THE SYSTEMIC RISK OF CONSOLIDATION IN THE CLOUD COMPUTING INDUSTRY

A DISSERTATION IN
Economics
and
Entrepreneurship and Innovation

Presented to the Faculty of the University of
Missouri-Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
ANDREW D. COTTON

B.A., Truman State University, 2010
M.A., University of Missouri-Kansas City, 2016

Kansas City, Missouri
2021

THE SYSTEMIC RISK OF CONSOLIDATION IN THE CLOUD COMPUTING INDUSTRY

Andrew Douglas Cotton, Candidate for Doctor of Philosophy Degree

University of Missouri-Kansas City, 2021

ABSTRACT

The purpose of this study is to examine the effects of consolidation within the cloud computing industry related to the reliability and availability of computing resources. This dissertation begins by assessing the scale and scope of the cloud computing industry leader, Amazon Web Services. Included in this assessment are a collection of case studies that reveal some of the unique transactions between actors in this industry. The next section uses a bowtie analysis to frame for discussion the key risks related to cloud computing. This framework is used to analyze how the economic risks of compromise and unavailability have changed with a shift from on premise computing to cloud computing. A normative systems analysis examines the policy considerations for addressing the consolidation in the cloud computing industry, and the social fabric matrix is applied to discuss the unique deliveries among processing institutions and between processing institutions and authorizing institutions. On the basis of the normative systems analysis, several policy implications are examined, including the extent to which government spending reinforces consolidation of power and risk within the cloud computing industry.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies have examined a dissertation titled "The Systemic Risk of Consolidation in the Cloud Computing Industry," presented by Andrew D. Cotton, a candidate for the Doctor of Philosophy degree and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

James Sturgeon, Ph.D., Committee Chair
Department of Economics

Scott Fullwiler, Ph.D.
Department of Economics

Linwood Tauheed, Ph.D.
Department of Economics

Brian Anderson, Ph.D.
Department of Global Entrepreneurship and Innovation

Sejun Song, Ph.D.
School of Computing and Engineering

CONTENTS

TABLES

Table Page

ILLUSTRATIONS

ABBREVIATIONS

Abbreviation	Abbreviated
ACH	Automated Clearing House
ATO	Authority to Operate
AWS	Amazon Web Services
C2S	Commercial Cloud Services
CDN	Content Delivery Network
CIA	Central Intelligence Agency
CISA	Cybersecurity and Infrastructure Security Agency
COPPA	Children's Online Privacy Protection Act/Rule
CVE	Common Vulnerabilities and Exposures
DDOS	Distributed Denial of Service Attack
DHS	Department of Homeland Security
FAR	Federal acquisition Regulations
FCA	Financial Conduct Authority
FCC	Federal Communications Commission
FedRAMP	Federal Risk and Authorization Management Program
FICA	Federal Insurance Contributions Act
FTC	Federal Trade Commission
FUTA	Federal Unemployment Tax Act
GETS	Government Emergency Telecommunications Service
IaaS	Infrastructure as a Service
IOT	Internet of Things
ISP	Internet Service Provider
IT	Information Technology
IT SCC	Information Technology Sector Coordinating Council
IT SSP	Information Technology Sector-Specific Plan
JEDI	Joint Enterprise Defense Infrastructure
NB	Normative Belief
NGN-PS	Next Generation Network Priority Services
NIPP	National Infrastructure Protection Plan
NSAIM	Normative Systems Analysis of Instituted Processes Method
OFPP	Office of Federal Procurement Policy
PaaS	Platform as a Service
SaaS	Software as a Service
S3	Amazon Simple Storage Service
SEC	Securities and Exchange Commission

SFM	Social Fabric Matrix
SFM-A	Social Fabric Matrix Approach
SLA	Service Level Agreement
SRECON	Site Reliability Engineers Conference
UPS	Uninterruptable Power Supply
WPS	Wireless Priority Services

ACKNOWLEDGEMENTS

I would like to thank my professors and committee members for inspiring me to pursue my doctorate, for guiding me to this topic and for seeing me through to completion.

I would like to thank my employer for funding this degree program.

I would like to thank my supervisors, colleagues, and friends who have had to accommodate my studies.

I would like to thank my classmates who helped me study for my classes and exams.

I would like to thank my parents for believing in me and encouraging me along the way.

Finally, and most significantly, I would like to thank my wife, whose faithful support and sacrifices have made this accomplishment possible.

