STABLECOIN PROPOSAL AND THE SNUT TASK FORCE PLAN

With regards to a possible U.K. CBDC, in October 2017, I requested my Member of Parliament, the Rt Hon Priti Patel MP (currently UK Home Secretary), to put forward a Prime Minister's Question (PMQ) in Parliament proposing my unique idea and recommendation for the U.K.'s own national stablecoin, the QE2-Coin: "Will the PM seize the international economic high ground for the U.K. in regard to the dramatic evolution of the field of cryptocurrencies, and announce that Britain will issue the first ever state-backed sovereign state initial coin offering (SSICO), the "QE2-Coin", to be used specifically to fix the U.K. housing shortage, the Government granting tranches of QE2-Coins to local councils, with a mandate to employ them vigorously to secure a rapid expansion in supply of badly-needed new starter and rental homes throughout Britain; and will the PM confirm that the Government understands that this world-first SSICO, putting billions of QE2-Coins into circulation applied productively towards the worthy objective of increasing the stock of affordable modern homes for the UK, will be a true 'magic money tree', non-inflationary, and not at all affecting the PSBR" [Castell (2019)].

I have subsequently put practical flesh on this idea, as described herein. I understand from my MP that these further details have reached the desk of the Rt Hon Boris Johnson MP, the U.K. Prime Minister. His response is awaited.

In brief, the QE2-Coin is to be a unique "digicoin": not simply a stablecoin, nor an asset-backed token, nor a fiat currency, nor a financial/investment instrument, but a "specialized national utility token", or SNUT.

It will have the following features:

- To be issued by and in the name of the people by the SNUT People's Trust Fund.
- Each QE2-Coin to be guaranteed as to its base exchange value, or BEV, by the Bank of England (or by the SNUT People's Trust Fund, well-capitalized and insured).
- The BEV to be: one QE2-Coin (QE2) will never be worth less than 0.5 GBP (i.e., 50 pence); thus, 1 GBP = 2 QE2 maximum.
- 20 billion QE2 to be issued through an unregulated/regulated ICO, or other workable digital currency mechanism (compliant with MiCA and FCA etc. provisions, as appropriate), i.e., 10 billion GBP fiat equivalent initially.
- The QE2 to be used solely within the housebuilding sector for new affordable homes, i.e., in the materials, labor, equipment, construction, fixtures and fittings, furnishings, services, utilities, realty agents, mortgage financing etc., home supply, and value chains.
- The U.K. Government and HM Treasury to support the QE2 vision and objectives of the SNUT People's Trust Fund by inter alia permitting taxes to be paid in QE2.

The initial scope and outline work plan of the QE2-Coin SNUT task force is proposed as:

1. Preliminary convening and defining of SNUT outline terms of reference.
2. Creating the core management team under Dr Stephen Castell.
3. Configuring the SNUT QE2 work groups: funding, SNUT People's Trust Fund, technical and operational requirements, design, creation, validation, ICO planning, ICO implementation, regulatory and legal, government, executive and legislative, construction industry liaison, roll-out, operational guidance, monitoring, and oversight and accountability.
4. Development of project, action, and management plans.

There will be regulatory and legal issues to be addressed, not least regarding the U.K. Government's passing a law making the QE2 legal tender:

(i) for payment for employment, goods, and services in the U.K. housebuilding sector (carefully defined in the legislation); and

(ii) prohibiting the QE2 to be purely deposited in banks, or invested in markets: "It's for spending; not for keeping, collecting, lending, borrowing, saving, or investing."

Discussion with companies in the U.K. housebuilding sector (having a stock market valuation of around £50 billion) suggests that there will be an enthusiastic welcome for this affordable homes SNUT. A leading real estate consultancy, with over 50 years' experience in London's commercial property, said "This QE2-Coin is a brilliant idea, and when you consider the appetite for a continuing post-Covid 'work from home' economy, it could readily be adapted to implement a newly re-imagined commercial workplace-home property sector, too."

There is much already available in the literature addressing, explaining, and discussing the micro-economics of U.K. housebuilding. It is not the purpose or place of this introductory article to go into how the QE2-Coin fits into any micro-economics model of the sector, and I leave that to others and/or for another day [Thangavelu (2021), Ball (2003)].

4. AFFORDABLE AND STARTER HOMES FOR U.K. CITIZENS: THE SOCIALLY USEFUL CRYPTO-ECONOMICS OF THE QE2-COIN

There is no doubt that social housing and the affordable and starter homes sectors in the U.K. are in dire need of attention, assistance, and improvement. Young people starting out, the geographically, economically, and socially disadvantaged, and the poorly-paid, least skilled, and less financially capable: these groups of U.K. inhabitants have, for generations after the post-war generation, been faced with a paucity of quality and fit-for-purpose housing choices, at affordable prices, either for rent or for purchase. That most basic of human needs, an appropriate, sound, well-built, safe, secure, healthy, and comfortable home, has never been universally and satisfactorily met for U.K. inhabitants.

The U.K.'s Housing Minister, Michael Gove MP (currently holding the government posts of Minister for Intergovernmental Relations and Secretary of State for Levelling Up, Housing and Communities), has himself recently declared that the quality of some social housing in Britain is "scandalously poor" [Barker (2021)].

It appears that fulfilling this most fundamental need of anyone for "a home, and a reasonably good one, at an affordable cost" may be unachievable without new thinking. There have been 18 U.K. Housing Ministers since 1997 and none has managed to establish any initiative, policy, or workable implemented plan to fix the affordable and starter homes shortage; that does not auger well for the U.K. Government's current ambition for circa 300,000 new homes per annum, as set out in its recent revised National Planning Policy [IH Reporters (2020)].

Without new ideas, innovation, and a powerful vision, it seems unlikely that any U.K. government policy will evolve to rectify this situation, and be able to narrow the significant, and routinely growing, gap between U.K. housing supply and housing need – let alone come near fully meeting that need. According to Grissell and Kerley (2021): "Across the country, there are more than 27 million homes, but many more are needed. In the 30 years to 2021, three million fewer properties were built than in the previous 30. The population, however, has increased by more than nine million." Recent research by the House of Commons reports that "In order to reach the Government's target of 300,000 new homes per year, annual net supply would need to reach levels 39 percent higher than in 2020/21" [Wilson and Barton (2022)].

In summary, throughout the post-WW2 decades, despite many political manifesto pledges for reform and repeated central government attempts at encouragement of the home construction industry, there has in reality been a constant and growing new affordable U.K. homes blight, characterized by woefully under-target new housebuilding and poorly executed government stimuli. These have been negatively complemented by continuing archaic and unwieldy productive land use regulation, overly-protective zoning restrictions, and tediously long, convoluted, and uncertain planning permission processes.

Consistent national land management and oversight, with a substantive vision for homing U.K. inhabitants, and a coherent, effective implementation strategy for the future, have been sorely lacking, whatever the party in government. Land use and home construction day-to-day controls and restrictions continue to be essentially devolved to a multiplicity of local authorities, most having little capability or resources to execute a "citizens quality homing vision". At the same time, these local authorities are charged with the responsibility for finding locally disadvantaged and homeless people whatever dwellings, of whatever quality, that may be available, in severely limited quantities, at great cost to the public purse, from almost exclusively private sector housing suppliers and owners.

What is more, this dearth in availability, exacerbated by the rapidly increasing costs, of any suitable stock of private sector dwellings are factors currently creating near-crisis difficulties and challenges for U.K. local public servants expected to grapple with homing the homeless. The latest U.K. house price figures from property portal Rightmove reveals "the biggest monthly jump in pounds ... recorded in its 20 years of data-gathering" with "the highest annual rate of growth since September 2014" [Michael and Howard (2022)].

The really socially useful and valuable stablecoin, the QE2-Coin, to be spent in the U.K. housebuilding sector economy, and not converted into any inert non-economically productive asset or instrument, will positively address these U.K. homes availability and cost factors, fueling economic activity, livelihoods, and growth, enhancing optimism, confidence, jobs, and profits in the UK housebuilding sector specifically focused on providing affordable homes. Furthermore, with the right vision and implementation thereof, this could well promote and provide smart, carbon-neutral, or even carbon-positive, homes as well [Richardson and Coley (2019)].

5. THE QE2-COIN: TECHNICAL DETAILS

Uniquely, the QE2-Coin is a "limited life utility token", meaning that it will have a smart contract baked into it, with functionality coded to "dematerialize" any QE2-Coin token instance, taking it out of existence if it does not get spent and used socially usefully within a defined time period.

For purposes of manifestation, a maquette or prototype QE2-Coin has been created as the basic Ethereum crypto-token QE2. As QE2 evolves, details will be dynamically available at www.QE2Coin.com. Its initial technical configuration may be examined at Etherscan (Kovan Testnet Network).[2]

6. NEXT STEPS

To anyone, like the author, who has been active in financial market systems and capital instruments and products innovation (fintech) for a considerable length of time, it has always seemed odd that bitcoin sought only to mimic existing "traditional" currency, payment, and asset concepts. Leaving aside the technical innovation of utilizing a blockchain consensus mechanism and cryptographic security architecture for creation and anti-counterfeiting, bitcoin and other cryptocurrencies are generally rather quaintly "financial and economic old school" in business application concepts and implementation. In 1995, long before bitcoin, when I conceived my own "electronic cash unit" (ECU) [Castell (1995)], I always had in mind using the untrammeled "whatever you want it to be" imaginative software-coded processes of computer systems-based digital cash to embed novel functionality and utility "within the very 'digicoin' itself."

One imaginative and innovative element of this inherently flexible functional capability of computer systems-based digital cash, not present, nor able to be implanted, in bitcoin or any other standard blockchain-architected cryptocurrency, can be a baked-in algorithm, a smart contract, that makes a digicoin's existence "intrinsically time-dependent" and this embedded "use it or lose it" algorithm design will be utilized for the QE2-Coin.[3] "It's for spending; not for keeping, collecting, lending, borrowing, saving or investing." Utility, not vacuity!

[2] https://bit.ly/3hkxtU2

[3] Possible outline algorithm stub:

```
EXISTENCE: IF DEMATERIALSE FLAG (QE2-COIN (N)) EQ 'FALSE' THEN GOTO OUT
  WRITE QE2-COIN (N) TO BLOCK (BLOCK-COUNT)
  TIME-COUNT= TIME-COUNT+1
             IF TIME-COUNT > LIFE-SET AND WALLET-SPEND (QE2-COIN (N)) EQ 'FALSE'
             THEN GOTO KILL
                    PROCESS UTILITY (QE2-COIN (N)) RETURN
                    BLOCK-COUNT=BLOCK-COUNT+1
                    GOTO EXISTENCE
  KILL: SET DEMATERIALSE FLAG (QE2-COIN (N))
OUT:
```

It is interesting that the recent “Executive Order on Ensuring Responsible Development of Digital Assets” signed by the President of the United States includes in the definition of a “stablecoin” mechanisms “algorithmically controlling supply in response to changes in demand” [Biden (2022)].

The U.K. Prime Minister’s response to the SNUT Proposal is awaited. However, it is undeniable that there is a severe and persistent affordable starter and rental homes shortage in the U.K., and, whether or not the U.K. government decides to engage with SNUT, there is no reason why the QE2-Coin initiative should not proceed. Young people starting out, the geographically, economically, and socially disadvantaged, and the poorly-paid, least skilled, and less financially capable should undeniably be given the chance to receive the home-provision benefits of the financial transformation that the uniquely imaginative QE2-Coin crypto-economics will deliver, irrespective of the incomprehension, inaction, or incapability of the U.K. government.

Visionaries and entrepreneurs in FinTech, the crypto community, the investment world, and the property sector are welcomed to understand, share, and support the vision, power, and advantages of the innovation of the “QE2-Coin, at last a really socially useful stablecoin”, and join in developing, evolving, and implementing the SNUT Plan for fixing the U.K. starter and rental homes shortage.

REFERENCES

Ball, M., 2003, “Markets and the structure of the housebuilding industry: an international perspective,” Urban Housing and Property Markets 40:5/6, 897-916

Biden Jr, J., 2022, “Executive Order on Ensuring Responsible Development of Digital Assets”, The White House, March 9, https://bit.ly/3vVXFgi

Barker, N., 2021, “Quality of social housing ‘scandalously poor’ says Gove,” Inside Housing, October 4, https://bit.ly/3vwwCYq

Bloomenthal, A., 2022, “What determines the price of 1 Bitcoin?” Investopedia, February 26, https://www.investopedia.com/tech/what-determines-value-1-bitcoin/

Castell, S., 1995, “What the ECU stands for,” Computing, July 20

Castell, S., 2019, “A UK central bank digital currency (CBDC): a fresh new BOECoin, a digital version of Sterling, or something else entirely? Governor Carney may be open to the role a Digital currency may fulfil,” Digital Bytes, November 20, 6-10

Castell, S., 2021a, “Slaying the crypto dragons: towards a CryptoSure trust model for crypto-economics,” in Patnaik, S., T. S. Wang, T. Shen, and S. K. Panigrahi (eds.), Blockchain technology and innovations in business processes, Springer

Castell, S., 2021b, ““I, Bitcoin”: as told to Stephen Castell,” The World Financial Review, June 16, https://bit.ly/3C1VZmf

Fed, 2022, “Federal Reserve Board releases discussion paper that examines pros and cons of a potential U.S. central bank digital currency (CBDC),” January 20, https://bit.ly/35BujbF

Ferreira, A., 2021, “The curious case of stablecoins – balancing risks and rewards?” Journal of International Economic Law 24:4, 755–778

Fry, J., 2021, “Why stablecoins are not just important for the crypto market,” Digital Bytes, February 17, https://bit.ly/3JZmE62

Grissell, L., and P. Kerley, 2021, “In need of modernisation? The UK’s housing crisis close up,” BBC News October 1, https://bbc.in/3lwQfnc

Grobys, K., J. Junttila, J. W.Kolari, and N. Sapkota, 2021, “On the stability of stablecoins,” Journal of Empirical Finance 64, 207-223

Haar, R., 2021, “Why do bitcoins have value?” Time, September 7, https://bit.ly/3vpgmIN

IH Reporters, 2020, “A timeline of the 18 housing ministers since 1997,” Inside Housing, February 2, https://bit.ly/3JYScZR

Lipton, A., A. Sardon, F. Schär, and C. Schüpbach, 2020, “Stablecoins, digital currency, and the future of money,” Work in Progress, MIT Press, April 30, https://bit.ly/3BYn0a9

Michael, A., and L. Howard, 2022, “UK house prices – latest news,” Forbes, February 21, https://bit.ly/3t8ahO4

Pierce, H. M., 2022, “Statement on settlement with BlockFi Lending LLC,” February 14, U.S. Securities and Exchange Commission, https://bit.ly/3434fGg

Richardson, J., and D. Coley, 2019, “Labour’s pledge for carbon neutral homes will require a revolution—but it’s possible,” Quartz, November 7, https://bit.ly/3hkQFkB

Thangavelu, P., 2021, “How microeconomics affects everyday life: renting an apartment,” Investopedia, July 26, https://bit.ly/3lJ1Ob3

Wilson, W., and C. Barton, 2022, “Tackling the under-supply of housing,” Research Briefing, House of Commons Library, February 4, https://bit.ly/3sso9Um

CYBER

1R2
C500
C503
25.0F6K
RTC27
RX35
SKW23
C-100
16V
CFG9002V3
U
1576903

A SEMANTIC FRAMEWORK FOR ANALYZING "SILENT CYBER"

KELLY B. CASTRIOTTA | Global Cyber Underwriting Executive, at Markel Corporation[1]

ABSTRACT

Insurers developed property and casualty insurance policies prior to widespread computerization and the prolific use and transmission of electronic data. Many such insurance contracts did not expressly address cyber exposures at the time of their initial creation. In 2015, the Prudential Regulatory Authority (PRA) formally introduced a theoretical problem of "silent cyber" to the insurance industry, contemplating catastrophic cyber scenarios with not only a potentially powerful impact on dedicated Cyber insurance portfolios, but also on traditional insurance portfolios. The issue soon became a reality in the wake of the expansive losses associated with the NotPetya attacks of 2017.

In response to the requests made by the PRA to insurers to manage "silent cyber", Lloyd's of London introduced a mandate to eliminate "silent cyber" on all Lloyds policies, first charting a course for the transformation of insurers' contractual wording to more appropriately address cyber risk. This article discusses the general concerns around "silent cyber" as presented by the PRA, the challenges of defining cyber risk across the insurance industry, and steps taken to rectify the silent cyber issue. The article then explores the idea that the silent cyber problem is at its core a semantic one rather than one of risk perception. The article concludes by offering solutions as to a semantic framework under which to analyze and address "silent cyber".

1. INTRODUCTION

Historic buildings are worth preserving not only because of their cultural significance, but also because they can be a potential source of revenue.[2] It may occasionally make economic sense to rebuild certain architectural structures in the face of new environmental threats or newfound recognition of the ways that existing threats impact aging structures.[3] However, there are alternatives to a destroy and rebuild approach, one of which is retrofitting older buildings with new materials or design features.[4] Seismic retrofitting, for example, is the act of performing engineering treatments such as preservation, rehabilitation, restoration, and reconstruction, to improve a historic building's ability to withstand earthquakes.[5] With appropriate retrofitting, contemporary architects can maintain older buildings by implementing and layering emerging design technologies upon older ones, thereby maintaining the integrity of cultural structures.[6]

[1] The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of Markel Corporation or any of its subsidiaries or holdings. All content within is for general informational and academic research purposes only and not intended as legal advice. Article republished from original source, Castriotta, K. B., 2021, "A sematic framework for analyzing "silent cyber"," Connecticut Insurance Law Journal 27:2, 68-104.

[2] See Sigmund, Z., V. Ivanokovic, and A. Braun, 2011, "A challenge of retrofitting a historical building," 2nd WTA International PHD Symposium Building Materials and Building Technology to Preserve the Built Heritage, at 1.

[3] See Hutchinson, T., 2012, "Retrofitting is expensive – let's demolish and start again," The Guardian, April 3, https://bit.ly/3sYnnh9.

[4] See Sigmund, supra note 2 at 2.

[5] See id.

[6] See id.

We can look at the issue of "silent cyber"[7] in a similar light. The insurance industry[8] has developed and maintained a prolific body of contractual architecture (policies) that has created a legacy of meaningful risk transfer products for customers. Among those products is the relatively emergent Cyber insurance product, specifically designed to cover certain aspects of so-called "cyber risk". The insurance industry has historically paid losses associated with their insurance products and remained profitable.[9] Similar to the architectural community, the insurance industry also occasionally encounters emerging appreciation of the catastrophic[10] reach of specific threats. In recent years, one such concern is the wide reach of cyber risk[11] and with it, concerns as to whether the insurance industry will be able to withstand an event like a malware attack on the United States' power grid.[12] Compounding this fear is a recognition that perhaps "silent" cyber exposure will extend beyond the realm of monoline Cyber[13] insurance portfolios and threaten the sustainability of traditional[14] lines of insurance coverage. Specifically, the industry is concerned about risks that it failed to consider and to adequately price for cyber losses (attritional[15] or otherwise).

As such, the industry has, on the one hand, a vast set of traditional risk transfer products not specifically engineered to withstand such cyber risk, and on the other hand, an emerging set of risk transfer products (and in some cases, services) that have been intentionally created to address cyber risk. This article proposes that one solution to the concerns regarding "silent cyber" is to "retrofit" traditional insurance products with language and other normative concepts borrowed from standalone Cyber products.

> “*New ideas must use old buildings.*
>
> **Jane Jacobs** – *The life and death of great American cities* ”

A prerequisite to solving the problem of "silent cyber" is the adoption of a consistent semantic framework to be implemented across an insurance enterprise. This approach will ultimately lead to better evaluation and quantification of cyber exposure within any specific firm's insurance portfolio and across the industry. The framework should be flexible enough to adapt to the iterations of the Cyber insurance product sold today and in the future. In turn, this article will offer a definition of "silent cyber" that can be used to determine what should and should not be covered by non-Cyber policies. Such a semantic framework focuses on the "nesting"[16] of Cyber and non-Cyber policies, and emphasizes that losses that are covered by Cyber policies should not be covered by non-Cyber, and vice versa (unless done so intentionally). Just as auto and homeowner's policies "nest" together by covering mutually exclusive risks, the same should be true of Cyber and non-Cyber policies. To accomplish this, non-Cyber policies should continue to cover losses where a cyber-as-a-peril is involved in the causal chain of a loss and there is a physical alteration to the structure of tangible property. By contrast, traditional policies should not cover any losses that are in fact covered by current Cyber insurance policies.

[7] When used as a noun, the term "silent cyber" will appear in quotations, but when used as an adjective, the phrase will appear without quotations.

[8] The phrase "insurance industry", when used throughout this article, is to be construed broadly to include businesses that partake in the underwriting and procurement of insurance or reinsurance products.

[9] For a quick snapshot of 2020 profitability, see "Visualizing the 50 most profitable insurance companies in the U.S.," August 10, 2020, https://bit.ly/34Vf6lo. For a historic view, see Lynch, J., 2016, "The property/casualty landscape profitability, growth – disruption?" Insurance Information Institute, September 26, https://bit.ly/3uPBV55. For a forward-looking view, see Shaw, G., and N. Baumann, 2020, "2021 insurance outlook: accelerating recovery from the pandemic while pivoting to thrive, Deloitte, December 3, https://bit.ly/3sF3V8N.

[10] When this article refers to "catastrophic" losses, this is generally intended to mean the same as correlated losses, systemic losses, or accumulated losses – all losses other than attritional losses. See infra at 15.

[11] See Reinsurance News, 2017, "Swiss Re highlights role of re/insurance in cyber risk," March 6, https://bit.ly/3LyeR0M.

[12] See Trevor Maynard, et. al., "Lloyd's emerging risk report – 2015," Innovation series: The insurance implications of a cyber-attack on the U.S. power grid," Centre for Risk Studies, University of Cambridge Judge Business School, https://bit.ly/36dfKvs.

[13] References to cyber-specific insurance policies are denoted with capitalized version of the word "Cyber". References to cyber-as-a-peril (or hazard) are denoted with a lower-case version, "cyber".

[14] References to "traditional lines" or "traditional property and casualty" insurance policies include the broad array of products to cover bodily injury, property damage, liability, and professional risk developed prior to 1990.

[15] References to "attritional losses" are those losses other than losses associated with catastrophes. When I refer to expected losses, non-systemic losses, or non-catastrophic losses, I am referring to attritional loss.

[16] In this context, the "nesting" of sets of insurance policies refers to policies that, as a rule, complement each other, by covering specific aspects of a risk, but not the same aspects of a risk.

The article ends with prescriptive view of how to view cyber risk: by embracing the Cyber insurance product framework that the industry has developed. To reach this conclusion, this article will examine the current semantic frameworks offered (as set forth by the PRA and other regulatory bodies) and the problems with having disparate frameworks for such, and offer potential solutions to be implemented on a firm-by-firm basis.

2. CYBER AS A COVERAGE

Three conceptual coverage parts comprise a contemporary Cyber insurance product: (1) third-party liability coverages; (2) first-party coverages; and (3) business interruption coverages (which are technically first-party coverages, but of a specific "time element" nature). Each respond to a variety of cyber incidents, spanning from cyberattacks on one's own network, to system failures and other outages, to cyberattacks on a network provider's system (herein "cyber event"). Liability coverages are typically offered as follows: privacy and security liability, media liability, regulatory coverage, and payment card industry (or "PCI") coverage. The first-party coverage includes incident response (including call center costs, credit monitoring, and related mitigation costs), cyber extortion payments, and restoration costs. The business interruption part typically includes coverage for the costs of interruption of business due to a cyber event, whether the event is perpetrated upon the policyholder itself or a business upon which a policyholder depends. This often includes the reputational costs associated with a cyber event.

The coverages[17] are a good place to find a common understanding of what the industry considers to be covered or potentially covered cyber loss. For example, liability coverages naturally respond to the legal costs and the damages (judgments, fines and penalties, or settlements) that arise from a cyber event. First-party coverages tell us in detail what cyber losses a business may suffer. For instance, an incident response insuring agreement tells us about the costs incurred to engage a host of service providers that are needed to respond when there is a security or privacy incident. These include breach counsel, privacy counsel, credit monitoring services for customers, forensic providers, and public relations firms. The extortion and restoration agreements provide coverage for ransomware payments made to cyber criminals and the costs of a cybersecurity firm to restore one's data (and in some cases, hardware). And finally, the business interruption coverages tell us that companies may undergo loss of income and even loss of contractual or other business opportunities due to a cyber event.

3. FROM ABERRATION TO AGGREGATION

The next step is to elucidate industry concerns surrounding "silent cyber". The insurance industry has been formally discussing the issue of "silent cyber" since 2015, with most crediting the PRA as the initial regulatory catalyst for the movement towards eradicating "silent cyber" in insurance portfolios. In many ways, the silent cyber problem has existed well before 2015, following a history of professional advice as to where to find cyber coverage under traditional insurance policies.[18] For example, until around 2014,[19] commercial general liability policies rarely included concepts or language specific to cyber risk and even then, they were specifically focused on privacy exposures associated with computer hacking (as opposed to other security and business threats). Conflicts between insurers and policyholders developed over the applicability of coverage as they applied to emerging situations, such as whether coverage existed for damage to data and whether data was tangible property.[20] Other examples of such disputes include those where policyholders sought coverage under property policies because of power outage events (impacting computerized systems) under a theory of "loss of use or functionality", even where the outage did not

[17] Note that the insuring agreements of a Cyber policy provides a normative view of what constitutes cyber loss, even though Cyber policies typically only extend to financial loss (defined as pure economic loss that would be reflected as loss in a balance sheet only). To achieve a more nuanced picture of what constitutes cyber loss, we could also look to the common exclusionary language in Cyber policies. This article will not address common exclusions in Cyber policies.

[18] See Clarke R., 2013, "Cyber liability: where to find cyber coverage," Insurance Journal, January 28, https://bit.ly/3JueAd4.

[19] In 2014, ISO introduced endorsements "addressing the access or disclosure of confidential or personal information," https://bit.ly/3rOLKhV.

[20] See, e.g., West Bend Mutual Ins.Co. v. Krishna Schaumburg Tan, Inc., 2020 Ill. App. LEXIS 179, at 12 (Ill. Ct. App. Mar. 20, 2020) (holding that under a general liability policy, coverage part b, "publication" encompasses the act of providing plaintiffs fingerprint data to a third party, alleged to be in violation of the Biometric Information Privacy Act (Act) (740 ILCS 14/1 et seq. (West 2014)); Eyeblaster, Inc. v. Fed. Ins. Co., 613 F.3d 797, 803 (8th Cir. 2010) (describing invasion of privacy and deceptive practices allegations from the installation of advertising tracking software on a non-consenting plaintiff, and finding "loss of use" of computer allegations fell within "tangible property" terms of general liability policy); Am. Guarantee & Liab. Ins. Co. v. Ingram Micro, Inc., No. 99-185 TUC ACM, 2000 WL 726789, at *3 (D. Ariz. April 18, 2000) (describing how a power outage knocked out systems, causing loss of data and loss of software functionality, and the court found there was "property damage" per CGL terms). Compare, Am. Online, Inc. v. St. Paul Mercury Ins. Co., 347 F.3d 89, 97–99 (4th Cir. 2003) (finding that data, information, and instructions are not "tangible property," and that an "impaired property" exclusion precluded coverage for loss of use of tangible property that is not physically damaged), with Zurich Am. Ins. v. Sony Corp. of Am., No. 651982/2011, 2014 N.Y. Misc. LEXIS 5141, at *67–72 (N.Y. Sup. Ct. February 24, 2014) (describing how an insured sought coverage under CGL terms for alleged transmission of private information by hackers and finding no coverage).

amount to actual physical damage.[21] Much of the focus of these disputes focused on underwriting and drafting intent. In other words, did the policy wording offer coverage for a cyber loss, even though the insurers did not price the policy to cover this type of risk? In this type of scenario, underwriters did not necessarily contemplate losses caused by cyber threats and, therefore, the definition of loss expanded beyond the intended scope of coverage.

The conversation about unexpected cyber losses began to morph after the PRA performed a cross-industry survey regarding cyber risk in 2015.[22] The initial PRA findings were grim, including the finding that the failure to account for cyber exposure in traditional insurance lines was material and likely to worsen with time.[23] The PRA also found that the industry was hamstrung from taking appropriate corrective action due to a lack of effective cyber exclusions, lack of clear strategy and risk appetite, and an insufficient grasp of aggregation and tail potential of affirmative cyber.[24] As the focus of the PRA findings revolved around potential catastrophic losses, the conversation circles about "silent cyber" broadened from the plaintiff's bar to the C-suite of insurance companies.[25]

A secondary catalyst for this broader conversation was the series of cyberattacks in 2017, known as NotPetya,[26] which amounted to more than U.S.$10 billion in losses.[27] As the loss picture of the NotPetya[28] attacks sharpened in 2019,[29] the concerns shifted from attritional losses (usually due to aberrations in coverage) to mountainous aggregation[30] issues. Aggregation concerns arise when multiple policies or multiple lines of coverage offered to an insured (either by design or inadvertently) are triggered from a single event, and as such, there is an accumulation of loss across product lines underwritten by any one insurer. "Silent cyber" poses a particular aggregation challenge to insurers because monoline Cyber policies are often the only policies underwritten to cyber risk. As aggregation concerns relate to "silent cyber", underwriters underestimate the accumulation risk within a product line or for a specific insured across multiple product lines due to the possibility that traditional policies may unexpectedly respond to cover such losses. A large scale, or geographically expansive cyberattack could impact multiple insureds and multiple policies, both traditional and Cyber-specific.

4. SEEKING NORMATIVITY

The PRA's definition of "silent cyber" evolved over the course of its surveys, findings, and publications. Some versions rely on normative concepts of "cyber risk", "cyber exposure", or "cyber-related losses", while others rely on terminology commonly used and defined in Cyber-specific insurance policies. By describing the issue of silent cyber both with normative cybersecurity concepts on the one hand and with Cyber policy concepts on the other, the PRA was touching upon the two main categories of silent cyber loss: ensuing loss and cyber product loss. Both categories are important. A definition of Cyber product loss allows insurers to effectively treat situations in which overlapping coverages are inadvertently provided. A definition of ensuing loss is equally important, given that non-Cyber policies will in fact respond to losses caused by cyber risk.

[21] See Am. Guarantee, 2000 WL 726789, at *2 (describing an electrical outage, where an insurer said there was no "physical damage" pursuant to "all risks" policy language, yet finding that "physical damage" is not restricted to physical destruction or harm of computer circuitry but includes loss of access, loss of use, and loss of functionality); see also, National Ink & Stitch, LLC v. State Auto Property & Casualty Insurance Company, 2020 U.S. Dist. LEXIS 11411 (U.S.Dist. Ct., Maryland) (holding that loss/corruption of electronic data and software and reduced efficiency of computer systems due to a ransomware event amounted to direct physical damage under BOP policy). But see Ward Gen. Ins. Servs., Inc. v. Emp'rs Fire Ins. Co., 114 Cal. App. 4th 548, 554–55 (Cal. Ct. App. 2003) (finding no coverage for costs of recovery of data or business interruption because there was no loss of, or damage to, tangible property).

[22] See Letter from Chris Moulder, Director of General Insurance at Bank of England, PRA (Prudential Regulatory Authority), August 10, 2015, https://bit.ly/3gNE09v (including questionnaires as to cybersecurity and resilience, cyber insurance, and conduct).

[23] See Moulder, 2016 Letter, infra note 39 at 1.

[24] See id. at 1–2.

[25] See Consultation Paper CP39/16, "Cyber insurance underwriting risk", Bank of England: Prudential Regulation Authority (Nov. 2016) at 5, https://bit.ly/34YjeRC (noting that the responses to its investigation were made by the following roles within insurance firms: Chief Underwriting Officer, Chief Risk Officer, Chief Actuary, Lead Cyber Underwriter, and Head of Exposure Management).

[26] See Krebs on Security, 2017, 'Petya' ransomware outbreak goes global," June 27, https://bit.ly/3Jri1kZ.

[27] See generally Abraham, K., and D. Schwarcz, 2021, "Courting disaster: the underappreciated risk of a cyber-insurance catastrophe," Connecticut Insurance Law Journal (forthcoming) (discussing the prospect of cyber incidents having the potential to simultaneously cause very large losses to numerous firms across the globe, thus resulting in a cyber "catastrophe"); see Willis Towers Watson , 2020, "The problem of silent cyber risk accumulation," February 25, https://bit.ly/3rOt3ul. See also Mondelez v. Zurich, No. 2018L011008, 2018 WL 4941760 (Ill.Cir.Ct) (subject litigation filed by Mondelez).

[28] See Johansmeyer, T., 2019, "Could NotPetya's tail be growing?" Verisk, https://vrsk.co/34AudRN (referring to a PCS study that NotPetya's economic losses were estimated at U.S.$10 bln by 2017).

[29] See Ward, C., 2020, "Cyber turned inside-out: three years after NotPetya," Carrier Management, June 17, https://bit.ly/3HORfSZ (estimating U.S.$10 bln in losses associated with NotPetya, but with estimated U.S.$3 bln in insurable losses from policies other than cyber dedicated lines).

[30] The term "aggregation" is used synonymously with the term "accumulation" throughout this article.

Put in terms of cause and effect, the "ensuing loss" category of silent cyber loss addresses "cyber" as a peril[31] or as a hazard and refers to losses[32] that flow from such cyber perils or hazard. In other words, as humanity grows increasingly dependent upon computers and digitization, the mere use of a computer or computer-operated technology will result in losses from otherwise covered perils. Another way to put this is that a computer is somewhere involved in the causal chain of the loss, even if the computer was not the sole cause[33] or the proximate cause[34] of the loss. This type of "silent cyber" is where a loss is caused by or results from computer-related acts or events, but where such cause does not change the nature of the expected loss under any given policy (but may change the magnitude or frequency of such loss). The exposure is typically "silent" due to the structure of all-perils policies. An example of ensuing loss is where a hacker (cyber incident) exploits a vulnerability in a computerized device that ignites a fire (a traditionally covered peril), which causes property damage to a building (an ensuing loss).[35] Historically, this type of incident would be covered under a property policy that covers damage to a building caused by fire, a covered peril, regardless of the use or involvement of a computer. Accordingly, there is no apparent mismatch between the policy offering and the underwriting intention in terms of type of risk, even though the policy's language may fail to expressly discuss computer-related technologies.

The other category of "silent cyber" relates to Cyber as an insurance product. This version of "silent cyber" is where the losses covered by a non-Cyber policy stemming from a cyber event overlap with losses specifically covered by a Cyber insurance product, against the insurer's intention that traditional policy and Cyber policies "nest" together to cover mutually exclusive sets of losses. In these cases, the cyber-related acts or events result in loss that is a change to the nature or the characteristics of expected loss under a traditional insurance policy. The result is tantamount to the type of coverage one would normally find in the insuring agreements of a Cyber policy. Such losses often come as a surprise to the underwriter, are brought under a novel theory of loss (from the perspective of the insurer), and were not factored into the underwriting process when pricing and terms were quoted. Put another way, such losses are aberrations as to what is underwritten to and ultimately modeled by pricing or CAT actuaries for that specific product line. This type of silent cyber loss has to do with "cyber," not as a normative concept of cyber risk, but as a normative concept of a distinct type of insurance product line ("Cyber"). An example of this is where a retailer experiences a cyberattack whereby the personal data of many customers is exfiltrated, including correlated bank account information. The banks, who must now re-issue all affected credit cards to consumers, proceed to sue the retailer-insured to recover the costs of the cards (Cyber product loss). Consequently, the insured alleges that this is a form of damage to tangible property due to their limited usability (novel loss theory).[36]

Earlier iterations of the PRA's definition of "silent cyber" have combined the two views of the phrase: one, having to do with "cyber" as a cause of loss, and the other, having to do with "Cyber" as a type of insurance coverage. In a 2016 advisory, for example, the PRA explained that it was investigating the question of underwriting risks emanating from affirmative[37] Cyber insurance policies, but also "from implicit cyber exposure within 'all-risks'[38] and other liability insurance policies that do not explicitly exclude cyber risk. This latter type of cyber risk is referred to as 'silent' cyber risk..."[39] In this characterization, the PRA focuses on scenarios where

[31] Peril, Black's Law Dictionary at 524 (2nd pocket edition 2001). Black's defines "peril" as follows: 2. Insurance: The cause of a loss to a person or property. Compare with, Black's definition of hazard: "The risk or probability of loss or injury esp. a loss or injury covered an insurance policy." Id. at 316.

[32] Generally speaking, "ensuing losses" are losses that follow from an incident that causes direct physical loss or damage.

[33] Sole cause, Black's Law Dictionary at 89 (2nd pocket edition 2001). Black's defines "sole cause" as follows: The only cause that, from a legal viewpoint, produces an event or injury. If it comes between a defendant's action and the event or injury at issue, it is treated as a superseding cause.

[34] Proximate cause, Black's Law Dictionary at 88 (2nd pocket edition 2001). Black's defines "proximate cause" as follows: 1. A cause that is legally sufficient to result in liability. 2. A cause that directly produces an event and without which the event would not have occurred. Id.

[35] See Wagensiel, P., 2011, "Printers can be hacked to catch fire," Scientific American, November 29, https://bit.ly/34GFNdZ (relaying findings by Columbia University researchers that attackers may spread malware causing printers to overheat and catch fire).

[36] This is a novel theory of loss because it involves an allegation that cards are damaged based on "loss of use" versus actual physical damage to the card, particularly because the cards were physically useable after the attack. In other words, users could physically swipe their affected credit cards, albeit not without consequence. See Target Corp. v. ACE American Ins. Co., et al, 2021 WL 424468 at *7 (D. Minn. Feb. 8, 2021) (holding that Target could not obtain coverage from its CGL to replace credit cards after a data breach under a "loss of use" theory as the cards diminution in value did not amount to loss of use).

[37] Affirmative Cyber policies are insurance policies that specifically respond to a variety of so-called "cyber incidents," including ransomware attacks, viruses, ddos attacks, but also to computer system failures, supply chain interruptions, and exfiltration of private data (both digital and non-digital).

[38] All-risks policies refer to traditional property and casualty policies that respond to all perils unless specifically stated otherwise.

[39] See Letter from Chris Moulder, Director of General Insurance, Bank of England, PRA, to CEO's [of various insurers], at 1, November 14, 2016, https://bit.ly/3p9XaLt. See also, Consultation Paper 39/16, supra note 25 at 5.

cyber exposure is implicitly covered within all-perils insurance policies. The reason why this would be an area of "silent cyber" is because such all-perils policies would readily have been developed, standardized, and well-established prior to the computerization of society. As such, the policies did not contemplate that the use of a computer to cause harm could be a peril, simply because computers were not in commercial use at the time the language was initially developed.[40] More appropriately, the cyber aspect was not so much silent as it was absent. Notably, these traditional policies were also first developed prior to the invention of a standalone Cyber policy. So, underwriters could not have possibly considered whether the type of loss would be redundant with an affirmative Cyber insurance product.

Later, in a 2017 Supervisory Statement, the PRA defined cyber insurance underwriting risk as "the set of prudential risks emanating from underwriting insurance contracts that are exposed to cyber-related losses resulting from malicious acts (e.g. cyberattack [sic], infection of an IT system with malicious code) and non-malicious acts (e.g. loss of data, accidental acts or omissions) involving both tangible and intangible assets."[41] Here, the PRA introduced a dichotomy between malicious and non-malicious behaviors that recurs in Lloyd's wording[42] developed to address "silent cyber".[43] In other words, a prudential risk – or a non-silent risk, rather – is one that is intentionally underwritten to and priced for, whereas with silent cyber exposures, one of those two elements is absent: underwriting intent as to cyber risk or pricing as to cyber risk.[44] In the same 2017 Supervisory Statement, the PRA simplifies the definition of non-affirmative cyber as: "insurance policies that do not explicitly include or exclude coverage for cyber risk."[45] Given that here the PRA is referring to insurance policies, which are contractual arrangements commemorated in writing, it follows that one of the primary issues of "silent cyber" is an issue of language – specifically, the failure of the underwriter to clearly express whether 1) cyber perils are covered and 2) that coverage is the same kind of coverage found in an affirmative Cyber insurance product.

One of the major issues with the PRA's earlier definition of "silent cyber" is that it attempts to define cyber underwriting risk in relation to a normative concept of "cyber risk" – a concept that the PRA does not define.[46] As such, in evaluating its portfolio's cyber risk, the carrier is then left to determine whether "cyber risk" is the same as "cyber underwriting risk" and in turn, whether this equates to "cyber-related losses" or is something else altogether. A lack of construct in this regard leads to ambiguity in insurers trying to assess, measure, and course correct as to cyber exposure across product lines. If the PRA is going to characterize a type of risk as prudential, there also must be some foundational concept of what that risk is (and what it is not).

In the PRA's Policy Statement[47] referencing a concept of "cyber risk", the PRA also explained that the definition of "silent cyber" should be understood as the equivalent of a concept of "non-affirmative cyber".[48] Here, the PRA departs from a definition of "silent cyber" that is entirely dependent upon a concept of "cyber risk" per se. According to the PRA, "silent cyber" and "non-affirmative cyber" can be used interchangeably.[49] Four of the thirteen respondents to the PRA's Consultation Paper pointed out that the use of the term "silent" cyber risk is problematic and may create ambiguity in future arbitration or litigation cases.[50] Moreover, two respondents suggested that the term "non-affirmative" cyber risk should be used instead

[40] See generally Lloyd's Wording Repository, https://bit.ly/3JwxfFr.

[41] Supervisory Statement SS4/17, "Cyber insurance underwriting risk," Bank of England, PRA, at 5, July 2017, https://bit.ly/34T0tPF.

[42] See generally Lloyd's Wording Repository, https://bit.ly/3JwxfFr.

[43] See Supervisory Statement SS4/17, supra note 41. But see, Consultation Paper CP39/16, "Cyber insurance underwriting risk," Bank of England, PRA, November 2016) at 5, https://bit.ly/34Z1Dt7 (PRA defines cyber underwriting risk is as the set of prudential risks emanating from underwriting insurance contracts that are exposed to losses resulting from a cyberattack).

[44] See Supervisory Statement SS4/17, supra note 41.

[45] See id. at 5.

[46] There is no known standardized definition of the term "cyber risk". I have come across a variety of definitions of cyber risk. See, e.g., CRO Forum, 2014, "The cyber risk challenge and the role of insurance," paragraph 3, December 2014, https://bit.ly/33l1Bv9; Committee on Payments and Market Infrastructures and International Organization of Securities Commissions, 2016, "Guidance on cyber resilience for financial market infrastructures," June, https://bit.ly/3sFpwhl.

[47] See Policy Statement PS15/17, "Cyber insurance underwriting risk," Bank of England, PRA, July 2017, https://bit.ly/3JrQGyZ. Policy Statement SS4/17 is responsive to Consultation Paper (CP) 39/16 "Cyber insurance underwriting risk," including Supervisory Statement (SS) 4/17 "Cyber insurance underwriting risk," which sets out the PRA's final expectations regarding the prudent management of cyber insurance underwriting risk. Id. at 1.

[48] Policy Statement PS15/17, supra note 47 at 5. See also, Supervisory Statement SS4/17, supra note 41 at 5-7.

[49] See Policy Statement PS15/17, supra note 47 at 5. See, Supervisory Statement SS4/17, supra note 41 at 5 (stating "non-affirmative cyber risk, i.e., insurance policies that do not explicitly include or exclude coverage for cyber risk. This latter type of cyber risk is sometimes referred to as 'silent' cyber risk by insurance professionals.") Other definitions of "silent cyber" exist. For an example, see Guidewire's definition in "Silent cyber scenario: opening the flood gates," October 2018), https://bit.ly/34BcXMe ("We define "silent cyber" exposure as the potential for cyber risk to trigger losses on policies where coverage is unintentional, unpriced, or both. "Unintentional" coverage means not explicitly excluded or affirmed (with any applicable sublimit)").

[50] Policy Statement PS15/17, supra note 47 at 5.

whereas one respondent suggested a distinction based on whether a cyberattack is a named peril or not.[51] Finally, one respondent suggested that the distinction between "silent" and "affirmative" should be completely removed and instead referred to "cyber risk exposures". As a result, the PRA agreed that the use of "non-affirmative" cyber risk would be less ambiguous and adopted the use of affirmative cyber risk (insurance policies that explicitly include coverage for cyber risk); and non-affirmative cyber risk (policies that do not explicitly include or exclude coverage for cyber risk).[52] This is important because it points to one of the PRA's major concerns: aggregation.[53] Specifically, PRA seeks to identify the potential for "clash" (wherein an insurer can experience excessive covered losses due to one insurable event).[54] In its equivocation of "silent cyber" as "non-affirmative cyber", the PRA's reference point is not only "cyber risk" per se, but affirmative Cyber coverage, meaning, an actual cyber-specific product offered by the insurance market.

Others who have attempted to define "silent cyber" also embrace the two distinct concepts: normative risks from cyber-as-a-peril and Cyber as an insurance product. For example, the European Insurance and Occupational Pensions Authority (EIOPA) utilizes a definition of "silent cyber" akin to the PRA's definitions: "Non-affirmative cyber risk refers to instances where cyber exposure is neither explicitly included nor excluded within an insurance policy. The latter type of cyber risk is also referred to as 'silent' cyber risk."[55] Like the PRA, the EIOPA's definition of "silent cyber" both references a concept of cyber risk and refers (albeit loosely) to a Cyber insurance offering. Unlike the PRA, the EIOPA attempts to define "cyber risk". The EIOPA's methodology for this exercise involved asking participants[56] for their enterprise's definition of cyber risk, while providing a cyber risk definition from the Financial Stability Board (FSB) Cyber Lexicon[57] as an initial reference.[58] The results of the EIOPA's survey varied widely with some groups relying on FSB definitions, some on the American Association of Insurance Services (AAIS) definitions, some relying on regulatory concepts, and others not having a working definition whatsoever.[59]

The EIOPA concluded[60] that having a clear and common set of definitions would foster a more productive dialogue regarding cybersecurity challenges, including quantification methods for "silent cyber". Its straightforward observation aligns with the PRA's findings regarding disparate opinions as to the amount and severity of cyber risk within traditional lines of coverage. As discussed, this divergence in view likely stems from a lack of a collective semantic framework. What the EIOPA, the PRA, and the FSB overlook, however, is the idea that an established semantic framework already exists and is fully accessible to insurers. The sector has already built a strong framework based upon a series of normative constructs and definitions that comes close to a fully formed concept of "cyber risk" vis-a-vis its current Cyber product offerings.

5. CYBER INSURANCE AS THE SEMANTIC SOLUTION

What if, instead of relying upon definitions derived from outside the insurance industry to address "silent cyber", the insurance industry drew upon its own resources as a normative guide for cyber risk? Even though a market-standard monoline Cyber policy will typically only provide coverage for financial loss (and does not typically extend to bodily injury and property damage), insurance carriers can still refer to the insuring agreements of such a standalone policy to formulate a comprehensive idea as to what "cyber risk" means, both to the insurance industry and to its policyholders.

[51] Id.
[52] Id. at 5–6.
[53] Clarification of wording alone will not stymie the impact of catastrophic cyber losses to any single insurance firm. However, clarifying the wording and "channeling" the coverages to the appropriate products may serve to gain better or more accurate outputs from cyber models.
[54] See Supervisory Statement SS4/17, supra note 41 at 7 (describing minimum standards for insurers to incorporate cyber insurance underwriting risk stress tests that explicitly consider the potential for loss aggregation (e.g., via the cloud or cross-product exposures) at extreme return periods (up to 1 in 200 years)).
[55] EIOPA, 2019, "Cyber risk for insurers: challenges and opportunities," European Insurance and Occupational Pensions Authority, at 18, https://bit.ly/3oMIFxK.
[56] See id. at 3. Participants included 41 large (re)insurance groups across 12 European countries representing a market coverage of around 75% of total consolidated assets.
[57] See generally Cyber Lexicon, Financial Stability Board, November 12, 2018, https://bit.ly/3Jq2JwF. The FSB developed a cyber lexicon in November 2018, in part, to assess and monitor financial stability risks of cyber risk scenarios.
[58] The Cyber Lexicon defines cyber risk as "the combination of the probability of cyber incidents occurring and their impact." Id. at 9 (adapted from Committee on Payments and Market Infrastructures-International Organization of Securities Commissions, International Association of Insurance Supervisors (CPMI-IOSCO, ISACA) Fundamentals and ISACA Full Glossary).
[59] See Cyber risk for insurers, supra note 55 at 7 (emphasis added).
[60] Id.

Since its earliest iterations, the Cyber policy offering has evolved to stay fit for purpose. The coverage will continue to evolve as offerings expand and contract in response to the threat environment, customer needs, and the performance of affirmative Cyber portfolios.[61] However, there are two main reasons to rely upon cyber concepts that are already formulated in a Cyber insurance coverage policy. One is that the Cyber insurance policy is developed from a set of norms that the industry already accepts, some of which was directly in reaction to the threat environment experienced by actual companies, so it is a good place from which to establish common dialogue.

A second reason is that the industry's preoccupation with "silent cyber" is due in large part to the potential "clash" risk involved with having accumulative and redundant cyber coverages available to the same client or subject to the same cyber event, unbeknownst to the underwriters. Namely, of the two theories of "silent cyber" loss, the Cyber product loss is the more pressing aspect of the silent cyber problem. By its very definition, ensuing loss from a cyber event is likely contemplated by the underwriter and priced for accordingly. And because the cyber event is one event among others on the causal chain, as opposed to being the single event on the causal chain, ensuing loss has an anchor to a time and place type peril (e.g., fire), which helps to anchor the loss in a predictable pricing manner. On the other hand, cyber risk as a form of product loss is where insurers can start to see the pronounced effects of accumulation across a portfolio. Because Cyber as a product loss refers specifically to covered losses under affirmative Cyber policies, where traditional policies respond to the cyber perils in the same type of way as Cyber policies, there is a real potential for an insurer to have significant limits exposed to a cyber event at significantly reduced pricing. Accordingly, if Cyber product loss accumulation is the more prominent concern of "silent cyber," correcting traditional policy language to eliminate (or at least price for) redundant Cyber coverage becomes the first priority[62] of the "silent cyber" solution. To accomplish this objective, an enterprise must be well-versed in the mechanics and semantics of a typical standalone Cyber offering.

[61] See Jones, J. H., 2021, "AIG introduces ransomware co-insurance and sub-limits at 1.1 cyber renewals," Insurance Insider, January 8, https://bit.ly/3HPPcOR.

[62] A secondary component of the "silent cyber" solution is the capability to accurately map and quantify the areas of Cyber product losses, regardless of the original intent of the underwriter at the time of binding. Quantifying this accumulation exposure can be done more meaningfully if insurers map cyber exposures to the general categories of insurable Cyber losses throughout their portfolios.