CHAPTER I

INTRODUCTION, PROBLEM STATEMENT, AND BACKGROUND

In 1961 John McCarthy, speaking at the MIT Centennial, posited that computing might someday become a "public utility" (Garfinkel, 1999). Cloud-based computing services are evidence that such a future is present. Firms and organizations across industries have come to depend on cloud-based software to carry out their business. Consistent with other utilities, the new cloud computing industry rewards both size and incumbency. Amazon Web Services (AWS) is a sizeable, early competitor in cloud computing and has come to account for over a third of all cloud computing revenue (Synergy Research Group, 2018). Gartner (2018) predicts that this segment of computing services will grow by over 20% in 2018. As new firms and industries embed cloud computing into more and more of their processes, the broader economy becomes increasing dependent on the cloud computing infrastructure hosted by fewer and fewer companies. A service as complex as cloud computing is prone to accidents and failure. This tendency is apparent in the history of complex systems and has been made manifest in the short history of AWS. This study will examine how a failure of AWS would affect the network of firms and organizations that depend on it, and how this failure may spillover into the broader economy. It will examine the potential causes of such a failure, the impact of such a failure, and evaluate the tools policy-makers have to reduce the likelihood and mitigate the impacts of this critical new infrastructure becoming unavailable.

While not necessarily the first-mover in the cloud computing space, Amazon.com, Inc. was a very early provider of cloud services on a large scale. In July of 2002, Amazon.com launched Amazon Web Services, Inc, as a subsidiary of the online retail company. This subsidiary now provides cloud-computing resources to companies, organizations, governments, and individuals, world-wide. These services include computing, data storage, content hosting, DNS resolution and a variety of other services. The breadth of AWS services and customers is wide and continues to grow.

Cloud computing has been widely adopted for a variety of applications in a variety of organizations, both public and private. These organizations are incentivized by reliability, simplicity, flexibility, and cost savings. These incentives reinforce the economies of scale enjoyed by AWS. Cloud services can provide reliable, geographically diverse backups for data storage and processing, even to the extent that they can replace a company's on premise data center in its entirety. While some companies are using cloud services to free them selves from the overhead, complexity, and burden of an on premise data center, other firms have grown without the need to stand up on premise servers in the first place. Even, and especially, information technology firms are increasingly able to setup to provide services without owning or controlling any physical information technology infrastructure, apart from a small number of personal computing devices. These cloud native companies will be considered in depth later in the case studies. By eliminating the need for or augmenting local data centers, users achieve what Ambrust *et al.* (2010. p. 51) describes as "infinite computing resources available on demand." This on-demand scalability reduces up-front investment by allowing companies to start small and scale only as needed to meet growth or surges in resource demand. Effectively, this

drastically reduces the fixed cost component of computing, substituting a variable cost in its place.

Before cloud service providers, companies subject to surges in demand for computing resources would either have had to invest in infrastructure early or ration those resources when they are needed most. Those companies which could commit major capital investment, up-front, would build infrastructure commensurate with their peaks, and then leave a large portion of those resources idle between those peaks.

Cloud providers, by diversifying their client base, are able to reassign unused computing resources to other users, effectively smoothing much of the peaks and troughs and reducing waste. By so doing, these cloud providers reap significant economies of scale and can pass a portion of these savings to their customers. Startups and small businesses benefit the most, because they face lower up-front capital requirements, achieve massive scalability, and receive the benefit of the same security and reliability received by the most demanding of AWS's customers.

The aforementioned economies of scale made possible by sharing cloud infrastructure enable Amazon to invest in processes that so far have provided a very high level of reliability and met the security requirements of their customers. Amazon has implemented geographical redundancy and invested in a variety of other features that ensure the availability of their services to their customers. Many of these features are included in the fees that AWS charges, but heightened provisions are available for an additional price, or through third-party vendors. Additionally, some features discussed in Chapter IV are created by AWS customers and provided for free to other customers of AWS.

Amazon's current security practices have evolved from those practiced in its own data centers. Now these practices have been adapted to address the complications of co-tenancy in a shared infrastructure. Amazon and other cloud providers must continually anticipate the next threat. In such an environment, the question is not if Amazon will be compromised, but when, to what degree, and with what impact.

The failure of AWS to deliver expected services can result from a variety of root causes. For the purposes of this paper, Amazon Web Services fails if it can no longer provide the services expected by its customers and it becomes nonuseful. This study will focus on what happens if Amazon experiences a technical failure. The following types of technical failure will be addressed in this paper:

1. Software failure
2. Equipment failure
3. Infrastructure or facility failure
4. Malicious actor usurps service
5. Malicious actor disrupts service
6. Publicized compromise without disruption
7. Data Loss

Software failure occurs when the applications that comprise AWS cloud offerings experience a logical failure in which the code does not perform as expected. This would also include a failure of the virtual infrastructure which will be explained in greater detail in Chapter IV. Such a failure may be unintentional or may be caused by undetected malware deployed with the intent of causing an outage. Malware is a broad term to describe software intentionally designed to cause damage to a computer or a computer

network. Malware is often infectious, replicating itself and spreading among information systems to disable them. A variety of malware known as ransomware captures private data and holds it for ransom or blackmail purposes. A distributed denial of service attack (DDOS) is an attempt to bring down an online system by overwhelming it with traffic. In a DDOS, malware may be employed to take control of vulnerable information systems and use them to overwhelm another system. Equipment failure refers to the physical network and computing devices that comprise the AWS cloud. This could be caused by defective or aged equipment or could be the result of sabotage through a supplier or installer. Infrastructure or facility failure captures physical damage to the location of an AWS data center or on the power or internet connectivity on which such a data center depends. Infrastructure or facility failure would most likely be the result of a natural disaster, but could also refer to an intentional attack on power, network, or the buildings themselves.