To some, the following analysis may seem to presuppose that affirmative Cyber coverage is an accurate reflection of the real cybersecurity landscape. Certainly, there is an overlap as to the realities of cyber, as a peril, and "Cyber," as an insurance product as demonstrated further in the history of the Cyber product section of this article. Regardless, while it may be the case that an insurance policy is a kind of representation of the threat or peril that it purports to cover, it uses abstractions to describe both the coverage triggers and the losses.[63] Accordingly, that policy language accurately reflects the actual threat environment or encompasses all that can be imagined as "cyber risk", is less important than it is for the insurance policy to accurately reflect the intentional and insurable (whether potential or actual) cyber risk. By "insurable" cyber risk, I am referring to the causes of loss and the types of loss to be covered, as contemplated by the underwriter.

As such, the appropriate definition of cyber risk for "silent cyber" is simply the type of risk that insurers of affirmative Cyber are generally willing to cover at a given point in time. Of course, there is no one single standard for a standalone Cyber coverage offering now or in the past, and there continue to be changes in policy offerings across various firms, along with nuances of certain offerings. However, there are coverage norms from which the insurance industry can gain a better understanding of the risk landscape as it seeks to correct the problem of "silent cyber". In other words, what we are looking to do with "silent cyber" is align portfolios within insurance companies and across the insurance industry. To realize this goal, a common language and framework for understanding must be accessed from within so that the industry can retrofit its aging architecture of insurance terminology to confront this emerging risk.

6. CONCLUSION

Many organizations and government bodies are widely concerned about the risks associated with computers. The media attention does not allow the public to ignore cyber threats, albeit much of the attention is dedicated to individual attacks against disparate companies, and less of it is focused on events that would lead to widespread, cumulative, and catastrophic loss. Since the PRA's work on "silent cyber" in 2015, however, there has been increased awareness of correlated cyber risk, especially silent cyber exposures, and fears of underpricing for it within an insurer's portfolio.[64] Most stakeholders seem to agree that cyber risk is a risk that should be measured, priced, underwritten, and otherwise treated appropriately. So, how do we then reconcile the acute variations in understanding cyber exposures simply as differences in perception of risk? Instead, insurers must admit that there is an emerging consensus around the perceived severity of cyber risk. They must also recognize that the central issue of "silent cyber" is first and foremost a problem of semantics. When insurers and governing agencies have looked for a common language regarding cyber, they have looked outward, instead of looking inward. This has led to confusion and discord, which in hindsight was largely avoidable had the industry and its regulators used the nomenclature at its disposal.

Carriers' first step to addressing "silent cyber" has been to review and potentially alter policy wording with regard to cyber risk. Curiously, most of the characterization has been dedicated to insurers making efforts as to "clarifying intent".[65] The suggestion is that the intent the insurance company seeks to clarify is "subjective" intent.[66] The characterization is a strange one considering insurance contracts: 1) consist of a series of logical syllogisms;[67] and 2) are (for the most

[63] Wollner, K. S., 1999, How to draft and interpret insurance policies, at 80, Casualty Risk Publications (explaining how abstractions are useful in succinctly drawing together a series of concrete ideas into a single concept and in anticipating unforeseen circumstances).

[64] O'Connor, A., 2018, "Insurers' worst fear: cyber hurricane or silent cyber?" Insurance Journal, March 21, https://bit.ly/3HThHv4. 65 See Marsh, 2020, "Silent cyber: what it is and how you can cover cyber perils," August, https://bit.ly/3BjsEUb.

[65] *See* Marsh, 2020, "Silent cyber: what it is and how you can cover cyber perils," August, https://bit.ly/3BjsEUb.

[66] Notably, the EIOPA promotes a mutuality in this undertaking: https://bit.ly/3LKDk35 ("A mutual understanding of contractual definitions, conditions and terms, for both, policyholders and insurance undertakings. Clear and transparent cyber coverages are crucial from a consumer protection perspective. It is the role of industry and consumers associations to provide this clarity and align expectations on cyber insurance coverages to avoid the potential for coverage disputes and costly litigation.").

[67] Wollner supra 63 at 140 ("normalized drafting represents an attempt to bring the certainty of symbolic logic to the drafting process.").

part) standardized. As such, legal interpretation of contracts (especially ones that fit this linguistic structure) depend almost entirely on the plain meaning of the text with the assumption that there is in fact an objective meaning to be communicated and understood. In such an interpretive undertaking, questions of intent on the part of the drafters or ratifiers of the document are rare and reserved for coverage litigation.

If insurers would recognize the futility in arguing over whether they should have seen the problem of "silent cyber" coming, and if they would cease their public posturing over the "original" intent of the policy language, perhaps they could then turn their attention to retrofitting the wording to the realities of the current threat environment, giving this problem some further thought as the PRA had suggested. I propose that insurers (and deciding courts) acquire a deeper understanding of the plain meaning of the wording contained in Cyber insurance forms, take those concepts, and apply them to traditional wording. The best frame of reference for analyzing whether there is Cyber coverage lurking in a traditional policy (and therefore more broadly within a product line) is the coverage afforded by a standalone Cyber policy. Not only will this reveal the plain meaning of critical definitions that govern both cyber as a peril and Cyber as a coverage, but this understanding will be derived from the collective expectation of coverage from the insurance consumer point of view. In other words, if one wants to know if a non-Cyber policy offers Cyber coverage, one must first read and understand what an affirmative Cyber policy offers. From this vantage point, insurers can begin to assess and measure the extent of "silent cyber" within their portfolios.

CYBER RESILIENCE: 12 KEY CONTROLS TO STRENGTHEN YOUR SECURITY

SARAH STEPHENS | Managing Director, International Head of Cyber & FINPRO UK Cyber Practice Leader, Marsh

ABSTRACT

Cyberattacks continue to dominate news headlines, driven by an overwhelming increase in ransomware events alone. As cyberattacks become more prolific, related insurance claims follow, meaning Marsh have been able to identify a correlation between certain security controls and corresponding cyber incidents. Organizations are recommended to implement a number of cyber hygiene controls that are key to achieving cyber resilience and insurability. In this article, Marsh presents 12 recommended cybersecurity controls including their characteristics and requirements.

1. INTRODUCTION

Cyberattacks continue to dominate news headlines, driven by a surge in ransomware events, which increased by an overwhelming 148 percent in 2021.[1] The perpetrators of these attacks now demand multimillion-dollar ransom payments as they cripple a business's operations, bringing them to a standstill until a payment is made.

As cyberattacks become more prolific, related insurance claims follow, meaning underwriters have been able to identify a correlation between certain controls and corresponding cyber incidents. Through this analysis and the continuous examination of relevant data points, the insurance industry has a rich understanding of the technical steps that organizations can take to build their cyber resiliency.

However, due to the growth in attritional losses, insurers are now taking a much more cautious position. Insurers are tightening their underwriting terms, carefully analyzing all cyber insurance applications, and asking more questions than ever before about an applicant's cyber operating environment and risk controls.

The adoption of certain controls has now become a minimum requirement for insurers, with organizations' potential insurability on the line. Organizations are undoubtedly placing more emphasis on controls than ever before to help mitigate their ransomware risks and improve their overall cybersecurity position and resilience.

While these controls have been established best practice for several years, some companies are still struggling to adopt them – most often because they have been unable to justify the cost of implementation, did not deploy them comprehensively, or did not understand or see the need for controls. In many regulated industries, where cyber resilience controls have been required for years, the effort was often more about checking a box than enhancing security.

Organizations are recommended to implement a number of cyber hygiene controls that are key to achieving cyber resilience and insurability. In this article, we present 12 recommended cybersecurity controls and their characteristics and requirements.

[1] https://bit.ly/3KrXywK.

2. MARSH'S 12 RECOMMENDED CYBERSECURITY CONTROLS

2.1 Control 1: Multifactor authentication (MFA) for remote access and privileged or administrator access

- **What is this control?** Multifactor authentication (MFA) is an additional login security layer to verify a user's identity when requesting access to a computer resource. MFA requires the user to provide two or more pieces of evidence to be authenticated from the following categories: "something you know" (such as a password/PIN), "something you have" (for example, a cryptographic identification device or token), and "something you are" (for example, a biometric).
- **Why should this control be adopted?** According to Marsh's findings published in "The changing face of cyber claims 2021,"[2] 80 percent of all cyber incidents are malicious and often start with compromised user credentials. MFA is an essential part of a strong identity access management (IAM) strategy, preventing unauthorized remote access to computer resources.

  MFA should be enabled in all systems, applications, and accounts that are accessible remotely, for all access by privileged and administrative users, and for all access to critical or sensitive data. In many cases, correct MFA implementation can help prevent cyber incidents – such as a costly ransomware attack. Insurers are requiring organizations to be more cyber resilient, with MFA as a key starting point. Ultimately, this will strengthen their security and will assist them in becoming better candidates for cyber insurance cover. At a minimum, companies should look for the enforcement of MFA on:

  - critical assets
  - privileged accounts
  - remote applications.

- **What a company needs to do to put this control in place:** in order to implement this control accordingly, businesses are recommended to: (1) require MFA for all remote logins to the corporate network by using secure remote access, such as virtual private network (VPN) and remote desktop protocol (RDP); (2) require multifactor authentication and encrypted channels for all administrative account access, irrespective of a user's location; (3) require MFA for access to the most critical or sensitive data or systems, irrespective of a user's location; and (4) enforce complex long passwords that are longer than 14 characters and use upper and lowercase letters, numbers, and symbols.

  As a basis to meet and implement the aforementioned requirements, an organization is advised to: (1) identify all systems and applications that are accessible remotely, critical and sensitive data (as well as all systems and applications that it is stored on), and all high-privileged and administrative users; (2) implement risk-based authentication, which is a method of applying varying levels of stringency to authentication processes based on the likelihood that access to a given system could result in it being compromised; (3) combine the VPN and any remote solutions with MFA; (4) identify all corporate devices, especially those that accept biometrics (such as laptops and mobile phones), for potential use as an additional factor; (5) check local regulation regarding data protection, privacy, and biometrics data (there may be limitations in using private devices, or biometric data, as a means to achieve the additional factor); (6) deploy factors in all devices to avoid a compromise affecting them all; and (7) train and inform employees on the value of additional layers of security before implementation of MFA to reduce resistance and avoid any misunderstandings.[3]

2.2 Control 2: Email filtering and web security

- **What is this control?** Email filtering software is used to scan inbound or outbound email traffic for undesired content. This can be less harmful spam emails approaching the recipient regarding specific actions – for example, selling a product or asking for donations – or phishing emails that represent a serious cybersecurity threat. These detected emails would automatically be filtered out, so they do not reach the user, or be flagged so the user is sensitized to the potentially malicious or unwanted content. Via email security software, suspicious and potentially malicious email attachments are additionally tested in a secure "sandbox" environment. Web content filtering can be implemented by using either hardware- or software-based solutions, as well as tracking, and regulating access to websites that users are not supposed to enter. The reason for that can be because

[2] https://bit.ly/35zBP7B.

[3] Please refer to NIST 800-63 (https://bit.ly/3HQUMzG) for further guidance.

the content is subject to compliance regulations – as is often the case with material on gambling websites – or is suspected to be malicious content. Domain name system (DNS) filtering, meanwhile, is a special type of content sifting that uses the DNS layer to regulate website access based on IP addresses, in order to filter web use and reduce malware exposure.

- **Why should this control be adopted?** Malicious links and files are the primary way to insert malware into organizations' systems, or to steal user passwords, and eventually access critical systems. Web and email filtering is seen as a "first line of defense" in defending against email- or web-browsing-related cyberattacks, even before the users – the "second line of defense" – can fall victim to a phishing attack or enter websites with malicious content. Filtering is needed as email phishing is one of the top initial attack vectors leading to severe cyber incidents, especially ransomware attacks. Cybercriminals often use phishing campaigns to steal their victims' usernames and passwords, which provide the attackers with initial access to a victim's IT environment. By implementing email security and web-filtering technologies, a large percentage of potentially severe cyberattacks can be stopped at the outset. At a minimum, organizations should pre-screen emails for potentially malicious attachments and links, and to use tools to monitor web content to block access to vulnerable websites.

  Insurers are also imposing stringent cyber resiliency requirements on policyholders to evaluate them as insurable. The implementation of email security and web-filtering technology will allow an organization to improve its profile in relation to presenting its cyber risk to insurance underwriters. These controls are a key element of eligibility for cyber insurance cover as, if they are implemented, insurers predict seeing a decrease in the total number of severe – and therefore costly – cyber incidents.

- **What a company needs to do to put this control in place:** security controls related to malware protection, email security, and web-filtering that could be put in place can encompass the following: (1) using technology to scan and filter incoming emails for malicious attachments and links; (2) preventing macro-enabled files from running by default; (3) evaluating email attachments in a sandbox environment prior to user delivery, in order to determine whether files are malicious; and (4) using technology to monitor web content and to block access to malicious websites or web content.

2.3 Control 3: Secured, encrypted, and tested backups

- **What is this control?** Secure, available, and accurate backups are essential to ensure business resilience. Backups should be secured, preferably by isolating them from the network, or by implementing multifactor controlled access and encryption. Regular testing is also critical to ensure the integrity and availability of data.

- **Why should this control be adopted?** As organizations increasingly move to cloud-based backup solutions, attackers look for administrator credentials to gain access to them, before deleting or encrypting them. A lack of available backups increases the likelihood of a victim paying a ransom, in order to recover systems and data, as they have no other option. Regularly testing backups is critical – there is no point in having backups if they are unavailable, or incomplete, when you need to restore your systems. Regular tests also enable IT and business resilience teams to understand the complexity of the restoration process and identify external partners that may be required to assist them. It is not usually as simple as flicking a switch.

  Viable backups enable organizations to recover from attacks more quickly and effectively. In the case of ransomware, having backups reduces the leverage that threat actors have over the victim and can greatly reduce the need to pay a ransom. Where systems are encrypted, businesses are usually unable to operate and so incur significant business interruption losses. Secured backups can reduce recovery time and enable a return to business as usual more quickly than negotiating with threat actors for dubious decryption keys. Tested backups enable the business to place more trust in the backed-up data – errors or failures in the backup process will be picked up and rectified quickly.

- **What a company needs to do to put this control in place:** organizations should review their critical systems and assets, and ensure that backup procedures are adequate and tested regularly. It also is essential to ensure that one copy of the backup is stored offline and is unconnected from the network.

  Disaster recovery, business continuity, and incident response plans should be put in place to accurately document the process that would be taken to recover systems from backups. There are myriad different backup solutions and it can be difficult to assess the best provider and proposition for an organization. Focusing on providing a solution for the systems, data, and assets that are truly critical – the "crown jewels" – is a good place to start.

2.4 Control 4: Privileged access management (PAM)

- **What is this control?** Privileged access management (PAM) is a security technology that offers an elevated or "privileged" level of access to protect accounts, credentials, and operations. Privileged access differs from "normal" access because it can allow security or maintenance functions, system- or application-wide configuration changes, and the bypassing of established security controls through super user access.
- **Why should this control be adopted?** In terms of cybersecurity, humans are often the weakest link, making any organization vulnerable to an attack. PAM tools control privileged access of machines (systems or applications) for internal or machine-to-machine communication, including for people who administer or configure systems and applications. It runs on the principle of "least privilege", meaning the users only receive the minimum level of access required by them to perform their job functions. Components within a typical PAM solution monitor sessions that are used by administrator accounts and generate alerts for any anomalous session usage. Anomalies may include an account trying to access areas outside of its responsibility domain or outside of its window of operations.
- **What a company needs to do to put this control in place:** at the outset, an organization needs to identify the use case – that is, the actions or event steps it wants to invest in a PAM for. For example, it can adopt a risk-based approach to identify critical assets that are at the highest risk of exposure, as a result of the compromise of privileged accounts, and then only implement the solution for those assets. Once PAM is in place, to overcome any misconception about the solution, an organization can distribute content to its employees on its different components, their purpose, and why they are required as part of the overall cybersecurity mix. An organization also should establish a governance and monitoring program for PAM so that performance does not degrade over time. This should include setting selection and performance criteria for vendors and products and conducting post-implementation performance evaluations. Regarding the scalability of PAM, roadmaps for business growth can factor in additional relevant assets requiring this control, so that licenses are available to accommodate them when implemented.

2.5 Control 5: Endpoint detection and response (EDR)

- **What is this control?** Endpoint detection and response (EDR) is a threat detection and response mechanism for an endpoint – a remote device such as a desktop, laptop, mobile phone, server, or internet of things (IoT) that communicates with an internal network, externally. As endpoints are the entry points for virtually any type of malicious attack on a network, their monitoring is vital to detect and stop a strike before it spreads to the wider internal network. An EDR solution continuously monitors endpoints, collects data from devices, and provides a response based on defined rules.
- **Why should this control be adopted?** According to a study published by the Ponemon Institute in 2020,[4] 68 percent of organizations have experienced one or more endpoint attacks that successfully compromised data and/or their IT infrastructure. The same report noted that 68 percent of IT professionals found that the frequency of endpoint attacks had increased since the previous year. Monitoring of endpoints is critical to detect and stop an attack before it spreads to the wider internal network. Additionally, when EDR is in place, it monitors and records activity on those endpoints. That data can be analyzed to detect persistent threats, or "zero day" vulnerabilities – in other words, flaws not yet patched – that have been compromised. If a security threat is detected, the log can be reviewed to determine when that threat began, the scope of the compromise, and the root cause. For example, an organization that is the victim of a ransomware attack will take longer to recover without EDR in place. This is because it will not have visibility into the extent of the event, and specifically, on how many endpoints have been infected. The organization will be unable to detect if there are any payloads still operating on the backend, and if relevant configuration settings are working as expected, or need to be implemented. All of this means that the recovery effort will take longer and need to be more in depth.
- **What a company needs to do to put this control in place:** having a strong baseline of cybersecurity best practices usually enables an organization to implement EDR seamlessly. It is also vital to find an EDR solution that can provide the maximum level of protection while requiring the least amount of effort and investment,

[4] https://bit.ly/3HRpUPt.

ultimately adding value to your security team without demanding a lot of resource. Key aspects organizations should look for in a solution include: (1) endpoint visibility across all your endpoints (it should provide real-time visibility for you to view suspicious activities, even as they attempt to breach your environment, and stop them immediately); (2) a solution that collects a significant amount of telemetry from endpoints, so it can be mined for signs of attack with a variety of analytic techniques; (3) Effective endpoint detection and response requires behavioural approaches that search for indicators of attack (IOAs), so you are alerted of suspicious activities before a compromise can occur; (4) a solution that integrates threat intelligence, including details on the attributed adversary that is attacking you or other information about the attack; (5) a quick-response, solutions should operate in real-time, provide accurate alerting, and automate threat response (this requires detection engines that produce minimal false positives and the ability to set automated response policies); and (6) having a cloud-based endpoint detection and response solution is the only way to ensure zero impact on endpoints (this solution should smoothly integrate with current systems and provide intuitive remote access to controls).

2.6 Control 6: Patch and vulnerability management

- **What is this control?** Vulnerability management is a capability that identifies vulnerabilities on software and hardware devices that are likely to be used by attackers to compromise a device and use it as a platform from which to further compromise the network. Patch management is the systematic notification, identification, deployment, installation, and verification of an operating system and application of software code revisions. These revisions are known as patches, hot fixes, and service packs.

  Not all vulnerabilities have related patches. Consequently, a proper vulnerability management process will consider other methods of remediation, or temporary workarounds – such as software configuration change and employee training – to limit or isolate the exposure.

- **Why should this control be adopted?** Organizations will always have a certain level of risk due to vulnerabilities in their IT environments. A risk can be defined as the probability that a particular security threat will exploit vulnerability in a system.[5] Vulnerabilities can be exploited by several cyber threats in software and hardware devices that are frequently used by an organization. Consequently, without having a clear and continuous view of existing vulnerabilities, organizations will struggle to identify and respond to threats in a timely manner.

  On the other hand, each organization will have a unique risk tolerance based on its financial health, reputation exposure, and compliance requirements. Establishing a relationship between proper IT vulnerability management and risk tolerance is complex and can be hard to advocate for in front of a board, as management may not fully understand which IT vulnerabilities present the greatest risk. A proper patch and vulnerability management function will reduce, or eliminate, the potential for exploitation and involve considerably less time, effort, and money than the response following an exploitation. Unpatched vulnerabilities remain a leading cause of intrusions into systems, with hundreds of vulnerabilities revealed every month for multiple applications and systems. When technology environments are not patched in a timely fashion, attackers will seek to exploit vulnerabilities.

- **What a company needs to do to put this control in place:** while the implementation of a vulnerability management process is very complex, it can be summarized in five steps:[6] (1) **preparation**: conduct a vulnerability analysis, define the scope of assets, inform stakeholders and asset owners, and plan vulnerability scans; (2) **identification and detection of vulnerabilities**: this can be achieved through a vulnerability scan; (3) **definition of remediating actions**: to properly define the remediating actions, an IT risk assessment must be conducted (depending on the remediation, such as a patch or a change in configuration, software restrictions, and availability of solutions, different options can arise including mitigate (by implementing remediating actions) or accept (by launching an exception process[7] and investigating potential indicators of compromise (IOC)); (4) **implementation of defined actions**: deployment of the tasks identified in the previous

[5] According to NIST SP 800-16. However, there are multiple risk definitions. Although we have included the simplest way to define a risk, this definition considers the most important characteristics: probability of occurrence, threat or event, vulnerability, and impact.

[6] These steps are enlarged upon in the SANS Institute paper: “Implementing a vulnerability management process,” https://bit.ly/3Csg6u3.

[7] Exception process: a condition that is not aligned with formal security expectations as defined by policy, standard, and/or procedure – for example, a patch is not applied.

activities; and (5) **monitoring of vulnerabilities**: as new vulnerabilities arise every minute, committing to real continuous monitoring is essential to properly manage them.

Most vulnerabilities can be remediated by patching and updating systems. IT systems with known vulnerabilities, with constraints to install the patch, or without an available solution, can be investigated for the presence of indicators of compromise. And, if these systems are compromised, incident response and recovery plans should be initiated. In addition, an isolation plan for legacy systems can also be applied. Ultimately, if the IT risk assessment indicates a delay in remediation due to operational constraints, an exception process should be in place.

Insurance companies are also likely to require the following actions: periodic performance of a vulnerability analysis; performance of penetration testing (that is, a simulated cyberattack to check for exploitable vulnerabilities – at least annually); ongoing maintenance and updating of the information technology and communications landscape; patches with CVE[8] 8 or above to be applied in less than three to seven days, after their publication, on exposed IT systems; and non-critical patches are expected to be applied in less than 30 days after their publication.

The frequency and detail of these actions is tied to an organization's cyber risk profile, industry, and cyberattack environment.

2.7 Control 7: Incident response plans

- **What is this control?** Incident response plans document a "predetermined set of instructions or procedures to detect, respond to, and limit consequences of a malicious cyberattack against an organization's information systems."[9] They need to be in line with other available related plans and capabilities, including: an IT disaster recovery plan (DRP), which describes how an organization recovers data during and after a crisis or disaster; and a business continuity plan (BCP), which sets out how an organization ensures that essential business processes are available during and after a crisis or disaster.

  Incident response will only work smoothly when all relevant stakeholders are familiar with the response plan. Consequently, regular testing of procedures is an essential part of this control.

- **Why should this control be adopted?** Incident response plans are an integral part of increasing an organization's cyber resiliency. These programs are not isolated frameworks – they need to reflect the specific and unique risk profile of an organization and require integration within an overall cyber risk management strategy.

  In order to mitigate cyber risk, the first approach and line of defence will always be technical and organizational prevention measures. When cyber incidents do occur, it is crucial to detect them as early as possible, and respond to them in a fast and professional way. An up-to-date incident response plan and a trained team provides efficiency, speed, and quality in response to cyber incidents. When combined with a holistic organizational approach to the 12 key controls – as well as the implementation of appropriate technical controls and incident and disaster recovery (such as secured, encrypted, and tested backups) – an incident response program significantly helps to mitigate the impacts of a cyber event on operations and an organization's reputation.

- **What a company needs to do to put this control in place:** organizations are advised to encompass the following core capabilities in their approach to incident response planning and testing: (1) the incident response plan must contain defined processes and procedures for performing cyber incident handling, reporting, and recovery; (2) the incident response team members' roles, tasks, and responsibilities during a security incident must be clearly defined (additionally, strong definitions of escalation paths and decision-making processes/responsibilities are obligatory); (3) the parts of incident response that will be covered externally (such as IT forensic investigations) should be planned and documented, and the relevant contact information noted; (4) due to the significant uptick in ransomware incidents and their enormous loss potential, a specific response playbook tailored to the ransomware crisis scenario should be defined; (5) incident response plans are only valuable when the response team members are familiar with their roles and responsibilities, and when there is clarity on the underlying processes (an annual table top exercise should be conducted to train the team for specific scenarios, and to evaluate an organization's incident preparedness); and

[8] CVE stands for common vulnerabilities and exposures. It is a program launched by MITRE to identify and catalog vulnerabilities in software or firmware.
[9] As defined by the Computer Security Resource Center, https://bit.ly/35ZqwFi.

(6) the plans need to be reviewed and updated periodically, incorporating recent developments, such as staff changes and new anticipated threats.

2.8 Control 8: Cybersecurity awareness training and phishing testing

- **What is this control?** Cybersecurity awareness training is a control used to educate employees and IT users on cyber risks and threats. It helps them identify and recognize the various attacks, and equips them with the necessary information on how to protect themselves and their organizations by preventing events in the first place, and doing the right thing after an attack or an attempted breach.

  Phishing testing is part of a security awareness training program that simulates phishing attacks by sending bogus, but very realistic, phishing emails to employees to measure their awareness. It tests the effectiveness of security awareness training, by evaluating employee reaction to the emails, and determines the behaviors that require further improvement.

- **Why should this control be adopted?** Business is about people, process, and technology. Investing in securing process and technology is very important, but insufficient, if the human aspect is ignored. Businesses are operating in a world in which 95 percent of cybersecurity issues can be traced to human error.[10]

  Despite very advanced IT security, human factors such as workload, stress, lack of skillset, the increased use of the hybrid working model, and basic human nature can all lead to human error. However, this weakest link of the security chain can turn into the best layer of defense, when it gets the right focus and attention. The human element is also a concern for regulators. Some regulations – including, but not limited to, Payment Card Industry Data Security Standard (PCI-DSS), National Institute of Standards and Technology (NIST), HIPAA, and General Data Protection Regulation (GDPR) – may require employees to undergo regular security awareness training. In order to establish a secure culture, make people part of the cybersecurity program, comply with regulations, and ultimately protect an organization from the impacts of a possible cyber incident, cybersecurity awareness training and phishing testing have become extremely important.

- **What a company needs to do to put this control in place:** organizations should take the following actions when establishing cyber awareness training:

  1. Perform an annual analysis to identify gaps in their cybersecurity skillset and develop and implement training roadmaps and/or project plans to close identified gaps.
  2. Establish annual (at a minimum) cybersecurity training and a cybersecurity awareness program that:

  – Are mandatory for all employees, vendors/contractors, and third-party partners with access to the corporate network.

  – Train users to avoid common cyber risks and threats, such as social engineering and phishing.

  – Provide frequent – at least annual – updated content to embody the latest attack and social engineering techniques.

  3. Conduct, at least annually, internal phishing campaigns.
  4. Have a process to report suspicious emails to an internal security team to investigate.
  5. Have a process to respond to phishing campaigns.
  6. Tag external emails to alert employees that the message originated from outside the organization.

  NIST in the U.S. also focuses on security awareness and training under the "protect function" of its cyber framework. For further guidance please see: "NIST Special Publication 800-50".[11]

2.9 Control 9: Remote desktop protocol (RDP) mitigation and other hardening techniques

- **What is this control?** Hardening is the process of applying security configurations to system components including servers, applications, operating systems, databases, and security and network devices, in line with best practices. These configurations are defined in order to reduce an organization's surface attack by limiting the exposure of each platform on the internal network or that may be facing to the internet.
- **Why should this control be adopted?** Through hardening techniques, companies can minimize their attack surface by disabling unused or insecure services, mitigating vulnerabilities, and improving weak configurations that could be used by malicious actors to compromise their systems.

[10] Mee, P., and R. Brandenburg, 2020: "After reading, writing, and arithmetic, the fourth 'r' of literacy is cyber-risk," World Economic Forum Global Agenda, December 17, https://bit.ly/3HSrwlA.

[11] https://bit.ly/3HUL1QV.

- **What a company needs to do to put this control in place:** normally, organizations define a set of secure configurations for their main systems and services, based on best practices, commonly known as security baselines or hardening guides. A process is implemented to deploy these configurations, and review them periodically, in order to identify any misconfiguration or deviation. Although they vary between each platform, the configurations that commonly are part of these security baselines may include the following: user and access management; password policies; secure services and protocols; firewall configurations; network configurations; remote access; log management and audit policies; antivirus/antimalware protections; application control; security updates; encryption; and other platform-specific security configurations.

  To ensure the timely deployment of these configurations, organizations may use images of systems with security configurations or tools already applied and then perform a gap analysis periodically. An important topic that insurers are concerned about is the exposure of weak or commonly attacked protocols or services to the internet, such as remote desktop protocol (RDP), server message block (SMB), secure shell (SSH), file transfer protocol (FTP), and database ports. Organizations need to have a strict hardening process in order to eliminate the usage of these kinds of ports exposed to the internet. If they are needed, as a result of a specific business requirement, organizations should implement compensating controls to mitigate the associated risk.

  One of the most common barriers to implementing a hardening process is the absence of a comprehensive asset inventory, providing an organization with detailed knowledge of the technologies in place on the network, which may be supporting critical processes. Organizations are advised to define a structured change management process to deploy these security baselines. Without a proper process, some of these arrangements may affect the availability of the systems by disabling configurations that are required at the moment of deployment. They may need a deeper analysis in order to find a secure method to function or may even require a change on an application. Today, vendors and cybersecurity organizations are constantly releasing security baselines for the most common systems and services. The Center of Internet Security (CIS)[12] is one of the most important sources of baselines that all organizations can access.

2.10 Control 10: Logging and monitoring

- **What is this control?** In order to react to a cyberattack in a timely manner, organizations need to establish strong logging and monitoring capabilities that enable them to identify any suspicious activity on the network. These capabilities require specific knowledge, tools, and processes to be able to detect malicious activities. All of these factors are normally executed by a security operations center (SOC) or an external managed security service provider (MSSP). The SOC or MSSP may include different capabilities, depending on their level of maturity and sophistication.
- **Why should this control be adopted?** The current global threat landscape requires companies to not only implement a set of controls in order to protect their organizations from a cyberattack, but also to identify any suspicious activity that may indicate a potential attack in progress in a timely manner and that could trigger a cyber-incident response plan. This can only be performed by an adequate logging configuration on the main systems and applications of the company, the appropriate tools to collect, correlate, and alert in case of a situation, as well as the right team capable of analyzing and acting in case of an incident.
- **What a company needs to do to put this control in place:** companies are recommended to:
  1. Outline and implement the audit logs and systems or platforms to be monitored, including firewalls, intrusion prevention systems and intrusion detection systems, active directory, antivirus/antimalware, endpoint security technologies such as EDR and XDR, data loss prevention (DLP), applications, Microsoft 365, and other important platforms defined by the organization.
  2. Implement a security incident and event management system (SIEM) and integrate the main platforms into this system. Logs should be accessible for at least the last three months and backed up for a minimum of one year.
  3. Analyze the logs in the network and define a set of use cases or common patterns that the organization would like to monitor and react to, in the instance that they are found. The information should also be used alongside threat intelligence information.
  4. Define processes for reviewing, periodically, the administrators' or high-privileged users' activities on critical systems.

[12] https://bit.ly/3HMnFNr.

5. Define and train a team of professionals specialized in the monitoring of security events and incident response.

– Specific processes or playbooks should be defined in order for the SOC and MSSP to react if a cybersecurity incident is detected. If this service is outsourced, these procedures should also include the tasks that the organization would need to execute in order to contain, eradicate, and restore the operations to normality.

– Define and monitor key performance indicators for continuous improvement.

The development of adequate logging and monitoring capabilities may require significant investment and resources from the organization. In addition, establishing these capabilities requires continuous review to ensure that the processes put in place are able to detect suspicious activities in real life scenarios.

2.11 Control 11: Replacement or protection of end-of-life (EOL) systems

- **What is this control?** End-of-life (EOL) or end-of-support (EOS) products are those that reach the end of their lifecycle, preventing users from receiving updates. These products create risk because patches and other forms of security support are no longer offered by the vendor. Once the technology is unsupported, it will be exposed to unfixable vulnerabilities.

  The only fully effective way to mitigate this risk is to stop using the obsolete product and replace or upgrade it with a newer solution that continues to provide support. Where this is impossible, EOL/EOS systems will need to be protected by compensating controls, such as restricting access to those systems, ensuring they are not internet facing, and are "air gapped" – that is, physically isolated from other connected systems. EOL/EOS products and systems are often used by organizations with large legacy estates, particularly where systems are used to control operational technology (OT), which can be difficult and costly to upgrade regularly.

- **Why should this control be adopted?** Vulnerabilities in EOL/EOS products will remain unpatched and become increasingly exploitable by hackers looking for easy ways to gain access to systems. Known vulnerabilities are openly discussed on forums, and hackers are able to scan easily for EOL systems that continue to be in use. While open ports and email phishing remain popular attack vectors, known software vulnerabilities are also a common entry point, offering an easy route into systems. Once inside, hackers will try to gain access throughout a network, looking for valuable data to steal and systems to encrypt.

- **What a company needs to do to put this control in place:** ideally, organizations should stop using any obsolete products. If this is unfeasible, it is essential to ensure that legacy systems are protected. Limiting access to these products from outside the environment is a critical step – if attackers cannot reach a device, the risk of exploitation is significantly reduced. Where possible, network air gaps should be implemented. If this is not possible, a discrete network firewall and monitoring of data flows to obsolete servers should be considered. A good rule of thumb is to treat all access from the internet as untrusted. Steps can also be taken to limit the potential impact of compromise, such as preventing those EOL systems from accessing or storing critical and sensitive data or systems, meaning that a compromise of the EOL device would not be as damaging.

  Upgrading EOL systems and products will come with a potentially hefty price tag. For organizations with significant legacy estates and operational technology systems, an EOL product may mean that the whole system needs to be overhauled, upgraded, or replaced. Where organizations opt to continue to use the EOL product, the necessary protection and risk mitigation steps will require thorough implementation and will typically necessitate the collaboration of both the IT and OT security teams, and may also call for external expertise and tools. For manufacturers and other organizations with extensive OT systems, this implementation can be complex and time-consuming.

2.12 Control 12: Digital supply chain cyber risk management

- **What is this control?** The digital supply chain encompasses all information technology (IT) and operational service (OT) providers that together with an organization's teams deliver digital services. In terms of cyber risk, the digital supply chain poses an increasing challenge. There have been instances of various large digital supply chain vulnerabilities having major effects – for example, the recent Log4J[13] and Kaseya[14]

[13] https://bit.ly/3pObIAr.
[14] https://bit.ly/37hJubb.

vulnerabilities, or breaches such as the hacking campaign Operation Cloud Hopper.[15] Even the infamous NotPetya[16] attack resulted from a digital supply chain risk. A digital supplier can present the perfect entry point to hundreds of companies and their sensitive data. By successfully breaching a vulnerability within one single digital supplier, cyber criminals can gain access to a multitude of their clients' networks and devices. A robust framework for managing digital supply chain cyber risk is required.

- **Why should this control be adopted?** The continuously increasing digitalization and the use of information and communications technology to deliver critical functions have introduced new aspects of cyber risk that need to be managed. Organizations are consuming new and different digital services from various service providers, offering software packages to complete outsourced and software-as-a-service[17] products. At the same time, the fact that supply chains have become so global has created new risk, in terms of confidentiality, integrity, and availability.

  Given the pervasiveness of digital services, it is becoming increasingly complex to manage the digital supply chain. IT teams may not be aware of all services consumed by the organization, caused by the issue of "shadow IT",[18] but also, it is not always apparent what exactly constitutes a service and the potential vulnerabilities embedded in it. Cyber criminals use these digital supply chains as a mechanism for cyberattacks. Indeed, most software products rely on thousands of prewritten packages produced by vendors. The most commonly used third-party software supply chain components are highly-prized targets for cyber criminals, as breaching one digital service provider can allow access to its many customers. Hence, this control aims to protect the cyber risk heritage from digital suppliers by a set of activities focused on analyzing, managing, and responding to the cyber risk.

- **What a company needs to do to put this control in place:** organizations are advised to consider the following actions to manage digital supply chain risk:
  - Adopt a digital supply chain risk management framework, including risk rating of first-tier vendors/suppliers, based on an advanced risk quantification. This will help an organization take strategic decisions on risk management and capital allocation.
  - Implement a cybersecurity framework. This can include, but is not limited to:
    - Account management based on "zero trust" expectations and the "need-to-know" principle. Strict limitations of privileged and generic accounts apply.
    - Enforced appropriate risk-based multifactor authentication (MFA).
    - Engagement with the internal security operations center to develop specific use cases for monitoring third party accesses.
  - Develop and test an incident response playbook for vendor/digital supply chain scenarios and include third parties in this playbook.
  - Assess contracts, service agreements, and escalation protocols for each vendor or digital supplier.
  - Engage with the procurement department to include appropriate cybersecurity hygiene controls and responsibilities in new contracts and renewals. This can include security trainings and certifications.

3. CONCLUSION: WHERE TO FROM HERE

At Marsh, we believe the new cyber risk paradigm requires organizations to become more comfortable with the reality that the connective tissue of modern business is digital. As such, organizations need to adopt new methods of understanding, measuring, and managing cyber risk on a continuous basis. With discipline, foresight, and agility to shift focus, organizations can achieve improved outcomes, as we collectively embrace new cyber risks.

When an organization implements the recommended controls, they will either prevent, or be equipped, to respond to the majority of cyberattacks, in a way that minimizes their impact. They will be well prepared to defend themselves and feel more confident with their cyber resiliency. Given the current cyber landscape and the increasing threat to every organization, cyber resiliency can no longer be an afterthought or tick-the-box exercise – it has become a minimum requirement.

[15] https://reut.rs/3J9pVjn.

[16] https://bbc.in/3hQGWCA.

[17] Software-as-a-service (SaaS) is a software distribution model in which a cloud provider hosts applications and makes them available to end users over the internet.

[18] Shadow IT refers to information technology programs, projects, or systems implemented outside of the IT or information security departments

EUROPE'S PUSH FOR DIGITAL SOVEREIGNTY: THREATS, E.U. POLICY SOLUTIONS, AND IMPACT ON THE FINANCIAL SECTOR

LOKKE MOEREL | Professor of Global ICT Law, Tilburg University*

ABSTRACT

The European Union (E.U.) feels the threat of what is coined digital colonialism of the U.S. and China,[1] where the E.U. member states are increasingly dependent on digital infrastructures that are in the hands of a handful of dominant foreign market players. The digital identity of most European citizens depends on foreign email addresses, and a staggering 92 percent of European data reside in the clouds of U.S. technology companies, of which 80 percent are with five suppliers only.[2] Besides supply chain dependencies, these companies operate proprietary ecosystems, which offer limited interoperability and portability of data and applications, resulting in E.U. data being locked-in and having limited value for E.U. innovation.[3] Restoring Europe's "digital sovereignty" is now a core ambition of the European Commission (E.C.); however, achieving it at a time when digital technologies have become the battleground for the race for global leadership between the U.S. and China (aka the tech cold war) will not be easy. Both the U.S. and China regularly draw the national security card to justify stricter export controls of critical technology and bringing manufacturing back to their countries. Recent U.S. executive orders ensure that almost any ICT-related activity in the U.S. connected to China is now subject to regulatory review by the U.S. government. Not surprisingly, China is retaliating.

With the E.U. policy measures, the E.C. is aiming to pave a third way, in order to avoid falling into the trap of tech protectionism. Flagship initiatives discussed are the so-called European Data Spaces (bringing together E.U. data of specific industry sectors in order to unlock their value for E.U. innovation) and the GAIA-X project (achieving interoperability between cloud offerings to achieve the required scalability for AI-related innovations, without setting up European hyperscalers). All initiatives will also have a fundamental impact on the business models of the financial sector. This article discusses the threats to E.U. digital sovereignty in order to help the reader better understand the E.U. policy proposals and their disruptive impacts, which – as with any regulation – brings new requirements, but also opportunities for innovation.

* Lokke Moerel is also a member of the Dutch Cyber Security Council. This article is based on an earlier article: Timmers, P., and L. Moerel, 2020, "Reflections on digital sovereignty," E.U. Cyber Direct, January 15, https://bit.ly/3s7sz2K, originally written in assignment of the University of Utrecht 2020 Annual Constitutional Law Conference: Constitutional law in the data society.

[1] Kwet, M., "Digital colonialism: US empire and the new imperialism in the global south," Race & Class 60:4, 3-26.

[2] Amiot, E., I. Palencia, A. Baena, and C. de Pommerol, 2020, "European digital sovereignty: syncing values and value," Oliver Wyman, https://owy.mn/3LOpGf7.

[3] Digital Services Act package, Inception Impact Assessment, https://bit.ly/34TSe6u.

1. INTRODUCTION

Europe is one of the most digitalized societies and this has been accelerated by the COVID-19 pandemic.[4] Within no time, people worked from home and children were schooled online. It was amazing to see how quickly we were up and running again. However, as we become increasingly digitized, the vulnerabilities that come with it also increase. 2020 saw a 70 percent increase in internet-related crime, including COVID-19 scams,[5] a 150 percent increase in ransomware attacks exploiting work-from-home technologies,[6] hostile states trying to steal our COVID-19 research,[7] China and Russia pushing "fake news" to undermine our governments' COVID-19 responses,[8] and difficult-to-combat online conspiracy theories of anti-5G movements, stimulated by Russian infiltration.[9]

By now, the realization has set in that Europe's digital dependencies are so great that the digital sovereignty[10] of the E.U. and its member states is under pressure. The fears are justified, E.U. sovereignty (as the sovereignty of any state around the world for that matter) is under pressure due to a toxic combination of disruptive digital transformation (with winner takes all suppliers), exponential growth of cyberattacks (in which smaller countries and non-state actors now also enter the global battlefield), and rising geopolitical tensions, leading to a sovereignty gap.[11] Where at first digital sovereignty was discussed in the context of cybersecurity, military, and defense, the discussion now extends to concerns about the economy and society at large. The ultimate challenge is how Europe and its member states can retain control over their **economies** (control over essential economic ecosystems) and their **democracies** and **the rule of law** (trust in their legal system and quality of democratic decision-making) in the digital world.[12] Due to the multifaceted nature of the causes of the pressure on our digital sovereignty and rapid geopolitical developments, there is no one-size-fits-all solution. To be able to understand the series of E.U. policy initiatives to restore Europe's digital sovereignty, it is important to understand why Europe's ability to take decisions autonomously is under threat.

> *"We must have mastery and ownership of key technologies in Europe. These include quantum computing, artificial intelligence, blockchain, and critical chip technologies. (...) We need infrastructure fit for the future, with common standards, gigabit networks, and secure clouds of both current and next generations."*
>
> **Ursula von der Leyen** – *inaugural speech as president-elect European Commission (2019)*

Sovereignty is a political concept for which there is no generally accepted definition. Sovereignty is generally associated with territoriality, jurisdiction, a population, and authority with both internal and external recognition (legitimacy).

Internal legitimacy refers to the effectiveness of the state when executing governmental tasks (e.g., being in control of the electoral process and the criminal justice chain) and also the recognition by citizens of the government (having confidence in the rule of law).

External legitimacy concerns the recognition by foreign states and the autonomy of action toward such foreign states.

Strategic autonomy: if sovereignty is the goal, strategic autonomy is the means, i.e., the capabilities to decide on key aspects of the long-term future in the economy, society, and democracy.

[4] The European data economy continues to grow rapidly – from €301 bln (2.4 percent of GDP) in 2018 to an estimated €829 bln (5.8 percent of GDP) by 2025. IDC, 2020, "The European data market monitoring tool key facts and figures, first policy conclusions, data landscape and quantified stories," final study report, https://bit.ly/3BDBRGQ.

[5] FBI National Press Office, 2021, "FBI Releases the Internet Crime Complaint Center 2020 Internet Crime Report, Including COVID-19 Scam Statistics," Federal Bureau of Investigation, March 17, https://bit.ly/3p62T4V.

[6] https://bit.ly/3p29NrG.

[7] Grierson, J., and H. Devlin, 2020, "Hostile states trying to steal coronavirus research, says UK agency," The Guardian, May 3, https://bit.ly/3s8mLpN.

[8] Scott, M., 2020, "Russia and China push 'fake news' aimed at weakening Europe: report," Politico, April 1, https://politi.co/3LOate7.

[9] Lynas, M., 2020, "Anti-vaxxers and Russia behind viral 5G COVID conspiracy theory," Alliance for Science, April 8, https://bit.ly/3BGv0wj.

[10] For definitions see: Timmers, P., 2019, "Strategic autonomy and cybersecurity," E.U. Cyber Direct, May 10, https://bit.ly/3v67gAu.

[11] Timmers, P., 2019, "Challenged by 'digital sovereignty,'" Journal of Internet Law 23:6, 1, 18.

[12] See for in-depth discussion see Timmers, P., and L. Moerel, 2020, "Reflections on digital sovereignty," E.U. Cyber Direct, January 15, https://bit.ly/3s7sz2K.

2. WHAT ARE THE THREATS?

2.1 Disruptive digital transformation

Friends and foes agree that our society is undergoing a digital revolution (in official terms: the fourth industrial revolution) that will lead to a transformation of our society as we know it.[13] In addition to all economic and social progress and prosperity, every technological revolution also brings with it disruption and friction. The first law of technology is that it is not good, not bad, but also not neutral.[14] The new digital technologies (and, in particular, artificial intelligence (AI) and quantum computing) are in and of themselves already disrupting societies and create new vulnerabilities. Weakening control over innovation and knowledge can jeopardize sovereignty. For example, AI and encryption will play an increasingly crucial role in cyber resilience.[15] If there is not enough innovation, there will be new dependencies.

Current E.U. research investments in quantum computing and AI are dwarfed by the billions invested by the Chinese and U.S. governments,[16] combined with the investments from large U.S. and Chinese tech companies, such as Google[17] and Tencent.[18] Where foreign companies are at the forefront of (further) development and implementation of new technologies, such as AI and quantum computing, but also satellite and 5G networks, potentially new dependencies arise. These dependencies go beyond the specific technological applications themselves. For example, to be able to make large-scale use of data analysis by means of AI, enormous computing power is required. It is expected that the cloud infrastructure required for this will become the foundation for the European innovation and knowledge infrastructure. Maintaining control over this is an essential part of the E.U.'s digital sovereignty.[19]

EXAMPLE: AI AND CRYPTOGRAPHIC TECHNOLOGIES

With AI, bad actors can detect and exploit vulnerabilities automatically and on a large scale. However, AI is also expected to make it possible to automatically detect and restore vulnerabilities in software. We will, therefore, have to innovate to be able to keep ahead of bad actors.

Without proper encryption, we will not be able to protect the valuable and sensitive information of our governments, companies, and citizens. Current encryption will not hold against the computing power of future quantum computers. We will, therefore, have to innovate now to protect our critical information in the future. This is not only relevant for future information, but also for current information. Do not forget that currently hostile states systematically intercept and preserve encrypted communications in anticipation that these may be decrypted at a later stage and analyzed by deploying AI. We, therefore, have to invest in post-quantum encryption now in order to be able to protect strategic information that requires long-term protection.

2.2 Increasing cybersecurity threats

An important dimension of digital sovereignty is the cyber resilience of our critical sectors, processes, and data. The ever-increasing cybersecurity threats – in which smaller countries and non-state actors are now also entering the global battlefield[20] – undermine our digital sovereignty. These concern the entire spectrum of direct threats to our vital infrastructure (sabotage), systematic theft by foreign states of intellectual property from our knowledge-intensive industries (economic espionage), digital extortion (ransomware attacks), targeted misinformation (fake news), and systematic infiltration of social media to influence elections and democratic processes.

[13] For an accessible book, see Brynjolfsson, E., and A. McAfee, 2014, Second machine age: work, progress, and prosperity in a time of brilliant new technologies, W.W. Norton & Company, which gives a good overview of the friction and disruption that arose from the industrial revolution and how society ultimately responded and regulated negative excesses and a description of the friction and disruption caused by the digital revolution. A less accessible, but very instructive, book, on the risks of digitization and big tech for society is Zuboff, S., 2019, The age of surveillance capitalism, Public Affairs, [hereinafter: Zuboff (2019)].

[14] Kranzberg, M., 1986, "Technology and history: 'Kranzberg's laws'," Technology and Culture 27:3, 544-560.

[15] Van Boheemen, P., L. Kool, and J. Hamer, 2019, "Cyber resilience with new technology – opportunity and need for digital innovation," Rathenau Instituut, July 20, https://bit.ly/3LN7YsB. See also the Dutch Cyber Security Council Recommendation, 2020, "Towards structural deployment of innovative applications of new technologies for cyber resilience in the Netherlands," CSR Opinion 2020, no. 5, p. 3.

[16] See for an overview of U.S. and Chinese research investments, Smith-Goodson, P., 2019, "Quantum USA vs. quantum China: the world's most important technology race," Forbes, October 10, https://bit.ly/3sWJowv.

[17] In October 2019, Google claimed to have reached quantum supremacy with its Google quantum computer called Sycamore (https://go.nature.com/3JIJ9vL). On December 3, 2020, Chinese quantum computing researchers also claimed quantum supremacy (https://bit.ly/3vckY4W).

[18] Keen not to fall behind major U.S. tech firms in quantum computing, the Chinese company Tencent announced that it plans to invest U.S.$70 bln in infrastructure and quantum computing (https://bit.ly/3s7RkMc).

[19] Timmers, P., 2020, "There will be no global 6G unless we resolve sovereignty concerns in 5G governance," Nature Electronics 3, 10-12. See also the German "Industrial strategy 2030. Guidelines for a German and European industrial policy," (https://bit.ly/3t1c7Am) in which it is recognized that insufficient grip on new technologies poses a direct risk to the preservation of the technological sovereignty of the German economy.

[20] Sanger, D. A., 2018, The perfect weapon: war, sabotage, and fear in the cyber age, Scribe U.K.; Kello, L., 2017, The virtual weapon and international order, Yale University Press; Corien Prins also points out that the new digital weaponry is changing the (geopolitical) order: "The balance of power is shifting, now that smaller countries can also enter the global battlefield. Without having to engage in a large-scale military confrontation or actually enter the territory of another state. In short, it is relatively easy to develop great clout," https://bit.ly/3JOI8Td.

As far as cyber threats are concerned, digital sovereignty cannot be separated from the three basic principles of information security, also known as the CIA of cyber security: confidentiality, integrity, and availability. In these three domains, autonomy must be safeguarded, not only at the level of a specific system in a specific sector (such as an ICT system in the criminal justice chain), but also in the larger framework of the economy, society, and democracy.