The next four causes are the result of a malicious compromise of the AWS network or physical location. This could be caused by an insider abusing their access, a hacker gaining access, or either party gaining physical access to an AWS location and compromising the integrity, availability or confidentiality of AWS-hosted data and processes. As a result of such a compromise, AWS fails to deliver expected services. If a malicious actor usurps a service, it can use it to gain sensitive information or change processes to destructive or fraudulent ends. A malicious actor may simply disrupt a service and cause an outage. A malicious actor may also simply publicize their compromise of an AWS system without causing a disruption. This could be an act of terror or a publicity stunt. Finally, data loss occurs when unauthorized users access

customer data. In the context of this paper, unauthorized users could be internal to the customer's organization, internal to AWS, or entirely external to both organizations.

Researchers are using a few technical tools that would enable an evaluation of the footprint of AWS across a variety of websites by blocking packets from AWS IP addresses. This provides a picture of the internet sans-AWS. A recent paper on DNS concentration has leveraged the Herfindahl-Hirschman Index (Bates *et al*., 2018) to assess market concentration. Through an assessment of the most relevant literature, this paper aims to provide a clearer understanding of the tools already employed for this type of research.

Policy-makers have many options for mitigating the risks posed by the prominence of AWS. The U.S. Government could determine that AWS is a natural monopoly and regulate it like a public utility. Short of this, regulators might consider blocking mergers and acquisitions that would further consolidate this industry. The broader discussion of monopoly power, especially as it relates to Amazon's retail arm, is largely beyond the scope of this paper. Consumer information protection laws will be examined to determine their relevance to these latest technological advancements in the cloud computing industry. Additionally, public records and other records of strategic importance may be most efficiently hosted in the cloud, but this paper will highlight the need to assess what controls are and could be put in place to accomplish this. Critical computing resources should have sufficient redundancy, perhaps utilizing a secondary or tertiary cloud provider. Cloud hosting is not a like-for-like replacement of a government hosted data center. Countries will need to work together to ensure that international laws deter the development of malware and data theft. If one company comes to dominate key

infrastructure in the global information industry, the question arises whether there are sufficient protections to prevent that firm from destabilizing governments or international economies. It is useful to assess whether there are sufficient internal controls within that company and whether they are truly verifiable.

This dissertation begins by assessing the scale and scope of the cloud computing industry leader, Amazon Web Services. Included in this assessment are a collection of case studies that reveal some of the unique transactions between actors in this industry. The next section uses a bowtie analysis to frame for discussion the key risks related to cloud computing. This framework is used to analyze how the economic risks of compromise and unavailability have changed with a shift from on premise computing to cloud computing. A normative systems analysis examines the policy considerations for addressing the consolidation in the cloud computing industry, and the social fabric matrix is applied to discuss the unique deliveries among processing institutions and between processing institutions and authorizing institutions. On the basis of the normative systems analysis, several policy implications are examined, including the extent to which government spending reinforces consolidation of power and risk within the cloud computing industry.

CHAPTER II

LITERATURE REVIEW

What is the systemic impact of Amazon Web Services becoming unuseful? The existent literature related to this question only touches it tangentially. Because there is not a single body of research that is systematically specific to this question, an interdisciplinary approach is used here to position the present study within the relevant streams of research. This section opens with a public-sector perspective of risk as it relates to the information technology sector. The next stream of closely related research is in the applied information technology literature and deals with system outages and information security concerns. The final section begins with a strategic management and microeconomic perspective that addresses the appeal of cloud computing, the tendency toward consolidation, and then highlights risks introduced by consolidation, including issues of monopoly power and a potential lack of legal framework and precedence in this industry.

Cloud computing introduces new risks to the critical infrastructure of the global and national economy. The United States Department of Homeland Security (DHS) publishes the National Infrastructure Protection Plan (NIPP), which is divided into sector-specific plans. The Information Technology Sector-Specific Plan (U.S. Department of Homeland Security, 2016), here-after noted as the IT SSP, was most recently published in 2016 and while mentioning cloud computing, it does not offer a detailed treatment of this topic. It identifies cloud topics as an area for future risk assessments, including an analysis of consumer risk. While this assessment is lacking a detailed cloud treatment,

this sector-specific plan provides a framework by which the Department of Homeland security assesses risk in this sector.