For example, through a specific government ICT system, sovereignty can be undermined – think of stealing information from government officials for espionage purposes[21] (confidentiality) and cyberattacks on so-called industrial automation and control systems (IACS) in our critical infrastructure (availability). These systems are the specific targets of hostile states in order to make sabotage possible in the future as a means of pressure to achieve geopolitical objectives.[22] In these cases, we can translate digital sovereignty into direct requirements for ICT systems. These include requirements for security, threat detection, continuity (backup, disaster recovery), vendor lock-in (preventing dependence on a specific supplier), and access to data by foreign powers (encryption requirements). As indicated above, digital sovereignty, however, must also be translated into the broader state interest of economy, society, and democracy. Some examples to illustrate are listed below.

2.2.1 EXAMPLE: CONTROL OVER ESSENTIAL ECONOMIC ECOSYSTEMS

- **Economic espionage:** the systemic theft by hostile states of intellectual property and know-how of our high tech companies and universities undermines Europe's future earning capacity.
- **Cloud infrastructure:** we are becoming increasingly dependent on the digital infrastructures owned by a number of major foreign market players, which offer limited portability and interoperability of data and applications. For innovation with AI, you need large quantities of harmonized data and a lot of computing power to process these data. Individual companies do not have sufficient data to innovate and, therefore, the data of companies in a specific industry sector will have to be combined. This is currently difficult as the data of companies is stored in silos in the clouds of foreign tech providers. As a result whereof, these have limited availability for European innovation. Access to harmonized data and cloud-infrastructure will become the foundation for the European innovation and knowledge infrastructure. Maintaining control over this is an essential part of digital sovereignty.
- **Digital communications networks:** we are increasingly dependent on digital communications for the wellbeing of citizens and a strong economy. Think of video meetings and smart homes, but also new security-critical services such as smart energy grids, intelligent mobility systems, and remotely controlled care robots. The development and management of the underlying technical systems and networks (such as routers, switches, and DNS servers) are increasingly dominated by foreign parties. As a result, organizations and individuals have only a limited understanding of their dependencies on these parties and their systems, let alone control over them. This restricts our ability to decide autonomously and to act on how we set up our digital infrastructure and to which parties we want to entrust the transportation of our data.

IACS are the systems (hard- and software) that control our locks and bridges and ensure that energy and gas are distributed, drinking water is cleaned, and nuclear material is processed. IACS allow organizations to control their industrial processes locally or at remote locations and to monitor and process real-time data.

Vendor lock-in is caused by the fact that a supplier uses its own proprietary standards, which means that software and applications only work on its own platform, making a switch from one customer to another supplier costly or even impossible.

Portability is the ability of applications and data to be transferred – with reasonable effort – from one IT environment to another (the process of transfer, we call migration).

Interoperability is the ability of IT systems to work together with other IT systems, allowing data to be exchanged, and to use the data that has been exchanged.

[21] See, for an example: Bloomberg Law, 2020, "Chinese hackers targeted European officials in phishing campaign," September 2, https://bit.ly/3h3GxfN.

[22] For enemy cyberattacks on IACS in critical infrastructures, see: Gartner, 2019, "A report for the Dutch Ministry of Justice and Security, Cyber Security Research for Industrial Automation and Control Systems," August 21, https://bit.ly/3JKplbr; and the advice of the Dutch Cyber Security Council: "Advice on the digital security of Industrial Automation & Control Systems (IACS) in the critical infrastructure of the Netherlands," April 24, 2020 (CSC Advice on Cyber Security IACS), https://bit.ly/3BHIDeE.

2.2.2 EXAMPLE: CONTROL OVER DEMOCRATIC PROCESSES AND RULE OF LAW

- **Manipulation of election processes:** when our governments are not in control of important democratic processes like elections, it mainly affects the internal legitimacy of the state (the trust of citizens in the state). Where a state is not in control of the election process, because it has been infiltrated and manipulated by foreign powers, its external legitimacy may also be compromised. For example, during the pandemic, both China and Russia blatantly pushed "fake news" to undermine our governments' COVID-19 responses. This undermined not only the internal legitimacy of our governments, but also their external legitimacy. Whereas before COVID-19 China and Russia at least tried to hide their involvement in cyberattacks, they are now doing so blatantly. It shows Europe's weakness; these states do not fear that retaliations will be forthcoming, undermining the E.U.'s external legitimacy. Not President Biden – after the SolarWinds and Colonial Pipeline incidents, Biden made cyberattacks firmly part of the political discussions between states and warned Russia and China that continued cyberattacks could lead to a "real shooting war."[23]
- **Infiltration of a vital government process:** can also undermine trust in the rule of law. Illustrative is an incident in Germany. In January 2020, Der Spiegel reported that the Berlin High Court (responsible for terrorism cases) had been systematically infiltrated by a Russian hacker group probably sponsored by the Russian government, identified as APT 28 (Advanced Persistent Threat). This hacker group had previously been held responsible for the infiltration of the German Bundestag. The attack focused on data exfiltration, accessing the entire database with identities of suspects, victims, witnesses, and undercover agents, and informants.[24] These types of infiltration both undermine a governments' internal and external legitimacy.

2.3 Increasing geopolitical tensions

Europe's sovereignty is affected by the increasing trade and ideological tensions between the U.S. and China. The new digital technologies have become the battleground for the race for global leadership between the two countries (aka the tech cold war).[25] The battle is mainly about leadership in the fields of 5G/6G, quantum computing, computer chip technology, and AI. Both the U.S. and China have chosen the route of tech protectionism, regularly drawing the national security card to justify addressing critical supply chain issues (exposed by the pandemic) by bringing manufacturing back to their countries,[26] imposing stricter export controls of critical technology, and stepping up controls of foreign direct investments (FDI).[27]

Other examples of geopolitically motivated measures are President Trump's ban on Huawei as a supplier of U.S. telecommunications infrastructure, and the restriction on Huawei to purchase computer chips produced with U.S. technology outside the U.S.[28] Rather than specific restrictions on Huawei, President Biden issued a presidential Executive Order (amending President's Trump earlier ban), ensuring that almost any ICT-related activity in the U.S. is subject to prior regulatory scrutiny for Chinese involvement by the U.S. government.[29] Not surprisingly, China is retaliating.[30]

These examples show that the E.U. and its member states are limited in their sovereignty by geopolitically motivated measures taken by the U.S. and China. The E.U. increasingly finds itself the piggy-in-the-middle in a bipolar world, which hampers the E.U.'s policy options. This plays a role throughout Europe in, for example, the choice of suppliers for 5G equipment, for which Huawei was initially an important potential candidate. As a result, 5G, a critical digital infrastructure, is likely to become more expensive as the multivendor choice decreases. Over time, restrictions will likely extend to other equipment, such as Huawei servers that support cloud services, the presence of Chinese suppliers in the Internet of Things (IoT), cameras, airport scanners, and other surveillance equipment, and drones of Chinese origin.

[23] Manson, K., 2021, "Biden warns cyber attacks could lead to a 'real shooting war,'" Financial Times, July 28, https://on.ft.com/35me5Du.

[24] Kiesel, R., A. Fr hlich, S. Christ, and F. Jansen, 2020, "Russische Hacker könnten Justizdaten gestohlen haben," Der Tagesspiegel, January 28, https://bit.ly/3v8l1xB.

[25] https://bit.ly/3v5G1Gr.

[26] FACT SHEET: Biden-Harris Administration bringing semiconductor manufacturing back to America," The White House, January 21, 2022, https://bit.ly/3h7Da7G; 27; Congressional Research Service, 2021, "U.S. export control reforms and China: issues for Congress," January 15, https://bit.ly/3s7pe3D.

[28] See for President Trump's Executive Order 13959.pdf (treasury.gov) (https://bit.ly/3BHrvpJ); this EO is basically replaced by President Biden's EO, see next footnote.

[29] FACT SHEET: Executive Order addressing the threat from securities investments that finance certain companies of the People's Republic of China, The White House, June 3, 2021, https://bit.ly/33GprBz.

[30] https://nyti.ms/3LKjvbU.

Giving in to U.S. pressure will potentially in turn lead to further Chinese pressure on European governments, including threats of Chinese import restrictions on European equipment and products. This ultimately affects our digital sovereignty and makes it more urgent for us to develop our own offerings as well.

2.4 Data as a weapon

Concerns of the U.S. and China go beyond ICT-supply chain dependencies and extend to what their adversary can do with information about their companies and citizens.[31] By now, both consider access to each other's data a matter of national security (they consider data as a weapon).

Increased tensions were kicked off by President Trump banning popular Chinese apps – such as TikTok and WeChat – from the U.S. app stores because these would undermine the "national security, foreign policy, and economy" of the U.S.[32] The measures were announced as the necessary protection of U.S. citizens from the unbridled collection of their data by the Chinese government. The U.S. was not alone, the Indian government also announced its intention to ban large number of Chinese consumer apps, including TikTok, because they are a "threat to sovereignty and integrity" and undermine "national security".[33] Trump's ban on these Chinese apps was met with severe skepticism about his true motives; the ban was considered part of the trade war with China, more than based on true concerns about privacy of U.S. citizens. However, subsequent reports about the massive mining by China of Western social media data to equip its government agencies, military, and police with information on foreign targets, should also give us pause.[34] President Biden dropped President Trump's Executive Orders banning Chinese apps, only to replace them by an Executive Order that provides powers to protect sensitive data of U.S. citizens from foreign adversaries.[35]

In response, in November 2021, China issued two pieces of sweeping privacy legislations, both basically banning all exports outside China of "important data," being any data that may endanger national security or public interests. Reviewing the categories of data caught by this definition shows that it is difficult to envisage what data could still be exported (e.g., covered are already personal data relating to more than 100,000 citizens). More telling is the fact that China is even willing to crack down on its own tech companies in order to prevent data of Chinese citizens ending up in the U.S. In June 2021, when Didi, the Chinese equivalent to Uber, got listed on the New York Stock Exchange, Chinese regulators retaliated by banning the Didi app from the Chinese app stores, alleging that Didi was illegally collecting users' persona data. Didi is now in the process of shifting its shares from New York to Hong Kong, caught between China announcing stricter control over foreign listings of Chinese companies and the U.S. Securities and Exchange Commission (SEC) finalizing rules empowering U.S. regulatory authorities to delist Chinese companies if their auditors refuse to share information requested by them.

Note that concerns about large scale harvesting of social media data extend beyond individual privacy of citizens, they also concern protection of our collective data. Analysis of data of a large enough portion of a population will be predictive for the entire population. The E.U. General Data protection Regulation (GDPR), will, therefore, provide no protection here. For example, if sufficient E.U. citizens provide consent for analysis of their DNA by a Chinese company, this will potentially impact us all.

Concerns about the Chinese harvesting of social media data (via apps like TikTok) become more understandable when one considers that hereditary data (from DNA) can now be combined with socioeconomic data (information about how we live, what we eat, when we exercise and sleep). With information about heredity and environment, suddenly precision medicine will be possible, potentially bypassing doctors. China itself is well aware of the risks, and clamped down on any access to their biological data and samples.[36] Note that where both the U.S. and China limit data transfers, data exchange by the E.U. is increasingly becoming a one-way-street. In response, we see data localization requirements creeping in at, for example, the E.U. standard setting level for cloud services[37] and data export restrictions on non-personal data under in the draft E.U. Data Act (stricter even than under the GDPR for personal data).[38]

[31] Reich, R., 2021, "Data, not arms, the key driver in emerging US-China cold war," The Guardian, July 10, https://bit.ly/3BEydwx.

[32] Executive Order on addressing the threat posed by TikTok – The White House (archives.gov), August 6, 2020 (https://bit.ly/3LRNzlZ); New York Times, 2020, "Trump's attacks on TikTok and WeChat could further fracture the internet," September 18, https://nyti.ms/3sUMtxj.

[33] https://bit.ly/3H9xch8 34 https://bit.ly/3h62OcX; https://bloom.bg/3h6k7dP. 35 https://bit.ly/3sYaYJR.

[34] https://bit.ly/3h62OcX; https://bloom.bg/3h6k7dP. 35 https://bit.ly/3sYaYJR.

[35] https://bit.ly/3sYaYJR.

[36] https://bit.ly/3BD4AvD.

[37] See Position Paper of the Dutch Online Trust Coalition on regulatory developments at ENISA originating from the Cyber Security Act, https://bit.ly/33IyB0y.

[38] Which is scheduled to be officially published on 23 February 2022; see for the leaked version: https://bit.ly/3h9LHXD.

EXAMPLE: CONCERN ABOUT CHINA HARVESTING BIOLOGICAL DATA

In January 2021, it was widely reported in the U.S. media that at the outbreak of the pandemic, the world's largest biotech firm (based in China and with strong ties to the Chinese government) made an offer to the governors of six U.S. states to help build and run state-of-the-art COVID-19 testing labs against very favorable conditions.[39] So favorable indeed, that it seemed like an offer the states could not refuse. When the governors compared notes, however, they concluded that some offers are indeed too good to be true, the ulterior motive of the offer likely being to obtain biometric information of large parts of the American population to be used for Chinese DNA science, to develop vaccines and precision medicine. The offer lead U.S. officials to issue public warnings to hospitals and governmental agencies that "Foreign powers can collect, store and exploit biometric information from COVID tests."[40] The Chinese quest to control biodata and therewith control healthcare's future, is also called the new space race.

2.5 Referees do not win the match

The E.U. is behind in innovation, especially in AI innovation.[41] This is due to a lack of investment, but also because for a long time Europe thought that its laws and regulations would protect it. Until as recently as 2017, talking about European sovereignty was very much not done and Europe was in favor of the open liberal market economy and European research programs, for example, had to be "open to the world".[42] Europe trusted its regulatory power to protect E.U. values and the fundamental rights of its citizens. An example is the GDPR, the world's first sweeping omnibus law protecting the personal data of individuals. In a similar vein, the E.C. intends to be the first to issue omnibus AI regulation.[43] For a long time this has been a successful recipe, the E.U. is by now considered a regulatory powerhouse, where E.U. regulations have a strong effect also outside the E.U. (coined the Brussels Effect).[44] Case in point is again the GDPR. By now about 120 countries have followed suit and adopted omnibus data protection laws, of which 17 have explicit GDPR-like legislation.[45] There is even a call by leading tech companies to make GDPR the "law of the world".[46] Though successful from a regulatory perspective, the realization has set in that GDPR may succeed in protecting data of individual citizens, but not in protecting the E.U.'s economic ecosystem. GDPR actually hampers innovation. To start with, the rules are so strict and costly to implement that they are difficult for startups and smaller companies to implement. GDPR has in practice proven to be a strong competitive advantage of large technology companies.[47] In a similar vein, the prediction is that the draft AI Regulation will be so elaborate and costly to comply with that it will likely hamper E.U. innovation.[48] Second cause is that while E.U. research is open to the world, large data-intensive companies "hide" behind GDPR so as not to open up their data for research in the public interest. And finally, and most importantly, the realization has set in that rules do not protect if you do not innovate yourself: referees do not win the match.[49] Two examples to illustrate:

2.5.1 EXAMPLE: APPLICATION OF AI

GDPR requires that deploying an algorithm should not lead to discriminatory outcomes.[50] GDPR also requires companies applying algorithms for automatic decision-making – for example, automated rejection of a loan application – to provide individuals with meaningful information about the underlying logic and an explanation of the decision, so that they can challenge the decision.[51] At present, however, advanced forms of AI are still a black box – we do not know how algorithms arrive at their outputs. Innovation is, therefore, required to prevent discriminatory outcomes and ensure transparency and explanation.[52] In fact, innovation at major U.S. tech

[39] https://cbsn.ws/34ZQGrx.

[40] Ibid.

[41] https://bit.ly/3s9NZfL; https://bit.ly/3s7VGD4.

[42] "Horizon 2020 is open to the world," https://bit.ly/3BHIND1.

[43] European Commission Proposal for a Regulation laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act), (April 21, 2021), COM(2021)206 final.

[44] Bradford, A., 2020, The Brussels effect: how the European Union rules the world, Oxford University Press.

[45] https://bit.ly/3H4Bpm5.

[46] https://bit.ly/3JM9wBe.

[47] Yueh, J., 2018, "GDPR will make big tech even bigger," Forbes, June 26, https://bit.ly/33EyXVN.

[48] MacAfee, A., 2021, "EU proposals to regulate AI are only going to hinder innovation." Financial Times, July 25, https://on.ft.com/3sX7tn4.

[49] https://politi.co/34Zi0Gm; https://bit.ly/3s895ek.

[50] We regularly see in the news that the application of self-learning algorithms leads to discriminatory outcomes, see for example: Dastin, J., 2018, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters Business News, October 10, https://reut.rs/3sX7dV8.

[51] Articles 13, 14, and 22 (3) and Recital 71 of the GDPR. For information on these requirements, see Moerel, L., and M. Storm, 2019, "Automated decisions based on profiling: information, explanation or justification, that is the question!" in Aggarwal, N., H. Eidenmüller, L. Enriques, J. Payne, and K. van Zwieten (eds), Autonomous systems and the law, C. H. Beck.

[52] It is not easy to do this properly. See Moerel, L., 2018, "Algorithms can reduce discrimination, but only with proper data," Op-ed, IAPP Privacy Perspectives, November 16.

companies is currently geared toward cracking this black box and developing new de-biasing techniques.[53] Various media reported that Google has tackled the black box problem with "explainable AI",[54] which is expected to be a major competitive advantage going forward.

2.5.2 EXAMPLE: DATA TRANSFER RULES

In terms of control over European data, worrying from a sovereignty perspective is that U.S. intelligence agencies have certain powers for espionage and counterterrorism purposes to intercept foreign data in transit to the U.S. on transatlantic cables, and also have powers to collect data from U.S. cloud providers if they are hosted on servers in the U.S.[55] Two specific interception powers[56] have recently led the European Court of Justice (ECJ) in the well-known Schrems II judgment[57] to rule that U.S. law does not provide an equivalent level of protection to personal data of European citizens after being transferred to the U.S. U.S. law does not meet the requirements of the GDPR and the European Charter of Fundamental Rights of the E.U. The judgment has far-reaching consequences because in countries such as China, Russia, and India, the authorities have similar interception powers as the U.S. authorities. Also for these countries, data transfers are, therefore, under discussion. The ECJ leaves open the possibility for organizations to take supplementary mitigating measures that in specific cases address the shortcomings as a result whereof transfers can still take place.[58] Since U.S. intelligence agencies are not bound by contractual measures between the data exporter and importer, an obvious solution is to seek additional protection in data encryption. The data can then still be intercepted, but the foreign states can do little with these. Fact is that currently encryption is only possible for data at rest and for data in transit. Here too we see technical innovations in which data in use can also be encrypted (so-called homomorphic encryption).[59] U.S. cloud providers are the first to come up with practical applications here.[60] This form of encryption ensures that U.S. intelligence services do not have access to identifiable data, even when obtained when the data were in use. At the same time, it ensures that the providers themselves can analyze the data in order to generate insights. This innovation will, therefore, further strengthen the dominant position of these providers (see next section).

DATA CAN BE IN THREE STAGES:

Data at rest: the data are inactive and stored, for example in a database.

Data in transit: the data are transported over a network.

Data in use: the data are processed in an application.

Homomorphic encryption is a form of encryption that allows operations to be performed on the data without first having to decrypt it.

Exit and transition: customer dependencies often arise when contracts terminate because the customer needs the cooperation of the supplier for the transition of data and applications to a successor supplier (who in turn applies its own standards). For this purpose, specific protocols for "exit and transition" are already agreed upon at the conclusion of the contract.

2.6 Dependencies on dominant foreign suppliers

It will require little explanation that where governments and providers of critical infrastructure increasingly outsource their ICT systems, data storage, and processing to suppliers, new dependencies arise, especially if those suppliers are dominant market players.[61] The concept of digital sovereignty then also extends to the autonomy of our government and providers of critical infrastructure vis-à-vis these commercial parties, and where these are foreign parties, to their respective governments.

[53] The U.S. government is also making an effort. See, for an example of innovation in the field of explainable AI (also known as XAI), a project by the Defense Advanced Research Projects Agency (DARPA), Gunning, D., "Explainable Artificial Intelligence (XAI)," https://b.gatech.edu/3BHGVtZ.

[54] Kelion, L., 2019, "Google tackles the black box problem with Explainable AI," BBC, November 24, https://bbc.in/3p7DUhl.

[55] For a (still up-to-date) overview of the possibilities of interception by U.S. intelligence services of data of non-Americans, see Gorski, A., 2018, "Summary of U.S. Foreign Intelligence Surveillance Law, practice, remedies and oversight," American Civil Liberties Union Foundation, August 30, https://bit.ly/3JBHt7s.

[56] This concerns the powers of U.S. intelligence agencies under Section 702 of the Foreign Intelligence Surveillance Act ("FISA") and Executive Order ("EO") 12333.

[57] Case C-311/18, Data Prot. Comm'r v. Facebook Ireland Ltd, ECLI:EU:C:2020:559 (July 16, 2020), https://bit.ly/3BEB4FR.

[58] Ibid, See paragraph 133.

[59] See on this topic: Divatia, A., 2019, "Fact and Fiction of Homomorphic Encryption," Dark Reading, January 22, https://bit.ly/3p3aWzq.

[60] See for offer Microsoft: https://bit.ly/3v6oOwm; IBM: https://ibm.co/3BHVL3q, and Google: https://bit.ly/35jrcVW.

[61] The Dutch Scientific Council for Government Policy, in its advice "Preparing for digital disruption," 2019, Chapter 3, gives a good overview of the far-reaching digitalization of society, the strong interweaving of the digital domain and the physical domain, and the new vulnerabilities that this creates for core societal processes, WRR Advice Digital Disruption, https://bit.ly/34ZCbnx.

The international cloud providers compete on security and are best in class. The deployment of cloud solutions now offers so many advantages in terms of functionality (e.g., built-in data analysis tools), higher implementation speed, innovation, the possibility of collaboration, and often lower costs, that the use of cloud services is now also seen as "necessary for a well-functioning government", making government policy cloud first, both in the Europe as in member states.

In the market, there is a very limited choice of so-called hyperscalers (cloud providers with large capacity). As a consequence, currently 92 percent of the data of European companies and citizens reside in the clouds of U.S. technology companies, of which 80 percent are with five providers only.[62] European suppliers hardly appear in the picture.[63] These five players are now so big that if there is an outage of one of them, it is like a power cut, entire E.U. sectors will be down. If 10 years ago we would have asked ourselves whether this – in principle – would be a good idea, none of us would have answered in the affirmative. We would never put the switch of our power grid in the hands of a foreign company, and its government.

The dominance in market positions further leads to an imbalance between supplier and customer, with monopolistic behavior in contracts, price, service, and dependencies for the future (not only because of dependencies on contract termination (exit and transition), but also because making changes to standard solutions is difficult).[64]

The major market players offer limited interoperability and portability of data and applications. Because of their scale, they are able to use their own standards – often protected by intellectual property rights – and even to build a private internet infrastructure (including even their own submarine cables),[65] which makes them virtually autonomous both physically and legally and makes any interconnection difficult, both in terms of infrastructure and data exchange.[66] To prevent vendor lock-in, clients (including governments)[67] usually have a so-called multi-vendor strategy. However, under current market conditions, this is difficult to achieve.

The current expectation is that – without government intervention – the dominant positions of these market players will only increase. These market players are systematically expanding their ecosystem by integrating new functionalities into their services (such as cybersecurity and data analysis tooling), which will only increase vendor lock-in.[68] They are also able to attract the best talent worldwide and have almost inexhaustible access to capital. This enables them to continuously monitor innovations and startups, which they then take over at an early stage and integrate into their own offerings.[69]

These dominant positions (winner takes all) are a sign of the times and should not be taken as a given. As said, our society is undergoing a technological revolution, which brings along disruption and friction. History shows that whenever new technologies disrupt society, it needs time to adjust and regulators always play catch-up. At this time, the digital society is still driven by the possibilities of technology rather than social and legal norms.[70] These frictions will ultimately be addressed. For example, the first industrial revolution brought child labor, abuse of workers, and the skies of London were so full of soot that people fell ill. The barons of the new industry (steel, oil, copper, and coal) reigned supreme, with worsening inequalities due to their monopolist positions. Ultimately many new laws were introduced, most notably the first antitrust regulation, which broke up the monopolies. Illustrative here is that President Biden, when introducing his Executive Order on Promoting Competition in the American Economy,[71] made several references to the importance of abiding to the original principles of antitrust regulation also in the new digital economy: "It is the policy of my Administration to enforce the antitrust laws to meet the challenges posed by new industries and technologies, including the rise of the dominant Internet platforms, especially as they stem from serial mergers, the

[62] Amiot et al. (2020).

[63] Synergy Research Group, October 29, 2019.

[64] European Commission, 2020, "Communication: a European data strategy," February 19, https://bit.ly/3BJyYV1.

[65] Where even own submarine cables are laid, see for Google: https://bit.ly/34ZBmLt; and for Microsoft and Facebook: https://bit.ly/3v8RG7s.

[66] See farewell speech Jan Smits, https://bit.ly/3v8oBZy.

[67] See e.g., Cloud principles JenV, p.2, and European Commission/DIGIT (Appendix 3 – EU Cloud Policy).

[68] This problem is also called out by the European Commission, See European Data Strategy, p. 7. The financial sector (banks, supervisory authorities, etc.) also analyzes the strategic aspects of its own cloud policy. The European Securities and Markets Authority (ESMA) opened the consultation of its directive on cloud outsourcing on June 3. Steven Maijoor, the chairman of ESMA, explained, "Financial markets participants should be careful that they do not become overly reliant on their cloud services providers. They need to closely monitor the performance and the security measures of their cloud service provider and make sure that they are able to exit the cloud outsourcing arrangement as and when necessary." https://bit.ly/3JNPZjY.

[69] See about these practices: https://bit.ly/36lNAhl.

[70] Moerel, L., 2014, "Big data protection: how to make the draft EU regulation on data protection future proof," working paper, Tilburg University, https://bit.ly/3JQs5Et.

[71] https://bit.ly/3s72nFC.

acquisition of nascent competitors, the aggregation of data, unfair competition in attention markets, the surveillance of users, and the presence of network effects."

My point here is that governments around the world (including the U.S., China, and the E.U.) are currently considering their policy responses and antitrust investigations are underway on all continents.[72] Once these have done their work, the world will look very different indeed.

3. E.U. POLICY RESPONSE

An important upfront observation is that the E.U.'s mandate to safeguard the necessary form of sovereignty is limited. Although the E.U. can take initiatives in a large number of areas to strengthen "digital sovereignty", there is an important obstacle. In essence, the problem is that digital sovereignty soon touches on the national security of member states, which under the E.U. treaties is the prerogative of the member states. Where, however, the member states individually can no longer protect their sovereignty, the limited European mandate actually undermines national security.[73] E.U. digital sovereignty policy is, therefore, often framed in terms of the power of the E.U. to regulate the "internal market", while the real underlying denominator is protection of sovereignty. Where previously this would raise concerns among member states, we see an increased willingness to cooperate at the European level in the digital domain and to pool or share sovereignty.[74]

The second observation is that due to the multifaceted nature of the causes of the pressure on our digital sovereignty, there is no one-size-fits-all solution. Europe's sovereignty will have to be supported by a "smart" combination of measures acknowledging that becoming self-sufficient is not realistic for Europe, but also not desirable.[75] With the E.U. policy measures, the E.C. is aiming to pave a third way, aiming to avoid falling into the trap of tech protectionism. The policy is, for example, not to exclude foreign digital providers, nor for Europe to build its own hyperscalers. And rightly so, if you have concerns about vendor/data lock-in with current big tech companies, you will have similar concerns with their E.U. equivalent. Rather than blocking foreign suppliers, E.U. policy is about breaking through vendor/data lock-in by ensuring:

- **Interoperability of cloud infrastructure** in order to achieve the required scalability for innovations, without setting up its own hyperscalers.
- **Open data**, which makes it possible for an industry sector to combine its data in a common data space, to unlock their value for AI innovations.
- **Open source technologies**, which can be worked on collectively, and forked individually; the only way Europe will be able to match the R&D budgets of the tech giants, gaining both the benefits of scale and self-sovereignty.[76]
- **Federated solutions**, whereby data are not continuously copied, but remain at the source and are drawn on, where necessary, preserving privacy and self-sovereignty.

3.1 Increased cyber resilience and regulation of gatekeepers

Important building blocks of the E.U. sovereignty policy measures (but not further discussed here) are omnibus measures to increase the cyber resilience of critical infrastructures and services in Europe in the upcoming directive on the resilience of critical entities and the renewed Security of Network and Information Systems (NIS2) Directive.[77] Other components are proposals to better regulate the market power of gatekeepers providing core platform services (such as search engines, social networks, video sharing, and cloud computing services) in the Digital Markets Act[78] and increased requirements and liability of large online platforms related to the spreading of illegal content, misinformation, and targeted advertising practices in the Digital Services Act.[79]

[72] See for overview: https://bit.ly/3h62B9D.

[73] See on this paradox and potential solutions, Timmers, P., and L. Moerel, 2020, "Reflections on digital sovereignty," E.U. Cyber Direct, January 15, https://bit.ly/3s7sz2K.

[74] A telling example is 5G security, where the Member States asked the EC to draw up a joint direction for 5G security, even though the concerns in this area primarily concern national security. This was unthinkable not so long ago.

[75] See Timmers and Moerel (2020) for three approaches to achieve digital sovereignty: risk management, strategic partnerships, or working together on a global level to find solutions in the common interest (global common goods).

[76] Thompson, B., 2021, "Internet 3.0 and the beginning of (tech) history," Stratechery, January 12, https://bit.ly/3sag1Ib.

[77] European Commission, 2020, "Proposal for a Directive on measures for a high common level of cybersecurity across the Union, repealing Directive," (E.U.) 2016/1148, December 16.

[78] Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM/2020/842 final.

[79] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final.

3.2 Open data – open infrastructure – open source

The focus here is on the other policy initiatives – all dating from 2020 – and aimed at ensuring interoperability of E.U. data and cloud infrastructure, avoiding vendor/data lock-in.

3.2.1 OPEN DATA

The cornerstone is the E.U. Strategy for Data,[80] and specific for the financial sector: the E.U. Retail Payment Strategy[81] and the Digital Finance Strategy.[82] The European Strategy for Data aims to democratize access to data assets and drive data sharing in open digital ecosystems across the whole economy. It also aims to create a single market for data to be exchanged across sectors efficiently and securely within the E.U. in a way that fits European values of self-determination, privacy, transparency, security, and fair competition. The centerpiece of the European Data Strategy is the concept of European data spaces, bringing together E.U. data of nine defined clusters of organizations with common interests (including financial, health, and government), so that the scale of data required for innovation for a cluster can be achieved. The design of the data spaces will be based on full interoperability and data sovereignty, whereby users will be provided tools to decide about data sharing and access.[83] With the actual parties that generate the data regaining control, large hyperscalers will no longer be able to achieve vendor/data lock-in in their proprietary ecosystems. In this context also fits the Data Governance Act,[84] opening up public data for innovation through independent intermediaries and the draft E.U. Data Act, providing a harmonized framework for all data sharing, conditions for access by public bodies, data export restrictions for non-personal data, and portability and interoperability requirements for cloud services.[85] Where data spaces require many-to-many interactions, digital identity solutions and consent dashboards will become an inherent part of the design (E.U. digital identity solutions are further discussed in section 3.3, below).

3.2.2 OPEN INFRASTRUCTURE

Another flagship initiative is the GAIA-X project,[86] which is aimed at achieving interoperability between cloud offerings to achieve the required scalability of the cloud infrastructure for AI-related innovation, not by creating Europe's own vertical hyperscalers but by networking (making interoperable) the current European offer of cloud infrastructure, enabling clients to scale up within that network (i.e., scaling up in a horizontal way). This is achieved by setting common technical standards and legal frameworks for the digital infrastructure and standardizing contract conditions. This form of interoperability goes beyond portability of data and applications from one vendor to another to prevent vendor lock-in; it really concerns the creation of open APIs, interoperability of key management for encryption, unambiguous identity, and access management, etc. Cloud providers will be expected to offer a choice as to where (personal) data are stored and processed, without otherwise requiring storage in Europe. The GAIA-X project is not a comprehensive European policy, but it is a concrete realization of the open interfaces, standards, and interconnection needed for the European policy and explicitly based on principles of sovereignty-by-design. The project is open to foreign suppliers as long as they embrace the principles. From a digital sovereignty perspective, the GAIA-X project is a logical and promising initiative and is gaining more and more traction.[87] The expectation is that once the design principles are agreed upon, these may well become mandatory for all cloud services in Europe. Some of the elements (portability and interoperability requirements and data export restrictions for non-personal data) are already included in the draft E.U. Data Act.

Though the initial aim of GAIA-X is to achieve an open cloud infrastructure in an open market, we have recently seen that digital sovereignty concerns lead to an increased pressure to move to stand-alone E.U. cloud only solutions, whereby all E.U. data are stored in the E.U. only (unless the service requires transfer of data, e.g., in case of communication services). Rather than addressing sovereignty concerns in respect of

[80] European Commission, 2020, "A European data strategy," COM(2020)66, February 19.

[81] European Parliament, 2020, "Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the committee of the regions on a Retail Payments Strategy for the EU," https://bit.ly/3v3ZhnH.

[82] European Parliament, 2020, "Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the committee of the regions on a Digital Finance Strategy for the EU," https://bit.ly/3BOdxSY.

[83] See for overview of the data space design principles: "Design principles for data spaces," position paper, https://bit.ly/3p79v2O. 84 Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act), COM/2020/767 final.

[84] Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act), COM/2020/767 final.

[85] Which is scheduled to be officially published on 23 February 2022; see for the leaked version: https://bit.ly/3t4ExcC.

[86] "A Federated data infrastructure as the cradle of a vibrant European ecosystem," the GAIA-X project initiated by the German and French governments, October 2019, based on principles of sovereignty-by-design.

[87] In the Netherlands, a coalition of TNO and a number of industry associations are actively contributing to the GAIA-X project, https://bit.ly/3p7hbSx.

foreign cloud providers and data transfer issues at an E.U. policy level, we see data localization requirements creeping in at, for example, the E.U. standard setting level for cloud services[88] and data export restrictions on non-personal data under in the draft E.U. Data Act (stricter even than under the GDPR for personal data). Telling in this context is that Commissioner for the Internal Market Thierry Breton recently stated: "European data should be stored and processed in Europe because they belong in Europe."[89] It is not clear yet what the end result will be.

3.2.3 OPEN SOURCE TECHNOLOGY

The E.C. has an active open source software strategy, where open source solutions are preferred when equivalent in functionalities, total cost, and cybersecurity,[90] which facilitates decentralized and federated services that can be independently audited, contributing to public trust. Open source technologies can further be worked on collectively, which provides benefits of scale (combining the E.U. R&D to potentially match the R&D budgets of the big tech companies), but also ensures self-sovereignty as open source can always be subsequently forked individually for specific solutions.[91]

3.3 E.U. digital ID wallets

Part of the policy package is a proposal to create a modernized framework for a European digital identity,[92] based on self-sovereignty of European citizens. Member states will offer citizens and businesses "European digital ID wallets"[93] (digital ID wallets), which are stored as an app on smartphones and enable E.U. citizens to authenticate and access online services across the E.U. The digital ID wallets will be issued by a member state or by private entities (after their wallet is certified by accredited bodies designated by the member states). The digital ID wallets will enable citizens to do more than simply prove their identity: the wallets will also store proof of other personal attributes and credentials, such as driving license, education certificates, birth certificate, bank cards, a specific attribute to demonstrate you are older than 18 (to access certain websites), and further enable citizens to digitally sign documents with a qualified electronic signature (this is a higher level of identity proofing and security and is suited for banking transactions). This will be a big change. For example, when renting a car, an individual can prove possession of a driving license by sharing the attribute "in possession of a driving license" from the digital ID wallet, without having to actually provide a copy thereof. At the moment, citizens still have to login for each and every digital service with the vulnerable system of user name combined with password and manually enter and disclose (always the same) personal data. To simplify login, many websites offer citizens the option to authenticate via their account with one of the major foreign platforms, such as Facebook, Google, and Alibaba. This creates large concentrations of both business and personal data on these platforms, which has a direct impact on citizens' privacy and digital sovereignty.

Under the new regulation, large platforms will be required to accept the use of the digital ID wallets as well as all services that require strong customer authentication (SCA). The new regulation further restricts sharing of personal data to what is strictly necessary for the provision of the service, precludes the issuer of the wallet from collecting information on the use of the wallet, and prevents the issuer from combining personal data in the wallet with any other personal data in its possession, "unless the citizen expressly requested it".

Where data sharing across industries (in a so-called multi-to-multi-markets) becomes the norm, digital ID wallets will become a new intermediary function in the ecosystem, potentially disrupting current platforms. Not surprisingly, Apple has already included self-sovereign wallet functionality in its latest iOS 15, which may well meet the E.U. requirements.[94] The Apple ID wallet will be disruptive for the other large platforms (as these once were to others) and is expected to become its next big revenue source, more so than Apple Pay.[95]

Though the above restrictions on data collection and combining by issuers of the wallet may – at face value – seem detrimental to digital business models of issuers, the opposite

[88] See Position Paper of the Dutch Online Trust Coalition on regulatory developments at ENISA originating from the Cyber Security Act, https://bit.ly/3saeSQT.

[89] According to a POLITICO interview on September 1, 2020, https://politi.co/3JJJQoS.

[90] Communication to the Commission Open Source Software Strategy 2020 – 2023 Think Open, C(2020)7149 final, https://bit.ly/3BNhozx.

[91] https://bit.ly/3H8tZ1q 92 Proposal for a Regulation of the European Parliament and of the Council amending Regulation (E.U.) no. 910/2014 as regards establishing a framework for a European Digital Identity, COM/2021/281 final.

[92] Proposal for a Regulation of the European Parliament and of the Council amending Regulation (E.U.) no. 910/2014 as regards establishing a framework for a European Digital Identity, COM/2021/281 final.

[93] Defined in Article 3(42) as "a product and service that allows the issuers to store identity data, credentials and attributes linked to her/his identity, to provide them to relying parties on request and to use them for authentication, online and offline, for a service in accordance with Article 6a; and to create qualified electronic signatures and seals."

[94] Velasco, J., 2021, "Apple wallet with iOS 15 is close to replacing your wallet," Digital Trends, June 7, https://bit.ly/3sX2see; Apple, 2021, "Apple announces first states to adopt driver's licenses and state IDs in Wallet," press release, September 1, https://apple.co/3JlxnBl.

[95] https://bit.ly/3JlxehE.

is the case. Where many market players have to accept the digital ID wallet for authentication, having the channel to actually be able to request consent from users for data sharing becomes a competitive advantage in and of itself.

3.4 Impact on financial sector

Looking at these policy initiatives, these will have a fundamental impact also on the business models of the financial sector. The introduction of "open banking" in the revised Payment Services Directive (PSD2) and the E-Money Directive already lowered the barriers for non-banks (fintechs, big tech, etc.) to leverage the payment data of banks in order to provide value propositions on top of the payment infrastructure.[96] Financial institutions also complain that there is an increased use of the authentication solutions of big tech companies to access their payment processes, increasing their dependency on these providers and making it difficult to maintain the security of access to their services. In fact, banks complain about the gatekeeper function of big tech. However, due to the E.U. policy measures, what really is at stake is the banks' own gatekeeper function: "banks are no longer the sole manufacturers and distributors of payments and other financial products (e.g., loans) and hence risk losing their long-held dominance of the sector."[97] As often, once the insight is there, regulatory changes are also an opportunity. Instead of resisting the open banking and open data requirements, banks are well advised to embrace these and become open banks, facilitating (also) data driven transactions and many-to-many reach, for example, by allowing consumers to share energy data with loan providers.[98] As already well described by other authors, in this new data ecosystem banks could well leverage their customers' trust (and preserve customer contact and relevance) by becoming digital identity providers and data custodians.[99] As indicated above, digital ID wallets will quickly become a new intermediary function in the ecosystem, disrupting the gatekeeper function of the current platforms. The restrictions on issuers of wallets as to data collection and combining may seem detrimental, but actually create a channel to request consent from users in the first place (preserving customer contact and relevance).

The adoption of digital ID wallets will further accelerate digitalization in and of itself, e.g., will enable banks to rely on these digital identities to perform know your customer/anti-money laundering (KYC/AML) due diligence, facilitate executing banking documents, and use these identities to meet strong customer authentication (SCA) requirements under the revised Payment Services Directive (PSD2). Taking it one step further, banks could also become an active attribute provider for wallets, such as KYC/AML attributes, which can also be used by other service providers (against a payment). This will enable the banks to actually monetize their current KYC/AML efforts. Rather than frowned upon, this is actively encouraged by the E.C.[100] Other relevant attributes to be issued by banks could be source of funds, source of wealth, insolvency/bankruptcy risk, transactional behavior, banking relationship, etc. Where the European Central Bank is working towards a digital euro,[101] the digital ID wallet should in the future also facilitate payments with these digital currencies (digital currency wallet), including complex transactions like cross-border or multi-currency transactions. In this last scenario, all features of the E.U. digital policy will be combined: open banking, digital currency, digital ID wallets, and SCA under PSD2.[102]

4. CONCLUSION

History shows that whenever new technologies disrupt society, it needs time to adjust and regulators always play catch up. At this time, the digital society is still driven by the possibilities of technology rather than social and legal norms. This inevitably leads to social unrest and calls for new rules. An illustrative example here is that in 2010, Mark Zuckerberg (CEO and founder of Facebook (Meta)) caused quite a stir when he publicly announced that the end of privacy was in sight: "People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people. That social norm is just something that has evolved over time."[103]

[96] Zachariadis, M., and P. Ozcan, 2017, "The API economy and digital transformation in financial services: the case of open banking," SWIFT Institute working paper no. 2016-001, https://bit.ly/3If1vUY.

[97] Cortet, B., M. Bakker, P. Groen, and D. Hoppenbrouwer, 2021, "Establishing the trust anchor in the digital economy: The case for banks to become 'data custodians,'" Journal of Payments Strategy & Systems 15:2, 150-164.

[98] Ibid.

[99] World Economic Forum, 2016, "A blueprint for digital identity, the role of financial institutions in building digital identity," https://bit.ly/3BOaGcl; Wilson, M., 2021, "Commercialising open banking – digital identity, a key opportunity for banks?" https://bit.ly/3s8OeaP.

[100] https://bit.ly/3p5WAOO.

[101] Wagner, E., D. Bruggink, and A. Benevelli, 2021, "Preparing euro payments for the future: a blueprint for a digital euro," Journal of Payment Systems & Strategies 15:2; European Central Bank, 2021, "ECB publishes the results of the public consultation on a digital euro," press release, https://bit.ly/34X9pUA.

[102] Adams, M., L. Boldrin, R. Ohlhausen, and E. Wagner, 2021, "An integrated approach for electronic identification and central bank digital currencies," Journal of Payment Systems & Strategies 15:3.

[103] Johnson, B., 2010, "Privacy no longer a social norm, says Facebook founder," The Guardian, January 11, https://bit.ly/3p60fw0.

However, in March 2019 (following the Cambridge Analytica data analysis scandal), Zuckerberg requested that the U.S. senate regulate tech companies[104] and further announced a complete overhaul of Facebook's privacy features: "The future is private... and that's the next chapter for Facebook."[105] From privacy is dead to privacy is the future. My point here is that not only are technical developments moving fast, but also that social standards and customer expectations are evolving and that it will take years before we will have a somewhat clear and predictable new regulatory framework.

The threats to E.U. digital sovereignty have led to a flurry of E.U. digital policy measures, that will disrupt the digital landscape as we know it by working towards open infrastructure, open data, and application of open source technology. E.U. digital policy will have a fundamental impact on the business models of the financial sector. When E.U. policy has done its work, the world will look very different, though how it will look is anyone's guess. The financial sector has to be well tuned in to these developments to determine a digital strategy that can benefit from the new reality. Digital is not a communication channel or a specific expertise, it is, by now, the business itself. It is not possible to manage a company without knowledge of the business. For those tuned in, the E.U. digital policy may bring new requirements, but first and foremost many opportunities for innovation.

[104] Miller, J., 2019, "Mark Zuckerberg asks governments to regulate tech firms," Techspot, March 31, https://bit.ly/3vctrFk.

[105] See videos at: Hassan, A., 2019, "Zuckerberg promises 'complete overhaul' of Facebook geared towards user privacy at F8," ABC, April 30, https://abc13.co/3lhJBkC.

CONSTRUCTION OF MASSIVE CYBERATTACK SCENARIOS: IMPACT OF THE NETWORK STRUCTURE AND PROTECTION MEASURES

CAROLINE HILLAIRET | Professor and Director of the Actuarial Science engineering track and Advanced Master, ENSAE and CREST

OLIVIER LOPEZ | Professor of Applied Mathematics (Statistics), Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université[1]

ABSTRACT

This paper proposes a stochastic model to simulate massive cyberattack scenarios, taking into account the structure of the network as well as partial or full protection measures. Events, such as the recent COVID-19 pandemic, can rapidly generate consequent damages, and mutualization of the losses may not hold anymore. The framework is based on the multigroup SIR (susceptible, infected, and recovered) epidemiological model, which can be calibrated from a relatively small amount of data and through fast numerical procedures. As an illustration, we replicate the impact of a Wannacry-type event using a connectivity network inferred from macroeconomic data of the OECD. We show how this model can be used to generate reasonable scenarios of cyber events, and investigate the response to different types of attacks or behavior of the actors, allowing for the quantification of the benefits of an efficient prevention policy.

1. INTRODUCTION

With the growth of the digital economy, cyber risks are now one of the most important, if not the most important, threats facing the global financial system. The annual losses caused by cybercrime are estimated to be close to 1 percent of the world's GDP, U.S.$1 trillion. This threat has been amplified since the COVID-19 pandemic, as suggested by Kshetri (2020) and the French National Agency for Information Systems Security [ANSSI (2021)], which found a threefold increase in the number of reported ransomwares attacks between 2019 and 2020.

To face of cyber risk, insurance has a crucial role to play [Xie et al. (2020)]; and it is not only a matter of financial compensation, as cyber contracts generally include offers of prevention and assistance in the event of a loss [Romanosky et al. (2019)]. Nevertheless, quantifying the impact of this multi-faceted risk is a difficult task and a major concern for insurers. The extreme severity of some cyber events [Farkas (2021)] on the one hand and the potentially "systemic" nature of the risk on the other hand [Hillairet and Lopez (2021)] could endanger the principle of mutualization, which is at the heart of the insurance business. In particular, massive cyberattacks and contagion effects can lead to massive failures that can bring an economy to a halt, or at the very least jeopardize the solvency of an insurer. For example, the report by Cyence and Lloyd's of London [Cyence (2017)] estimates that the cost of an attack on a major cloud provider would be in the range of U.S.$15 billion to U.S.$121 billion, with an estimated average loss of U.S.$53 billion. The Wannacry or NotPetya episodes are also warning signs of massive cyberattacks, whose estimated costs are in the billions of dollars. It is important to note that even if the damages of each individual incident are low, the simultaneous occurrence of a large number of incidents in a massive attack can result in very high cumulative costs.

[1] The authors acknowledge funding from the project "Cyber Risk Insurance: actuarial modeling", Joint Research Initiative under the aegis of the Risk Foundation, in partnership with of AXA, AXA GRM, ENSAE, and Sorbonne Université.

In this paper, we propose a general and flexible framework to model the dynamics of a cyber contagion and to simulate accumulation scenarios, with a focus on their impact on an insurance portfolio. In order to take into account networks effects in the contagion [Fahrenwaldt et al. (2018)], we adopt a multi-group SIR model (susceptible, infected, and recovered) [Beretta and Capasso (1988), Guo et al. (2006), Magal et al. (2018)]. These types of compartmental models are commonly used to describe biological epidemics since McKendrick (1925), and have already been applied to several actuarial applications [Chen and Cox (2009), Lefèvre et al. (2017), Garrido and Feng (2011)]. Special attention is paid to the quantification of the impact of prevention and quick reaction to diminish the cost of such a massive cyber episode.

2. EPIDEMIOLOGICAL MODELS WITH NETWORKS EFFECTS

In order to propose a simple and flexible approach, we propose to model the strength of the cyber pandemic on the global population. Subsequently, the impact on an insurance portfolio is considered, assuming that contamination is more likely to come from outside the portfolio than from inside. This seems reasonable, based on the fact that a portfolio is in fact small when compared to the global population among which the cyber epidemic spreads.

2.1 Model on the global population

The construction of accumulation scenarios is based on stochastic epidemiological contagion models adapted to the context of cyber risk, similar to the virus contagion models like those used for the COVID-19 pandemic. Barrier measures, such as vaccinations, are replaced here by other preventive measures, such as identifying and correcting vulnerabilities. The risk of the saturation of intensive care services is replaced by the risk of being unable to provide all the necessary assistance to the insured, which could lead to an aggravation of the total costs.

Nevertheless, despite the analogy between cyber and biological epidemics, there are still differences, particularly in terms of timescales, parameter values, and the nature of the risk. Consequently, the existing models need to be adapted to the cyber context. In particular, the heterogeneity of the population (for example, in terms of security levels or of assets that can be targeted by hackers, etc.) may have an important impact on the spread of the contagion. Thus, our model relies on a multi-group SIR model (susceptible, infected, recovered) [Kermack and McKendrick (1927)]. In this model, the population is decomposed into d categories (for example, representing different sectors of activities), and the population[2] within each category $j \in [1, d]$ is split into three groups $[s_j(t), i_j(t), r_j(t)]$, where for any date $t \geq 0$:

- The "**susceptibles**" $[s_j(t)]$ are the entities in sector j (at date t) that can be impacted by the ongoing cyberattack.
- The "**infected**" $[i_j(t)]$ are former susceptibles of sector j that became "infected" by the cybervirus and that are contagious.
- The "**removed**" $[r_j(t)]$ are former infected of sector j that stopped participating in the contamination (because, for example, countermeasures have been adopted).

Then the dynamics of the population in each group is given by the following systems of ordinary differentials equations (presented in the Appendix A), where,

- The matrix $\mathbf{B} = (\beta_{k,j})_{1 \leq k,j \leq d}$ (not necessarily symmetric) conveys the information on how class k contaminates class j. This matrix is the key element of the model to capture the network topology.
- The vector $\mathbf{A}(t) = (\alpha_j(t))_{1 \leq j \leq d}$ represents a latent form of attacks (not contagious).
- The vector $\mathbf{H}(t) = (\eta_j(t))_{1 \leq j \leq d}$ represents a protection component against the threat, which diminishes the rate of new infections through time.
- The vector $\Gamma(t) = (\gamma_j(t))_{1 \leq j \leq d}$ represents the recovery rate.

By recovery, we do not mean "full recovery" (that is retrieving the same level of activity): the timescale for full recovery may be much longer than the duration of the crisis (weeks or months, compared to days). Note that this model encompasses wider situations than cyber contagion, such as, for example, a break in the supply chain (in such situations, matrix **B** generates a chain of dependence between different sectors of the activity).