A major theme underlying the IT SSP is the degree to which the other sectors of critical infrastructure rely on the IT sector. This degree of dependence is increasing with the advent of the internet of things (IOT) or the phenomenon of analog devices utilizing the internet to return feedback from sensors. As physical processes increasingly rely on network connectivity and even processing power from the cloud, there is an ever-increasing risk of IT incidents spilling over into other sectors. The IT SSP acknowledges the possibility of these kinds of cascading effects but without demonstrating the specific coupling between IOT and Cloud computing.

The National Infrastructure Protection Plan describes a sector partnership model designed to stimulate collaboration between the private and public sectors. Private sector firms may join the IT Sector Coordinating Council (IT SCC) which provides a forum to discuss sector protection issues. In 2016, both Microsoft and Google were members of the council, but Amazon was not. As of October 16, 2018, neither Google nor Amazon were members (IT Sector Coordinating Council, n.d.).

In the IT SSP, DHS shares its vision for the IT sector: "To achieve a sustained reduction in the impact of incidents on the sector's critical functions" (U.S. Department of Homeland Security, 2016, p. 9). This perspective provides an approach to IT sector risk, which will be adopted in the following chapters—specifically that incidents are inevitable and that at the system level the primary concern is to limit their impact when they do occur.

Charles Perrow addresses systemic risk in several of his works. In *Normal Accidents*, Perrow (2011) provides a framework for understanding system risk introduced by complexity and tight coupling. A case is made that in the presence of both complexity and tight coupling, accidents are inevitable and catastrophes are only a matter of time. Perrow (2007) relates increased scale and consolidation to increased risk in the context of complex systems, devoting a chapter to information technology systems. Perrow (2008) introduces modularization as a means of mitigating the risks of consolidation. Perrow does not address cloud computing directly, but the concepts of complexity, coupling, scale, and consolidation can be readily applied to cloud computing infrastructure and the processes dependent thereon.

Zheng *et al.* (2013) provides a systematic survey of public cloud outages. They provide an education framework for learning from cloud service outages. They offer four high level sample lessons from the outages researched. The first lesson is that uninterruptible power supplies (UPS) are more interruptible than most data centers anticipate, due to a lack of testing of the system. The second lesson is to be pessimistic about every component and prepare for failure. The third is to avoid chain reaction situations in which an outage in one component is able to spill over into others. Finally, another lesson is to share outage details with customers and others in the industry so that new best practices can be developed. This article highlights the myriad of causes behind cloud outages and provides some guidance for prevention and mitigation.

Van Eeten and Bauer (2012) address how poorly-aligned incentives create the potential for internet-generated mega-crises. They describe how botnet herders use computers infected with malware to overwhelm internet infrastructure. In the event of a

botnet attack, the owners of the infected machines are often not the targets of the attacks and have little incentive to protect their machines from being used in this way. The speed with which cloud based computers can be deployed make them an especially desirable target to bot herders.

Yanpei Chen, Vern Paxson, and Randy H. Katz (2010) explain that many predators would be willing to pay a premium to leverage cloud computing to host malware. While the cloud is more expensive than simply establishing a botnet using the computers of unsuspecting people and businesses, the superior computing and scalability may make it worthwhile for well-funded criminals. Their overall view is that cloud computing is primarily a new twist on the same old security issues. They mention the concentration of security expertise in cloud providers, but also note that sharing of infrastructure can also intertwine the fates of unrelated companies. They relate an incident in 2009 where FBI agents seized the physical assets of a data center because one of the shared tenants had engaged in illegal activity. Other tenants had their businesses interrupted and in some cases temporarily closed as a result. Cloud redundancy may make those types of seizures less damaging, but it does highlight how sharing infrastructure intertwines the fates of unrelated organizations. It also illustrates the difficulties faces by law enforcement when trying to obtain evidence from complex technical systems.

Haimes *et al.* (2014) assesses the risks facing cloud computing technology from the perspective of a complex interconnected system of systems (Yacov Haimes, Barry Horrowitz, Zhenyu Guo, Eva Andrijcic, and Joshua Bogdanor, 2014). This article determines that it is riskier for companies to use the cloud than to leverage their own on

premise data centers, but this conclusion hinges on two main assumptions. First, this article assumes that cloud providers utilize the same security standards employed by owners of on premise data centers. Second, they assume that cloud providers face the same type of intrusion or threat. This paper will later propose that neither assumption is valid for such an analysis.

In 2017, Amazon made headlines for an outage of their S3 storage service. Weise (2017) chronicles that event, while placing that occurrence within the context of other AWS outages. The article goes on to explain that firms can minimize the business process impact of such an outage by adopting a multi-provider cloud strategy. Weise includes statements from two analysts from the technology consulting firms, Forrester and Gartner. The Forrester analyst explains that most firms are hesitant to invest in cloud redundancy given its cost. Gartner's analyst takes it a step further suggesting that only very large, paranoid companies utilize a multi-provider strategy. This demonstrates that most companies prefer the risk of cloud outage to the high cost of prevention.

Cancila *et al.* (2016) highlights the business trends in cloud technology. They explain the Cloud-First mindset, which began in the Federal government and asks software developers to give cloud solutions primacy in an attempt to minimize IT costs on premise. It has since become a buzzword in application development shops and in the literature. This paper also recommends a multi-provider strategy by contracting with multiple cloud providers and/or a cloud broker. The authors warn of the risks of lock-in and unavailability. Lock-in occurs when software applications are designed around the infrastructure offered by a specific cloud provider and are not easily transferred to a competing cloud provider. Unavailability occurs when that specific cloud provider is

unable to deliver the services on which the business depends. These risks appear in the literature and in the recommendations of technical consultants but also appear to be generally ignored by practitioners.

Several studies examine the adoption of cloud technologies. Ambrust, Fox, Griffeth, Joseph, Katz, Konwinski, and Zaharia (2010) tout the efficiencies of scale of cloud computing and cite the ease and speed of scaling applications in the cloud. They make a compelling case for the opportunities of the cloud, especially for startups. West (2010) examines the cost savings for municipal users of cloud technology and roughly estimates savings to be between 25 and 50 percent. Tak, Urgaonkar, & Sivasubramanian (2011) perform an NPV cost analysis of running different types of applications in the cloud. They show that some applications will experience cost savings in the cloud while others may not. Workload intensity, growth rate, storage capacity, and software licensing are the key determinants of whether or not cost savings will take place. Adekunle *et al.* (2012) highlights the advantages of cloud computing and demonstrates how even small companies with limited budgets now have access to the same quality of service from the cloud as much larger firms were able to achieve form their on premise data center solutions or from the cloud. The International Data Corporation conducted their 2013 Cloud Security Survey (IDC 2013) and found that nearly 60% of organizations agreed that cloud service providers provide better security than their own IT organizations. This body of literature makes a compelling case for the advantages of cloud computing for individual companies and why they are incentivized to adopt cloud computing.

Gartner (April 25, 2018) offers a guide for designing a public cloud exit strategy. This guide observes that most organizations are adopting cloud services but note that few

consider an exit strategy. They go on to explain that this is because such provisions are expensive, time-sensitive, and operationally and technically challenging. The lack of exit-strategy represents either an oversight, an assumption of the clouds resiliency, or the believed infeasibility of leaving the cloud. This would be especially true of firms that no longer (or never did) have the on premise infrastructure and personnel to run it. This further demonstrates that firms are over-reliant on their single cloud providers and would be ill-prepared to adjust their strategy in the event that AWS becomes unuseful.

Karen Petrou (2018) likens the mounting consolidation of risk in internet infrastructure to the consolidation of risk in large Banks in 2008. Large technology companies are increasingly interwoven into our financial system and are often not regulated accordingly. This article was popularized by Penny Crosman in *American Banker* (2018). Petrou highlights the lack of inter-operability between cloud providers as a shortcoming of this unregulated public utility.

Bates *et al.* (2018) examine market consolidation in the Domain Name System (DNS) in "Evidence of Decreasing Internet Entropy: The Lack of Redundancy in DNS Resolution by Major Websites and Services". DNS resolution is the service that associates the characters of a web address with the servers that host the site. If the DNS server for a site is down, that site cannot be reached. This study shows that despite the relatively minor cost of registering a website with multiple DNS providers, many companies only have one. This is a single point of failure that can cost the company a lot in the case of an outage. They found that this was especially problematic for AWS users, of whom 87% were undiversified in their DNS providers. The authors also note how centralization in internet infrastructure is likely to invite attack. This article supports the

idea that firms are either ignorant of this risk, or will not accept even minor costs in order to ensure against outlier infrastructure events. If the consolidation in the cloud computing services is comparable to what is here shown for DNS resolution, there is reason to suspect that firms are introducing avoidable systemic risk.

Any research on economies of scale will begin to overlap with the literature on monopoly power. Within this literature, Amazon has been the subject of several studies. In the 2018 Jackson Hole Symposium, hosted by the Federal Reserve Bank of Kansas City, the theme was Changing Market Structures and Implications for Monetary Policy. Changing market structures here referred primarily to the scale of large corporations, especially in technology. Five of the six sessions that comprised this conference addressed Amazon specifically, but none addressed the role of Amazon as a cloud services provider (John Van Reenen, 2018, Nicolas Crouzet, 2018, Antoinette Schoar, 2018, Yuriy Gorodnichenko, 2018, & Stephen S. Poloz, 2018). AWS made up 9.82% of Amazon's revenue in 2017 and 105.48% of its total profit, effectively subsidizing the retail operations (Amazon.com, Inc., 2017a). Van Reenen addresses how Amazon's scale enabled the development of superior proprietary software in the context of its retail operations, but fails to mention that this proprietary software became the basis of Amazon Web Services and is now provided as a paid service to the public.