At the global population level, the total number of victims from a cyber incident is computed by solving a fixed point equation whose solution can be easily determined numerically [Hillairet et al. (2021)]. Then, measuring the total number of infected individuals in each group of the population (depending on the starting point of the infection) allows us to better understand

[2] Assuming the global size of the population is constant (equal to N), which seems reasonable for a cyber crisis that only lasts a few days.

the impact of connectivity between classes and to quickly calibrate or assess the impact of such an episode.

2.2 From the multi-group SIR to the impact on an insurance portfolio

The multi-group SIR defined in Section 2.1 describes the dynamic of the cyberattack on a large population. On the other hand, an insurance portfolio is of a smaller size and can be understood as a random sample of individuals from the global population. Denoting T_m the infection date of a policyholder m (belonging to category $x_m \in [1, d]$), T_m is then a random time characterized by its hazard rate λ_{T_m} (that may be infinite):

$$\lambda_{T_m}(t) = \lim_{dt \to 0+} \frac{P(T_m \in [t, t + dt] \mid T_m \geq t)}{dt}$$

λ_{T_m} reflects the severity of the cyber-contagion at a global level, depending on the category x_m; it is given by the probability of selecting a newly infected individual among the individuals of the global population, that is

$$\lambda_{T_m}(t) = \lambda(t,j) = n_j(t)\,\{\alpha_j(t) + \textstyle\sum_{k=1}^{d} \beta_{kj} i_k(t)\} \text{ if } x_m = j$$

Then the average number of infected policyholders of category j in the portfolio (denoting n_j the size of category j in the portfolio) is given by:

$n_j(1 - \exp\{-\int_0^\infty \lambda(t,j)dt\}) = n_j\, v_j$ with a variance of
$n_j \nu_j(1 - \nu_j)$ [Hillairet et al. (2021)].

In addition to a partial protection (for example by increasing awareness of the threat) modeled through the parameter **H**, in some cases a perfect protection is possible, by implementing patches or antivirus. We model this by an independent random variable C_m that represents the time at which the policyholder m implements security changes that make them immune to the attack. As for T_m, C_m is modeled through its hazard rate λ_{C_m} and acts like a censoring-variable: denoting $\delta_m = 1_{T_m \leq C_m}$, $\delta_m = 0$ indicates that immunity has been acquired, before contamination has occurred.

The aim of this paper is to analyze the impact of the network structure and of partial or full protection measures on the spread of the attack. But before we deal with that, one important and challenging task that needs to be undertaken is calibrating the model, or at least determining reasonable numerical values for the parameters of the equations in Appendix A. We now describe the heuristic we have developed to mimic a Wannacry-type incident and its propagation, with a network structure based on OECD data.

3. NUMERICAL IMPLEMENTATION

Determining reasonable values for the parameters is a difficult task due to the lack of public data on the network structures as well as on the real-time evolution of a cyber crisis. We first consider the model in Appendix A with no reaction (that is $\eta_j = 1$ and $C_m = \infty$ for all j and all m).

3.1 Connectivity between sectors

We give an example of calibration of the network based on macroeconomic data of the OECD [OECD (2018)], to identify the dependence between some sectors of activity, namely the categories of mining, manufacturing, energy, construction, and services. Although we admit that OECD data do not provide a very accurate vision of the connectivity between these sectors, our aim is to determine a reasonable benchmark and to show that plausible parameters may be obtained through the use of a relatively small amount of data. Assuming that the digital flow between these categories is somehow proportional to the economical flow, and after a normalization by the number of companies in each category, we obtain the following connectivity matrix $\mathbf{B_0}$, with the sum of all coefficients equal to 1 [see Lopez et al. (2021) for more details on the computation of $\mathbf{B_0}$].

3.2 Simulation of a Wannacry-type event

In the dynamics described by equations in the Appendix A, we consider the contagion matrix $\mathbf{B} = \beta\mathbf{B}_0$, where parameter β captures the intensity of the contagion, is calibrated on a cyber

Table 1: Normalized connectivity matrix B_0

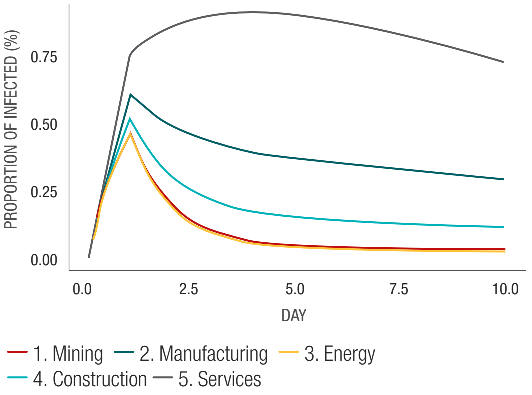
	MINING	MANUFACTURING	ENERGY	CONSTRUCTION	SERVICES	TOTAL
MINING	0.0634	0.2927	0.0449	0.1427	0.1255	0.6692
MANUFACTURING	0.0063	0.0527	0.0027	0.0108	0.0351	0.1076
ENERGY	0.0135	0.0370	0.0571	0.0150	0.0452	0.1679
CONSTRUCTION	0.0019	0.0068	0.0007	0.0141	0.0091	0.0326
SERVICES	0.0003	0.0042	0.0004	0.0017	0.0161	0.0227
TOTAL	0.0855	0.3934	0.1057	0.1844	0.2309	1

event similar to Wannacry. The Wannacry attack [May 2017, see Mohurle and Patil (2017)] is particularly emblematic due to the important number of computers infected around the world [more than 300,000 according to Chen and Bridges (2017)]. The attack consisted of a ransomware introduced into the systems through a well-documented vulnerability of Microsoft Windows [EternalBlue exploit, see Kao and Hsiao (2018)]. In this Wannacry episode, the susceptibles were computers vulnerable to the Eternal Blue exploit, but whose total number is hard to track – in fact, even the exact number of computers equipped with a given operating system is impossible to obtain. Consequently, we rely on indirect information about the total number of victims, the length of the episode (approximatively 10 days), and its dynamic (namely the timeline of the payments of ransoms, which is publicly available due to the use of the Bitcoin protocol). To ignite the epidemic, we consider a burst of infections caused by the hackers that strike the victims at uniform rate α_0 during one day: $\alpha_j(t) = \alpha_0 1_{t\leq 1}$ for all j. We take $\gamma = 1$, which corresponds to a fast containment (approximately 1 day) preventing the cyberattack to spread. This order of magnitude seems reasonable for the case of non-silent infections by malwares: once the victims identify they are attacked, links with the rest of the network may be easy to cut. This leads to the following set of parameters described in Table 2.

Table 2: Parameters used to simulate a Wannacry-type episode

PARAMETER	VALUE
α_0	7×10^{-3}
β	1.845×10^{-5}
γ	1
N	4,064,279

Figure 1: Evolution of the proportion of infected – Uniform bombing



3.3 Numerical results

We first compute the evolution through time of the infected in each category, as reported in Figure 1. We can observe that the peak of infections is not located at the same time (it is achieved later for services, with a slower decay).

We then investigate the vulnerability of the different sectors, by concentrating the initial attack on a given sector j (that is $\alpha_j(t) = \alpha^{(j)} 1_{t\leq 1}$, and $\alpha_k(t) = 0$ for $k \neq j$). To make things comparable, we take $\alpha^{(j)} = \alpha_0/p_j$, where p_j is the proportion of sector j in the global population. We compare it to the case of a uniform attack α_0 on all sectors. The proportions of companies (sector by sector) affected by the epidemic, depending on the targeted sector, are given in Table 3.

We observe that the mining sector seems to be the most contagious one. This can also make sense from a supply-chain modeling perspective. Nevertheless, this high contagiousness is to be tempered by the small population size of this sector.

4. IMPACT OF REACTIONS TO THE ATTACK

4.1 Reactions providing partial protection

We first consider the case where, during the crisis, a reaction of some categories can occur to lower the infection rate and to reduce the impact of the episode. In the Wannacry case,

Table 3: Proportion of infected sector by sector, depending on the targeted sector

TARGETED SECTOR	MINING	MANUFACTURING	ENERGY	CONSTRUCTION	SERVICES
Uniform attack	1.06%	4.11%	0.99%	2.07%	8.86%
Attack on Mining	99.70%	12.69%	1.36%	5.49%	20.37%
Attack on Manufacturing	1.02%	16.01%	0.66%	3.05%	16.58%
Attack on Energy	0.93%	5.96%	64.08%	2.35%	12.93%
Attack on Construction	0.33%	2.49%	0.21%	6.60%	5.72%
Attack on Services	0.25%	2.59%	0.21%	1.01%	7.84%

Table 4: Impact of the reaction on the number of victims

ρ = 10%	s = 10,000		s = 50,000	
	TOTAL	COLLATERAL	TOTAL	COLLATERAL
Mining	99.80%	99.99%	99,83%	99.99%
Manufacturing	94.60%	96.99%	95.82%	97.82%
Energy	99.81%	99.98%	99.84%	99.98%
Construction	98.51%	99.40%	98.87%	99.59%
Services	73.10%	77.62%	80.40%	84.36%
ρ = 50%	s = 10,000		s = 50,000	
	TOTAL	COLLATERAL	TOTAL	COLLATERAL
Mining	98.97%	99.90%	99.14%	99.93%
Manufacturing	76.87%	86.55%	81.92%	90.19%
Energy	99.03%	99.88%	99.21%	99.92%
Construction	92.99%	97.14%	94.66%	98.03%
Services	30.04%	38.29%	45.65%	54.04%

Depending on the sector which reacts (only one sector at a time) and on the thresholds activating the reaction, in case of an uniform initial attack.

for example, a "kill switch" was identified [Mohurle and Patil (2017)] that made it possible to diminish its severity. To illustrate this, we assume that the threat draws the attention of category j and is considered worth taking measures only if a sufficient number (namely s) of victims have been hit. This translates into the model presented in Appendix A, by introducing the function η_j (corresponding to the reaction of category j) given by $\eta_j(t) = 1 - \rho\ \Sigma^{d}_{k=1}{}^{i}{}_{k(t)\geq s}$.

We consider two levels of protection, $\rho = 0.1$ and $\rho = 0.5$, and two different thresholds of reaction $s = 10{,}000$ and $s = 50,\ 000$. Table 4 shows the impact of reaction in case of a uniform initial attack, and when only one single sector reacts. The column "Total" shows the ratio between the number of victims if reaction, over the number of victims without reaction. The column "Collateral" shows the ratio of the number of victims in the sectors that do not react, over the number of victims in these sectors if there is no reaction at all.

One observes that the reaction having the most important impact is the one on the services sector. As this sector contains the largest number of companies, this reduction of the size of the cyber epidemic is first of all caused by the fact that fewer companies in this sector are infected, due to the reaction. But it is also interesting to notice that this induces effects in the other sectors too, since the collateral gains are quite important too.

4.2 Reactions providing full protection

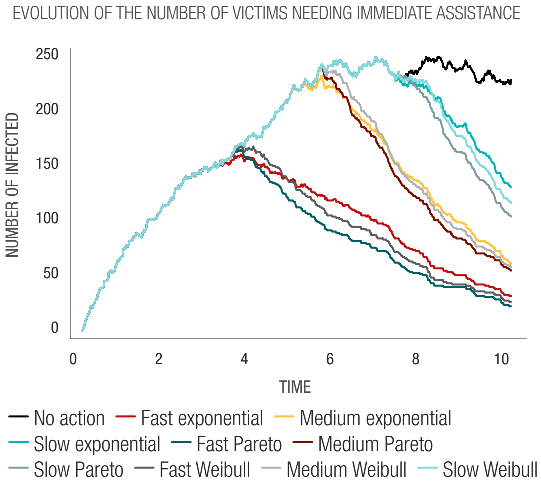
We now consider the case of an insurance portfolio of n policyholders representative of the global population. The policyholders have the possibility to implement (after some delay τ) an antivirus that provides immunity against the attack. This is captured by the random variable C (as in $\delta_m = 1_{T_m \leq C_m}$) modeled by three types of hazard rate:

- A translated exponential distribution. This means that, once the response has begun, the proportion of policyholders per time who update their security system is constant through time.
- A Pareto-type distribution. This corresponds to a situation where the vigilance of the policyholders decreases through time.
- A Weibull-type situation where there is a progressive attention devoted to this threat among policyholders.

In each case, the parameter τ represents the reactivity of the response. Figure 2 provides a simulated trajectory of the number of policyholders requiring immediate assistance, for $n = 10{,}000$ exposed policies and for three delays of reaction: a fast response ($\tau = 3$ days after the start of the event), a medium response ($\tau = 5$ days), and a slow response ($\tau = 7$ days).

The size of this peak can be of some concern, as pointed in Hillairet and Lopez (2021), since many cyber insurance contracts are supposed to provide immediate assistance to their policyholders when hit. However, a very high peak

Figure 2: Dynamics of the number of policyholders requiring immediate assistance



could lead to a situation where it might be impossible to deliver the service that was contractually guaranteed. In addition, if assistance comes too late due to saturation, this could increase significantly the amount of damages. We see that a slow response will hardly diminish the burden of the assistance teams, while a fast response in three days significantly reduces the magnitude of the peak of the attack.

5. CONCLUSION

In this paper, we propose a general and flexible model for constructing cyber-hurricane scenarios, taking into account some network structures and analyzing the impact of protection measures. In the numerical part, we use a rough connectivity matrix inferred from macroeconomic data of OECD and we mimic an event similar to the famous Wannacry episode. We emphasize the flexibility of the model, which can be easily adapted to various network structures and various scenarios. In particular, this model can be used to quantify the benefits of a reaction to such a crisis. Indeed, behavioral studies is determinant to evaluate the risk that the system collapses.

APPENDIX A: ORDINARY DIFFERENTIALS EQUATIONS MODELING THE DYNAMICS OF THE POPULATION IN EACH GROUP

$$\frac{\mathrm{ds}_j(\mathrm{t})}{dt} = -\eta_j(\mathrm{t})\ \{\alpha_j(\mathrm{t}) + \textstyle\sum_{k=1}^{d} \beta_{\mathrm{k,j}}\, \mathrm{i_k}(\mathrm{t})\}\mathrm{s_j}(\mathrm{t}),$$

$$\frac{\mathrm{ds}_j(\mathrm{t})}{dt} = \eta_j(\mathrm{t})\ \{\alpha_j(\mathrm{t}) + \textstyle\sum_{k=1}^{d} \beta_{\mathrm{k,j}}\, \mathrm{i_k}(\mathrm{t})\}\mathrm{s_j}(\mathrm{t}) - \gamma_j \mathrm{i_j}(\mathrm{t})$$

$$\frac{\mathrm{dr}_j(\mathrm{t})}{dt} = \gamma_j \mathrm{i_j}(\mathrm{t})$$

REFERENCES

ANSSI, 2021, “Etat de la menace rançongiciel,” Agence nationale de la sécurité des systèmes d’information, https://bit.ly/3JUq9L4

Beretta, E., and V. Capasso, 1988, “Global stability results for a multi-group SIR epidemic model,” in Hallam, T. G., L. J. Gross, and S.A. Levin (eds.) Mathematical Ecology (eds.), Springer

Chen, Q., and R. A. Bridges, 2017, “Automated behavioral analysis of malware: A case study of Wannacry ransomware,” 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 454–460

Chen, H., and S. H. Cox, 2009, “An option-based operational risk management model for pandemics,” North American Actuarial Journal 13:1, 54–76

Cyence, 2017, “Counting the cost – cyber-exposure decoded,” https://bit.ly/3tplQ2K

Farkas, S., O. Lopez, and M. Thomas, 2021, “Cyber claim analysis using generalized pareto regression trees with applications to insurance,” Insurance: Mathematics and Economics 98, 92–105

Fahrenwaldt, M. A., S. Weber, and K. Weske, 2018, “Pricing of cyber insurance contracts in a network model,” ASTIN Bulletin: The Journal of the IAA 48:3, 1175–1218

Garrido, J., and R. Feng, 2011, “Actuarial applications of epidemiological models,” North American Actuarial Journal 15:1, 112–136

Guo, H., M. Y. Li, and Z. Shuai, 2006, “Global stability of the endemic equilibrium of multigroup SIR epidemic models,” Canadian applied mathematics quarterly 14:3, 259–284

Hillairet, C., and O. Lopez, 2021, “Propagation of cyber incidents in an insurance portfolio: counting processes combined with compartmental epidemiological models,” Scandinavian Actuarial Journal 6, 1–24

Hillairet, C., O. Lopez, L. d’Oultremont, and B. Spoorenberg, 2021, “Cyber contagion: impact of the network structure on the losses of an insurance portfolio,” working paper, https://bit.ly/35mPyhH

Kao, D.-Y., and S.-C. Hsiao, 2018, “The dynamic analysis of Wannacry ransomware,” 20th International conference on advanced communication technology (ICACT), pp. 159–166. IEEE

Kermack, W. O., and A. G. McKendrick, 1927, “A contribution to the mathematical theory of epidemics,” Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 115:772, 700-721

Kshetri, N., 2020, “The evolution of cyber-insurance industry and market: an institutional analysis,” Telecommunications Policy 44:8, 102007

Lopez, O., L. d’Oultremont, and B. Spoorenberg, 2021, “Modeling accumulation scenarios in cyber risk,” Detra Note, Detralytics, https://bit.ly/3C2UdRN

Lefèvre, C., P. Picard, and M. Simon, 2017, “Epidemic risk and insurance coverage,” Journal of Applied Probability 54:1, 286–303

Magal, P., O. Seydi, and G. Webb, 2018, “Final size of a multi-group SIR epidemic model: irreducible and non-irreducible modes of transmission,” Mathematical biosciences 301, 59–67

McKendrick, A. G., 1925, “Applications of mathematics to medical problems,” Proceedings of the Edinburgh Mathematical Society 44, 98–130

Mohurle, S., and M. Patil, 2017, “A brief study of Wannacry threat: ransomware attack 2017,” International Journal of Advanced Research in Computer Science 8:5, 1938–1940

OECD, 2018, “Origin of value added in final demand,” https://bit.ly/3tfLoAo

Romanosky, S., L. Ablon, A. Kuehn, and T. Jones, 2019, “Content analysis of cyber insurance policies: how do carriers price cyber risk?” Journal of Cybersecurity 5:1

Xie, X., C. Lee, and M. Eling, 2020, “Cyber insurance offering and performance: an analysis of the US cyber insurance market,” The Geneva Papers on Risk and Insurance-Issues and Practice 45:4, 690–736

CYBER INSURANCE AFTER THE RANSOMWARE EXPLOSION – HOW IT WORKS, HOW THE MARKET CHANGED, AND WHY IT SHOULD BE COMPULSORY

JAN MARTIN LEMNITZER | Department of Digitalization, Copenhagen Business School

ABSTRACT

For two decades, the cyber insurance sector had been a niche sector of the insurance industry: tiny but boasting strong growth rates and enormous profit ratios. Yet, between 2019 and 2022, the cyber insurance industry has been devastated by the impact of the explosion in ransomware, causing huge payouts and escalating losses. Some insurers are now fleeing from the sector entirely. This article will shine some light on how the cyber insurance industry works and how it has responded to the ransomware impact. After discussing why insurers struggle with accurately pricing the cyber risks posed by the companies in their portfolios, it will explore the evidence in support of the claim that having cyber insurance improves a company's IT security. The final section offers a radical proposal to make cyber insurance compulsory for small- and medium-sized companies (SMEs) to tackle their known and longstanding issues with IT security. If combined with an externally established minimum IT security standard developed for SMEs and light regulation on insurance policies, this measure could transform IT security in thousands of companies and vastly improve their resilience against ransomware and other cyberattacks.

1. INTRODUCTION

For two decades, the cyber insurance sector had been a niche sector of the insurance industry: tiny, at less than 1 percent of the size of the greater property and casualty insurance market but boasting strong growth rates and enormous profit ratios [IST (2021)]. This growth accelerated further as many more businesses sought cover after the double shock of NotPetya and WannaCry in 2017. Yet, between 2019 and 2022, the cyber insurance industry has been devastated by the impact of the explosion in ransomware, causing huge payouts and escalating losses. Some insurers are now fleeing from the sector entirely.

This article will explain what the cyber insurance industry offers to clients, how it was hit by ransomware, and how it is responding. To explain how an entire branch of insurance could end up mispricing its products and underestimating risks, section 3 will look at how insurers set premiums and measure the cyber risks posed by the companies in their portfolios, and why they find the task extremely challenging. Section 4 will explore the evidence to support the claim that having cyber insurance improves a company's IT security. The final section will develop a radical proposal to make cyber insurance compulsory for small- and medium-sized companies (SMEs) to tackle their known and longstanding issues with IT security. If combined with an externally set IT security minimum standard developed for SMEs and light regulation on insurance policies, this measure could transform IT security in thousands of companies and vastly improve their resilience against ransomware and other cyberattacks.

2. WHAT IS CYBER INSURANCE?

Insurance policies covering cyber risk offer companies protection against the escalating costs related to a network breach or successful cyberattack and are sold either as part of a company insurance policy or (as is increasingly common) as a standalone cyber insurance policy. As such, cyber insurance is a risk management practice that transfers residual risk after all other available sensible measures to reduce an organization's cyber risk have been exhausted. Used wisely in conjunction with sensible IT security practices, cyber insurance can provide crucial cover against catastrophic breaches whose consequences might otherwise endanger the survival of the company.

Moreover, good cyber insurance policies offer much more than simply the chance to claim back damages. Next to the financial coverage, they provide access to support services that can be critical in containing and overcoming a cyberattack. Companies will be able to call a specific phone number 24/7 and request the immediate support of a team of sophisticated cybersecurity professionals at the insurer's expense. While the details vary between policies, an increasing number of them are also offering the services of specialists dealing with client data, GDPR exposure, and client management, as well as consultants for branding and media reputation that can support communications with the public and clients about the incident. Moreover, insurers provide quality control for incident responders: they will only call in IT companies who have proven themselves in previous assignments, while a single company looking for post-breach support will find it much harder to decide which IT service providers they can trust in their moment of crisis [Woods and Böhme (2021)]. SMEs will find it impossible to assemble a similar support set-up at short notice and at their own expense.

2.1 Why do so many companies choose not to have cyber insurance?

Although cyber insurance policies have been commercially available for more than two decades, less than 15 percent of organizations globally buy cyber insurance [IST (2021)]. The market is still lopsided and unsettled: the U.S. is by far the largest market for cyber insurance policies, with about 90 percent of all premiums written there, and Europe and Asia making up the remaining 10 percent [OECD (2017)]. One key reason for this difference is that starting with California in 2003, all U.S. states have introduced laws requiring notification of data breaches [Lubin (2019)]. Recent increases in European companies seeking coverage might, therefore, be as much driven by the introduction of E.U. data protection legislation in May 2018 (especially since GDPR comes with huge potential fines for data breaches) as it is by the increasing cyber threat.

One important caveat is that while cyber insurance is a widely used tool among large companies for managing their cyber risks, it remains a niche product for the many smaller- and medium-sized companies (SMEs) that make up a large part of the economy. There are several reasons for that: firstly, small company boards tend to believe that cyberattacks are something that happen to large companies and not to them. Unfortunately, the common view that SMEs are not targeted by ransomware gangs is manifestly false: in the first half of 2020, almost half of all cyber insurance claims came from SMEs [Cimpanu (2020)]. Predictably, direct personal experience of a cyberattack has been identified as a key driver of insurance uptake in this group [Bernard (2020)]. Moreover, company leaders find the wording and coverage details of cyber insurance policies highly confusing – privately, insurance brokers will agree [Insurance Journal (2017)]. Insurers are acutely aware that there are serious problems with the definitions used in the various policies to describe what kind of damage is covered and what is not, especially given the fast-changing market conditions [Rawlings (2014), Kesan and Hayes (2017)]. For example, the terms "data loss" or "data breach" may have different meanings in different policies, making them quite hard to compare [ENISA (2016b)]. Other terms, such as "cyber terrorism", are completely undefined [GAO (2021)]. Insurers know that a more unified approach to policy language would be preferable, but are wary of the huge, concerted effort that would be necessary across the industry. Moreover, a global solution is especially complex since different countries also have their own legal traditions, with specific legal concepts and insurance industry terms based on decades of court precedents. The wide variations of coverage and policy terms suggest a market that is still unsettled [Xie et al. (2020)].

Moreover, many policies list so many exclusions and duties for the policyholder that businesses get concerned about how easy it would be for an insurance provider to find negligence or other behavior breaching the policy [the model policy provided by the German insurance industry is a good example, see GDV (2017)]. This contributes to a general skepticism among smaller companies about whether such policies can be trusted and will pay out in full in the hour of need. There is some hope that the ransomware epidemic might provide some assurance here: cyber insurers ran into trouble with their portfolios because they paid out so much, not because furious hacked companies canceled their policies [Woods (2022)].

Crucially for budget-strapped small companies, signing up to cyber insurance can require serious effort. This is especially true if the company has never previously conducted a systematic assessment of its own network, patching procedures, and cyber risk exposure. In addition, company leaders know that insurers might demand the replacement of outdated software or IT infrastructure, which can result in considerable expenses. Taken together, these factors mean that while bigger companies have IT departments and usually at least some cyber insurance policy in place, many small companies that would benefit the most from IT guidance, support services, and financial cover do not.

2.2 Cyber insurance and the ransomware impact

Today, the daily reports of companies falling victim to ransomware are persuading companies of all sizes to apply for cyber insurance for the first time or raise the coverage limits on existing policies. Unfortunately, the escalating payouts caused by the ransomware problem have led insurers to make drastic changes to their portfolios. While the first wave of ransomware, targeting companies by encrypting all their data, could be countered by better backup practices, the second wave is practicing a double extortion approach: by threatening to leak stolen internal or client data (which may lead to substantial fines under data protection law, not to mention upset clients) the ransomware gangs are persuading companies to pay up even if they have recent backups. As it turned out, there is no easy fix to counter this extortion scam. In the first half of 2020, insurer Coalition experienced an increase in ransomware claims of 260 percent, with the average ransom demand rising by almost 50 percent [IST (2021)].