The legal environment for cloud computing is still being defined. Tina Cheng (2013) outlines the laws most relevant to the cloud environment. This article highlights the obsolescence of consumer privacy laws that were developed in the mid-80's and their poor application in the context of current technology. William Denny (2010) highlights the legal issues of privacy, jurisdictional confusion, and contributory liability for

copyright infringement in the context of cloud computing. Each issue represents situations in which the law fails to anticipate its application to cloud computing and results in legal decisions that differ from the intent of the law. Later sections of this paper will identify further gaps in the legal environment of cloud computing.

In summation the literature on risk has not kept pace with the latest advancements in cloud computing. Cloud outages are relatively common, but the incentives for organizations to think "cloud first" are sufficient for them to ignore these outages and the associated risks. While expert consultants advise multi-provider strategies, competitive firms ignore this advice in favor of saving money. Industry research firms document this phenomenon. Other internet infrastructure components have faced widespread outages due to the targeting of major providers. The IT SSP has not yet addressed the risk introduced by cloud computing and Amazon is not at the table for these risk discussions. The question is relevant: What is the systemic impact of AWS becoming nonuseful?

CHAPTER III

SCALE AND SCOPE OF AWS

This chapter will establish the scale and scope of AWS. First it will demonstrate how the cost structure of cloud infrastructure rewards large-scale firms and make a case that they tend toward natural monopolies. Next will be a brief examination of the extent of Amazon's footprint in the cloud computing industry, including an evaluation of Amazon's market share. After establishing the rough scale of AWS in the cloud computing industry, the focus will shift to scope. Scope will be assessed by considering representative case studies of organizations from a variety of industries and sectors that have become dependent on AWS for a portion of their operations. These case studies will also shed light on the degree to which these organizations are dependent on AWS and the precautions available to and taken by these companies to control the risk. The concept of lock-in will be demonstrated in many of these case studies. These case studies will elucidate the uniqueness of the cloud computing industry and the unique nature of cooperative competition that appears prevalent. By establishing the scale and scope of AWS, and examining the dependency of these sample organizations, this chapter will provide a basis for understanding the risk AWS poses to the broader economy. This will set the stage for Chapter IV in which are explored the consequences of Amazon becoming unuseful in the event of data loss or a failure to deliver expected services.

Cost Structures of the Cloud

The scale and scope of AWS can reasonably be attributed in part to economies of scale. For Adam Smith, the division of labor increased what he described as the

productive powers of labor. (Smith, 2013). For Smith, scale primarily allowed specialization. In the case of modern computing, the scale of computing in general has reached a point that has allowed for a significant division of labor to arise in providing computing services. This division of labor extends beyond the specialization of individuals to the specialization of entire organizations whose entire focus is in bringing specific cloud computing services to customers in other industries. Entire companies may specialize in one specific type of computing service and make that service available on a large scale. AWS has specialized in data storage, computing, and web-hosting as well as the networking and infrastructure which underlie and support each service. Few firms have internal information technology teams that can compete with the specialized skills, experience, and expertise of AWS.

Independent of specialization, microeconomic analysis analyzes economies of scale in terms of fixed and variable costs. In microeconomics, economies of scale are defined algebraically as when the marginal cost of producing a good or service is less than its average cost. Theoretically, the primary cause for economies of scale is an upfront investment in fixed costs that yields savings which are not offset by costs that increase with scale. For cloud computing service providers, there is a large fixed cost component comprised of land, buildings, equipment and software. Other sources of efficiencies that contribute to economies of scale are: bulk-purchasing and bargaining power; financial economies of scale resulting in lower interest or financing charges; relatively fixed marketing costs that are spread over an increasing number of sales; and the potential for economies of scale in the underlying technology. Another source of economies of scale are network effects, where the selection of a standard provides

benefits based on the number of other users who also adopt that standard. For AWS, the economies of scale are related to the underlying technology and the benefits gained by network effects. In order to understand how technology influences the scale of AWS, the next section will examine the design of AWS infrastructure and the technologies at work

The Technologies and Design of AWS Infrastructure

The primary functional unit of input for cloud computing is the server. Servers are simply computers that are housed together in a data center. These data centers can be replicated relatively easily without incurring increasing costs per data center. They can also be rented or contracted through other firms. For AWS, one or more data centers make up an availability zone and two or more availability zones make up a region. As of March 2019, AWS has 61 availability zones and 20 geographic regions around the world (Amazon.com, Inc., n.d.-a). The technology underlying this business model is scalable at every level in this hierarchy. Each server of a rack in a datacenter can be scaled for increased compute or storage within the confines of the current technology. Each rack can be expanded to hold more servers; each room, more racks; each data center, more rooms; each availability zone, more data centers; and each region, more availability zones. Due to the automated nature of server scaling, the burden of complexity introduced by scale is borne primarily by automated process and therefore does not significantly add to costs.

Another factor which plays into the advantage of large cloud providers is the network effect. A single data center owned by Amazon is more resilient because if it fails, it is backed up by other data centers within the availability zone and even data centers in other regions. In this way, additional data centers contribute both to overall

computing power, but also to the number of failover options and disaster recovery scenarios for the system as a whole. While costs do not increase with scale due to this network effect, the marginal benefits of adding an additional server, room, data center, availability zone, or region do decrease with scale. This is similar to the way that the value of a diversifying a financial portfolio decreases with each unique investment that is added (Evans & Archer, 1968).

Furthermore, in the same way that stock portfolios benefit from diversification of stock holdings, cloud providers diversify the demands on their system by attracting a wide variety of customers from a wide variety of industries. For example, a single firm may have processes that impose spikes on their computing systems. For example a financial accounting system may have much more data to process at the end of the month, quarter, or fiscal year, than they do day-to-day. In cloud hosted environments, however, these peaks may happen to align with troughs from other companies in other industries and the effect is a smoother, more predictable, and therefore more efficient demand on the systems.

Network effects also appear from a customer perspective. Customers benefit by using standard cloud services. By so doing, they have access to a wider variety of services that are created for the dominant platform. Additionally, when cloud customer hire staff to work with or maintain these services, they are more likely to find candidates with experience on the most dominant platform.

So far this section has established that the division of labor between firms has allowed AWS and other cloud providers to specialize in the provision of services through the cloud to customers who cannot easily replicate this expertise. These cloud providers

are likely to experience economies of scale in the algebraic microeconomic sense due to large up-front fixed costs that are spread over an increasing number of units of output. The network effect provides additional benefit to cloud providers that expand their business; diversification contributes to the efficiency and predictability of resource use; finally customers benefit from a single, dominant standard platform in the cloud computing industry.

Cost Structures for Amazon

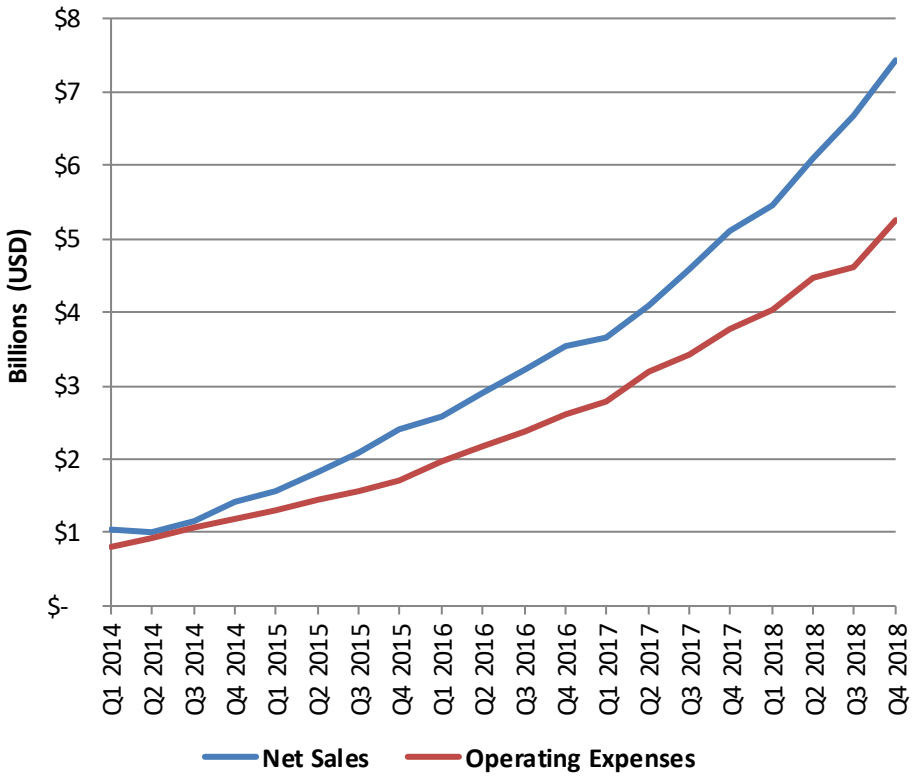
Having established the theoretical basis to expect economies of scale in cloud computing, and in particular for AWS, the next step is to demonstrate that this theory is not in contradiction to the stylized facts. This section will show that AWS revenues and costs are consistent with what would be expected by a firm currently benefitting from economies of scale. The most applicable publicly available and sufficiently granular data comes from quarterly press releases provided by Amazon.com, Inc. to report quarterly earnings. This financial data is broken out by Amazon.com, Inc.'s three business lines, United States Retail, International Retail, and Amazon Web Services. From this data, we can estimate whether average costs appear to decrease with increasing revenues.