That meant that insurers had to adjust their business models. Most of them raised premiums by 30-40 percent or more in the first half of 2021, decreased the maximum coverage limits on offer, or included new sub-limits for ransomware damage [Cohn (2021)]. In the third quarter of 2021, the price rises reported by Marsh reached an astonishing 96 percent for the U.S. market and 76 percent in the U.K., strongly suggesting that we have not as yet reached the peak of the ransomware epidemic [Marsh (2021)]. A report by the U.S. Government's General Accounting Office, published in May 2021, confirms this picture: while there is an increasing demand for policies by businesses and organizations, prices are much higher and coverage limits lower than they were in recent years. Some sectors that have been hit especially hard by ransomware attacks due to their highly sensitive data and known poor IT security practices, such as healthcare and education, are having real difficulties finding insurers that will cover them [GAO (2021)]. Following the highly publicized Solarwinds and Kaseya hacks, "managed service providers" (MSPs) are also experiencing similar problems. Given that they offer remote IT security management services for multiple clients' networks, the payouts when they are being hacked will be enormous. Consequently, they now face extremely high insurance premiums [Bay and Pruger (2021)].

Some insurers are even questioning the viability of the entire product, have stopped adding new customers to their portfolios, or decided to leave the market entirely [IST (2021)]. However, this phenomenon seems to be limited to smaller insurers who saw cyber insurance as an easy way to create income and growth by offering policies written and backed by major reinsurers and without investing in their own cyber expertise. As Woods (2022) states, "for the first two decades, the cyber insurance market rewarded entrepreneurial insurers who embraced uncertainty (or ignorance) while offering innovative insurance products." In other words, the ignorant got rich insuring the careless while the sun was shining. Then it rained, and hard, forcing many of these types of players to leave the market. That is why we might ultimately come to view these large ransomware events as a healthy moment for the cyber insurance market, when it matured and providers without deep knowledge of cyber risk who had previously pulled down prices or security requirements were weeded out.

Yet, this new, more mature market suggests that increased security requirements and higher prices are here to stay, as those insurers that stayed on have fundamentally reevaluated the risks they are taking on [IST (2021)]. In this new environment, it will become increasingly harder for small businesses to persuade insurers to provide them with the protection they need. The recommendations in the final section will address this problem, offering suggestions on how an externally set minimum cybersecurity standard for SMEs could provide the necessary clarity about mutual expectations. However, even if it might turn out to be a good thing for the market in the long run, this market contraction certainly raises questions regarding why cyber insurers were unable to see the wave of ransomware claims coming in advance. The next section will look at how insurers evaluate and price the cyber risks posed by the companies in their portfolios.

3. CYBER RISK ASSESSMENT

While it is not something usually mentioned in sales pitches, insurers have long known that company cyber risk is a very different beast to many of the other risk categories that they traditionally deal with. A key concern is that the usual approach of predicting future risks by amassing historical claims data has limited utility in cyber insurance. Companies only report network breaches if legislations force them to do so, and insurers do not share their claims data with competitors. Even if you have industry-leading knowledge about which industries were facing what chance of being hit by a cyberattack between 2005 and 2015, how much value does this information have for predicting the likelihood that a specific company will file a huge claim on their cyber insurance policy in 2023? Unsurprisingly, a key concern of the literature on cyber insurance is how to accurately price and manage cyber risk [Romanowski et al (2019), Khalili et al. (2019), Xu and Hua (2017)] given the lack, and limited reliability, of data on historic or recent claims and losses [Boyer (2020), Eling (2018), Marotta et al. (2017)].

Looking at how insurers gather data on their clients reveals a market split in two, with a high-end section offering bespoke arrangements for large businesses but demanding considerable scrutiny, and a budget product that is offered off the shelf to smaller customers who only need to undergo a very superficial audit before receiving their policies. At the high-end level, companies often buy so-called stacks or towers of insurance, where a huge coverage sum, reaching hundreds of millions of euros, is jointly guaranteed by multiple insurers and/or re-insurers. Consequently, insurers must make three separate decisions: 1) Do we want to insure this company; 2) what is the right price for insuring this company; and 3) where would we like to be in the tower: near the top, were we only need to pay out once the client claims their maximum coverage, or near the bottom, where we would be among the first to pay out but can command higher premiums?

Insurers collect data from multiple public or private sources on the company, send them detailed questionnaires about IT security practices and governance, and discuss the answers with the board and the IT department leadership. In some cases, they will also send one of their senior cyber underwriters to conduct an onsite audit [MacColl et al. (2021)]. This approach makes no economic sense for smaller companies, as the insurer would have to invest several years' worth of premiums to pay for this kind of extensive audit. Usually, smaller companies simply fill out a questionnaire, but insurance industry insiders do not like to discuss the level of scrutiny with which their answers are treated. After conducting dozens of interviews, as part of a wider research project on cyber insurance, Sullivan and Nurse (2020) conclude that almost no meaningful data on the IT security practices of small companies is gathered when signing them up for cyber insurance.

Insurers also use so-called outside-in rating agencies to assess company cyber risk. These companies will run a "vulnerability scanner" to scan a company network from the outside to identify vulnerabilities, patching regularity, open ports, and email security. This is in principle a very useful thing to do, as it mirrors the behavior of hackers and cyber criminals who run similar scans to identify potential victims. The rating agencies then employ an algorithm to quantify the results and combine them with data about the company from commercial providers or the dark web. The result is a "cyber risk rating score", which in theory allows the insurer or third-party risk manager to understand the company's cyber risk at a glance and base business decisions on this score [MacColl et al. (2021)].

Companies offering this kind of technology (such as BitSight, Security Scorecard, and RiskRecon) have seen huge growth in the insurance sector in recent years as their ratings offer a more comprehensive and reliable picture of a company's cyber risks than a short questionnaire. Moreover, the products are designed to be run at scale, meaning the cost of checking on an individual company is low. This explains why insurers were pioneer customers of these products before they began to become more popular in third-party risk and supply chain management.

Unfortunately, out-side in rating scores come with important inherent limitations to their scope and reliability. While a bad rating score makes it highly likely that there are serious cybersecurity issues at the company, a good rating does not necessarily mean that company IT security is handled well, and that the company poses a low cyber risk. The rating score can only include what is observable from the outside, or available in public or private databases. It reveals next to nothing about a vast range of key IT security issues within the company, ranging from systems and network configuration to staff training or incident response planning. Insurers need to know whether cybersecurity is something that is taken seriously by the company board and operationalized with clearly attributed responsibilities. Consequently, cyber risk ratings should not be used as the single data point to drive a business decision, especially since there are also issues with occasionally incorrect attributions of IP addresses to companies (so-called false positives) that are not rectified because the company in question does not know this rating exists. Yet, insurers will privately admit that this is happening for insurance decisions relating to small companies. Once cyber risk ratings are accepted as a de facto standard for third party risk management, financial, or investment decisions, we might even see a situation reminiscent of the corporate credit rating market where a small number of U.S. companies dominate the markets and set the standards for how companies are measured and evaluated [Lemnitzer (2020)].

4. DOES CYBER INSURANCE IMPROVE COMPANY IT SECURITY?

Having established how hard it is for insurers to assess company cyber risk and why companies struggle to find and buy the right cyber insurance policy for themselves, the question arises whether having cyber insurance has a measurable positive effect on company cybersecurity. Does having cyber insurance simply mean a company pays money to transfer risk and receive access to support services, or does it also tend to initiate a process that leads to improved cybersecurity performance? This paper is far from the first to pose that question, and as Woods and Moore (2020) note, there are two decades' worth of research on whether insurance improves security. In the absence of universally agreed and comparable measurements of company IT security performance, what researchers attempt to do is to find out whether a company is less likely to experience a network breach when it is insured.

Unfortunately, conducting such research comes with inherent methodological difficulties: there are no public registers of insured or uninsured companies, and the vast majority of company breaches are never reported to regulators or the public. Both insurers and their clients have good reasons to be rather private about data regarding market reach, insurance claims, or their experience with security breaches. It is also difficult to do comparative work since there are practical, as well as ethical, issues regarding maintaining a control group of uninsured companies to measure their susceptibility to cyber-attacks while trying to identify whether the sample of insured companies do better.

As a result, it becomes difficult to replicate, or even evaluate, the results of studies conducted internally by insurers, even if they are published and not reserved for internal use. For example, the U.S. insurer Corvus recently reported that a vulnerability scanning tool it makes available to its clients had led to a 65 percent drop in ransomware-related claims from April to September 2020 [Abrams (2020)], which would be a direct improvement in security performance as a result of an insurance policy, but this is not a peer-reviewed study tested for its methodology. Many similar studies exist, but insurers usually chose not to make them publicly available. For these reasons, an extensive discussion on whether cyber insurance improves IT security concluded that the lack of data meant that the question could not be resolved with any degree of certainty. However, MacColl et al. (2021) found “a solid body of theoretical arguments that cyber insurance could play a meaningful role in improving cybersecurity among businesses.”

Most experts agree and point to a number of factors: firstly, the mere act of applying for insurance cover usually entails a requirement to fully consider a company’s cyber risk exposure and conduct an audit of its IT infrastructure and network configuration. It is recommended that companies should regularly conduct such exercises, though not all companies do it in practice. Secondly, some insurance policies also provide free access to IT security products or advice, which could potentially mean a marked improvement in company IT security, especially if implemented properly. Thirdly, the biggest benefit of a good cyber insurance policy are the support services that are available to clients in the event of a breach. Employed successfully, they benefit three different groups at once: the company stands a much better chance of dealing with the breach successfully, the insurer invests in these support services to limit the size of the eventual claim, and the economy as a whole is more secure as a cyberattack that is quickly contained by professionals is less likely to spread to other companies or institutions. This example also highlights the methodological problem that arises when we use the likelihood of being breached as the key variable to determine whether having an insurance policy improves company IT security. If the only “success” parameter of cyber insurance is reducing breaches, every breach is a fail. However, companies will get breached and limiting the damage and preventing the spread down the supply chains can be a key benefit of a good insurance policy. This effect is not captured by just looking at how many insured companies still get breached.

Most recently, there is anecdotal evidence that companies that have been denied insurance due to the recent hardening of the market have responded by improving their IT security measures before returning to re-apply for coverage. This is a recent observation and there is no solid empirical study of it yet, but it supports the view that this mechanism might be exploited systematically to improve company IT security by making cyber insurance compulsory.

5. RECOMMENDATIONS: REGULATED POLICIES, EXTERNAL STANDARDS, COMPULSORY INSURANCE FOR SMES

While fire insurance or third-party car insurance is compulsory in most countries, cyber insurance is not. Outside of tightly regulated industries, such as finance or critical infrastructure, company owners can largely handle their IT infrastructure as they see fit. As states are unwilling or unable to take the matter of corporate IT security under direct control, the idea of using the insurance industry as a regulator in this field has emerged [Trang (2017)]. To overcome the issue that not all private companies might want to buy the insurance policies offered by their new "regulators", it was proposed to simply make cyber insurance compulsory [Miller (2019)]. A recent RUSI report suggested that the U.K. government should promote the sector by making cyber insurance compulsory for all companies competing for government contracts [MacColl et al. (2021)]. Interestingly, these demands tend to come from researchers rather than insurers, who fear the aggregate risk of large cyberattacks hitting many insured parties at once. The Danish market leader Tryg is an exception in this regard and published a white paper calling for compulsory cyber insurance in December 2019 [Hübbe (2019)]. Indeed, the greatest potential in using cyber insurance to improve company IT performance lies in making it compulsory for small- and medium-sized companies (SMEs). While large companies with sophisticated IT departments will be able to look after themselves in case of a network breach, the audit function and support services that come with cyber insurance can make a fundamental difference to the ability of SMEs to prevent, contain, or survive being hacked [Lemnitzer (2021)].

> *"Requiring SMEs to sign up to cyber insurance offers the best solution for changing the practices at a huge number of companies in a relatively short period of time."*

We have known for a long time that SME cybersecurity is typically poor, and that despite the well-publicized hacks of businesses across the world and numerous government awareness campaigns, the vast majority of SMEs do not practice proper cybersecurity. A recent Hiscox report on cyber readiness puts about 75 percent of companies into its politely-worded "novice" category [Hiscox (2019)]. Data from Germany suggests that half of all small companies still have no incident response plans or any staff members explicitly responsible for IT security, and over 70 percent conduct no IT security training for their staff. Only a fifth of the companies surveyed fulfil the most basic requirements for secure IT systems [GDV (2020)]. This is a major issue since any attempt to achieve resilience within a modern digital economy will fall flat if such a large percentage of companies remain vulnerable to the most basic malware. After many years of relying on awareness campaigns, we know full well that they will not cause the drastic change of approach by SME company boards that is necessary.

We need to try something new and requiring SMEs to sign up to cyber insurance offers the best solution for changing the practices at a huge number of companies in a relatively short period of time. Once insurance becomes compulsory, companies must meet the required minimum IT security necessary to obtain cover or face a fine. Consequently, the key element necessary for the success of compulsory cyber insurance is to accompany it with a clear, externally set minimum IT security standard that both insurers and companies can refer to. This task should not be left to the insurers – variation between providers creates confusion and unpredictability for clients, and insurers might find economic incentives to water down standards to win market share or arbitrarily exclude certain groups of companies perceived as too risky.

Instead, the standard should be set by a trusted external body. The procedures and controls established by the various cybersecurity standards developed by the U.S. National Institute of Standards and Technology or the International Organization for Standardization (esp. ISO 27001) are a challenge to fully implement even for large companies with skilled IT departments. For SMEs they are simply too demanding in organizational scope and technological sophistication. The measures required by this new standard must be feasible to implement without extensive specialist IT knowledge, and they must come at a cost point that is manageable for smaller companies. At the same time, they must be carefully chosen to achieve the highest security gains at the lowest price.

The U.K. National Cyber Security Centre attempted to provide such a universal minimum standard with its "Cyber Essentials" certification program for small businesses, which is already used as a reference point by U.K. insurers. It has just been updated and will now demand multi-factor authorization, password management, and tighter security regarding the use of cloud services [Hill (2022)]. Combined with least privilege principles, network segmentation, breach response, and mandatory staff training it could serve as a good starting point for any country considering a minimum standard for SMEs. Australia's National Cyber Security Centre has embraced a much more ambitious approach with three different levels of cyber maturity adapted to company size. An alternative route would be to build up a nationwide cyber risk rating system like the one currently set up in Austria, which combines a vulnerability scan, an onsite audit, and a bespoke standard to rate and compare companies [Cyber Trust Austria (2021)]. Originally created to allow critical infrastructure companies to monitor their suppliers, the ultimate intention is to cover all Austrian businesses.

Moreover, the clarity and uniformity that this new standard needs to achieve should be matched by corresponding improvements to the wording of cyber insurance policies. The E.U.'s insurance oversight organization, European Insurance and Occupational Pensions Authority (EIOPA), is now on record demanding that minimum standards for policies should be set externally (in other words, by regulators), with insurers then competing over price or by providing extra coverage and features [EIOPA (2020)]. Next to clear language, policies must also offer clear guarantees: first, a company meeting the minimum standard must be able to rely on their claims being paid out in full once they are hit by malware. That excludes common tricks such as hiding a much lower sublimit for ransomware-related damages deep in the small print. Second, policies must include quick, easy, and reliable access to support services once a breach has occurred. This is a point that has been overlooked in the relevant reports and the specialist literature but is vital if we look at the insurance sector from a public policy or national security perspective. Some policies still do not include such services, while others put access to them at the discretion of the insurer. Neither should remain since easy access to professional tech support is one of the main advantages cyber insurance offers to SMEs. Finally, it should no longer be legal for insurers to cover ransom payments made by their clients. With some states already moving in this direction, it would make no sense to extend compulsory cyber insurance to many thousands of companies while allowing these policies to be used to pay off cyber criminals.

6. CONCLUSION

While the story of cyber insurance has long been one of continuous growth, the sector is now experiencing its first proper crisis as ransomware claims led to huge losses on formerly profitable portfolios. This has caused a spike in prices and a hardening of market conditions, which has unfortunately inhibited the increased take-up of cyber insurance policies among smaller companies that we might have expected following the introduction of GDPR in 2018 and the escalating ransomware threat. However, this new "harder" market will almost certainly be a healthier market where more insurers will have a deep understanding of cyber risk and establish specific security requirements for their clients without the fear of losing business to more lenient competitors.

This is a good development, but it also makes it harder for SMEs to obtain cyber insurance just when they need it most. While it has proven difficult to show a direct empirical connection between having cyber insurance and improvements in company IT security due to data and methodological constraints, a good case can be made that the financial cover, technical support, and post-breach incident response services offered by cyber insurance would be hugely helpful to SMEs in particular. At the same time, the increasing focus on cyber risk supply chain monitoring in larger companies, particularly those that are part of critical infrastructure, means it is becoming increasingly common to demand proof of cyber insurance before signing a contract with a supplier, just at a time when many SMEs find it harder to access cyber risk coverage as conditions tighten [Glover (2022)].

Frankly, this group of companies is struggling to meet basic IT security standards and will struggle to obtain insurance in the new market conditions. Yet, this is not just a problem for the individual companies: as long as a large number of SMEs remain so vulnerable, their connections to business partners and clients of all sizes means the security of the digital economy as a whole remains compromised. Something needs be done to support them in an environment where the threat from ransomware and state hackers is so severe, and if done the right way, making cyber insurance compulsory for this group of businesses might be the game changer that is required.

Compulsory cyber insurance for SMEs is a radical idea, but given that none of the awareness campaigns that were tried over the years has had a significant impact on security standards in smaller companies and the threat level due to ransomware and supply chain hacks keeps rising, something radical must be done. Moreover, compulsory insurance is accepted without controversy in other parts of business life, such as fire insurance or third-party car insurance. If compulsory cyber insurance is combined with an externally set minimum security standard designed with SMEs in mind and appropriate regulation of cyber insurance policies, it might well be the single best lever there is to significantly improve IT security in many thousands of companies in a short period of time.

REFERENCES

Abrams, L., 2020, "Cyber insurer's security scans reduced ransomware claims by 65%," Bleeping Computer, September 22, https://bit.ly/3sOBbvH

Bay, K., and M. Pruger, 2021, "The future of cyber insurance: what to expect in 2022," https://bit.ly/3MttMtq

Bernard, J., 2020, "Overcoming challenges to cyber insurance growth: expanding stand-alone policy adoption among middle market business," Deloitte, March 16, https://bit.ly/3HJhfi5

Boyer, M., 2020, "Cyber insurance demand, supply, contracts and cases," The Geneva Papers on Risk and Insurance – Issues and Practice 45: 559–563

Cimpanu, C., 2020, "Ransomware accounted for 41% of all cyber insurance claims in H1 2020." ZDNet.com, September 10, https://zd.net/3vFREUO

Cohn, C., 2021, "Insurers run from ransomware cover as losses mount," Reuters, November 11, https://reut.rs/3sY7TLj

Cyber Trust Austria, 2021, https://bit.ly/3pLgC15

Eling, M., 2018, "Cyber risk and cyber risk insurance: Status quo and future research," The Geneva Papers on Risk and Insurance – Issues and Practice 43: 175–179

EIOPA, 2020, "Cyber underwriting strategy," European Insurance and Occupational Pensions Authority, February 11, https://bit.ly/36Zcwfg

ENISA, 2016, "Cyber insurance: recent advances, good practices and challenges," European Union Agency for Cyber Security, https://bit.ly/3lPYtXr

ENISA, 2016b, "Commonality of risk assessment language in cyber insurance," European Union Agency for Cyber Security, https://bit.ly/3HPP1lB

GAO, 2021, "Cyber insurance: insurers and policyholders face challenges in an evolving market," U.S. Government Accountability Office, May 20, https://bit.ly/3KVCaAv

GDV, 2017, "Allgemeine versicherungsbedingungen für die cyberrisiko-versicherung (General insurance conditions for cyber risk insurance)," Gesamtverband der deutschen Versicherungswirtschaft, https://bit.ly/373EG8V

GDV, 2020, "Cyber-Risiken im Mittelstand 2020, (Cyber risks in SMEs)," Gesamtverband der deutschen Versicherungswirtschaft, https://bit.ly/35vDHy9

Glover, C., 2022, "The ransomware crisis is making cyber insurance harder to buy," TechMonitor, January 24, https://bit.ly/3lPZw9P

Hill, M., 2022, "UK NCSC updates Cyber Essentials technical controls requirements and pricing structure," CSO Online, January 7, https://bit.ly/3CmZZ0X

Hiscox, 2019, "Cyber readiness report," https://bit.ly/3lKtJap

Hübbe, M., 2019, "Lovpligtig forsikring mod cyber-angreb," Jyllands-Posten, December 2, https://bit.ly/3lSto5o

IST, 2021, "Combating ransomware – a comprehensive framework for action: key recommendations from the ransomware task force," Institute for Security and Technology, https://bit.ly/3MuUyBx

Insurance Journal, 2017, "Why 27% of U.S. firms have no plans to buy cyber insurance," May 31, https://bit.ly/3lPi2z3

Kesan, J. P., and C. M. Hayes, 2017, "Strengthening cybersecurity with cyberinsurance markets and better risk assessment," Minnesota Law Review 102: 191

Khalili, M. M., M. Liu, and S. Romanosky, 2019, "Embracing and controlling risk dependency in cyber-insurance policy underwriting," Journal of Cybersecurity 5:1

Lemnitzer, J. M., 2020, "Do we need and EU cybersecurity ratings agency?" EU CyberDirect Blog, November 10, https://bit.ly/3pJcqPz

Lemnitzer, J. M., 2021, "Why cybersecurity insurance should be regulated and compulsory," Journal of Cyber Policy, 1-19

Lubin, A., 2019, "The insurability of cyber risk," SSRN, https://bit.ly/3pNfha7

MacColl, J., J. R. C. Nurse, and J. Sullivan, 2021, "Cyber insurance and the cyber security challenge," Royal United Services Institute (RUSI) occasional papers, June 28, https://bit.ly/3MunidP

Marotta, A., F. Martinelli, S. Nannia, A. Orlando, and A. Yautsiukhin, 2017, "Cyber-insurance survey," Computer Science Review 24, 35-61

Marsh, 2021, "Marsh global insurance market index – 2021 Q3," https://bit.ly/3MuCAPP

Miller, L., 2019, "Cyber insurance: an incentive alignment solution to corporate cyber-insecurity," Journal of Law and Cyber Warfare 7, 147-182

OECD, 2017, "Enhancing the role of insurance in cyber risk management," Organisation for Economic Cooperation and Development

Rawlings, P., 2014, "Cyber risk: insuring the digital age," Journal of the British Insurance Law Association 128:1

Romanosky, S., L. Ablon, A. Kuehn, and T. Jones, 2019, "Content analysis of cyber insurance policies: how do carriers price cyber risk?" Journal of Cyber Security 5, 1-19

Sullivan, J., and J. R. C. Nurse, 2020, "Cyber security incentives and the role of cyber insurance," Royal United Services Institute (RUSI) Emerging Insights series, https://bit.ly/3sR1Fg9

Trang, M., 2017, "Compulsory corporate cyber-liability insurance: outsourcing data privacy regulation to prevent and mitigate data breaches," Minnesota Journal of Law, Science and Technology 18: 389-425

Woods, D. W., 2022, "The evolutionary promise of cyber insurance," The FinRegBlog (Duke University School of Law), https://bit.ly/3vLXrs3

Woods, D. W., and R. Böhme, 2021, "How cyber insurance shapes incident response: a mixed methods study," 20th Workshop on the Economics of Information Security (WEIS 2021), https://bit.ly/35Du54b

Woods, D. W., and T. Moore, 2020, "Does insurance have a future in governing cybersecurity?" IEEE Security and Privacy Magazine 18:1, 21-27

Xie, X., C. Lee, and M. Eling, 2020, "Cyber insurance offering and performance: an analysis of the U.S. cyber insurance market," The Geneva Papers on Risk and Insurance – Issues and Practice 45, 690-736

Xu, M., and A. Lei Hua, 2017, "Cybersecurity insurance: modeling and pricing," Society of Actuaries, https://bit.ly/3Cn5VHh

This document was produced for information purposes only and is for the exclusive use of the recipient.

This publication has been prepared for general guidance purposes, and is indicative and subject to change. It does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (whether express or implied) is given as to the accuracy or completeness of the information contained in this publication and The Capital Markets Company BVBA and its affiliated companies globally (collectively "Capco") does not, to the extent permissible by law, assume any liability or duty of care for any consequences of the acts or omissions of those relying on information contained in this publication, or for any decision taken based upon it.

ABOUT CAPCO

Capco, a Wipro company, is a global technology and management consultancy specializing in driving digital transformation in the financial services industry. With a growing client portfolio comprising of over 100 global organizations, Capco operates at the intersection of business and technology by combining innovative thinking with unrivalled industry knowledge to deliver end-to-end data-driven solutions and fast-track digital initiatives for banking and payments, capital markets, wealth and asset management, insurance, and the energy sector. Capco's cutting-edge ingenuity is brought to life through its Innovation Labs and award-winning Be Yourself At Work culture and diverse talent.

To learn more, visit www.capco.com or follow us on Twitter, Facebook, YouTube, LinkedIn Instagram, and Xing.

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Gurgaon
Hong Kong
Kuala Lumpur
Mumbai
Pune
Singapore

EUROPE
Berlin
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Munich
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Hartford
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

WWW.CAPCO.COM

CAPCO
a wipro company