As its name implies, the output of AWS is services. Services can be difficult to quantify, but a close proxy for output of services is the dollar-volume of sales. In order for sales to be a useful proxy for output in the assessment of economies of scale, the unit-price of those services must be constant or decreasing. If the price of services were increasing, then increasing revenue with sales could be due to either the increased price or costs that decreased with sales. As of 24 September, 2018, AWS has reportedly decreased its prices 67 times since its inception (Rallo, 2018). This is consistent with the

findings of Zhang (2016), demonstrating an annual decrease in price of between 7.33% and 8.10% for comparable AWS services between 2009 and 2015. This decrease in price over time also aligns with a near constant increase in sales volume per quarter. Figure 3.1 shows the decline in average cost per dollar sold as sales volume increases. This negative correlation is statistically significant at the 95% confidence level. The elasticity of cost to sales is .9222. This means that a 100% increase in sales was associated with a 92.22% increase in costs over this period. This elasticity is visually represented in figure 3.2. This evidence suggests that as AWS scales up, it faces decreasing average costs even in the face of decreasing prices per unit of service.

Figure 3.1

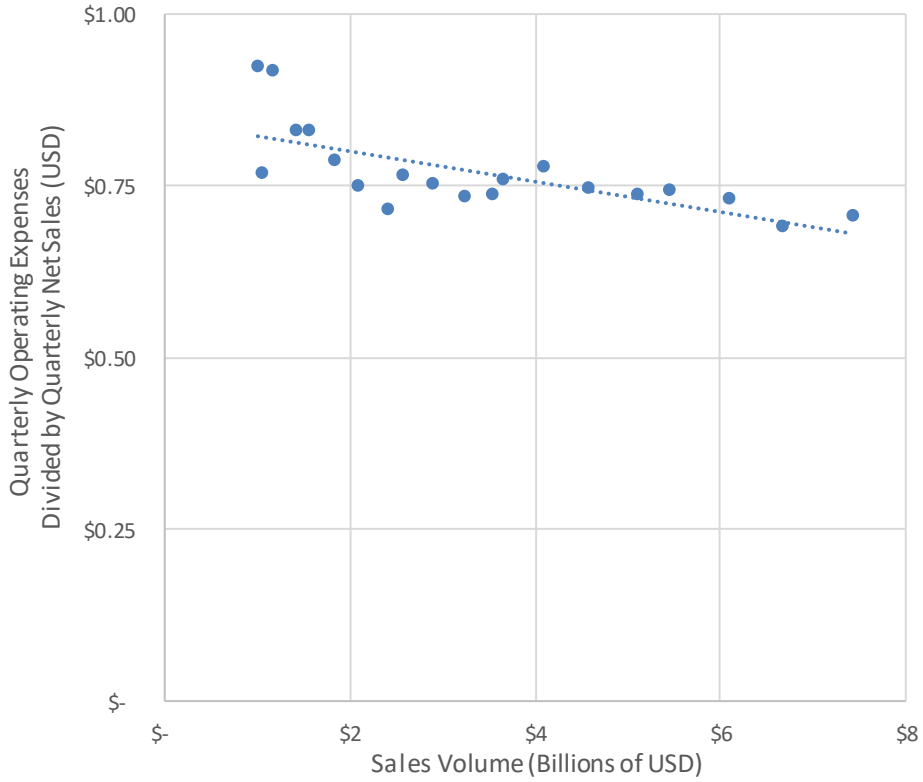
AWS Net Sales and Operating Expenses from Q1 2014 to Q2 2018



Source: Compiled from data in Quarterly Earnings Press Releases. Retrieved from https://press.aboutamazon.com/press-releases on March 25, 2019.

Figure 3.2

The Average Cost Per Dollar Sold by AWS (2014-2018, by quarter)



Source: Calculated by author, from Quarterly Earnings Press Releases. Data Retrieved from https://press.aboutamazon.com/press-releases on March 25, 2019.

Entangled in this analysis is the concept of technological efficiency and whether or not the decrease in AWS prices over time is more reflective of economies of the scale achieved during that period or the improved technologies that became available during that period. The question of whether or not AWS is passing along any of the rents gained by both technological efficiency and scale is beyond the scope of this work.

Having shown that AWS appears to benefit from economies of scale, the next question is whether or not this would inevitably lead to a monopoly. According to Posner,

(1978), a natural monopoly occurs when the entire demand of a relevant market can be met at lowest cost by one firm rather than two or more. By that definition, the cloud computing industry is likely a natural monopoly, but those conditions are not sufficient to assume that AWS will become a monopoly. Additionally, differentiation between competitors is significant, which may prevent consolidation in the market. One of the diseconomies of scope which Amazon has encountered is that AWS is sometimes unable to sell its cloud services to firms that its parent company, Amazon.com, Inc., competes with in other markets. For example, retail firms like Kroger, the largest grocery store chain in the United States, are intentionally avoiding doing business AWS. This is reported to be because AWS is owned by one of their top retail competitors (Levy, 2017).

The Scale of Amazon

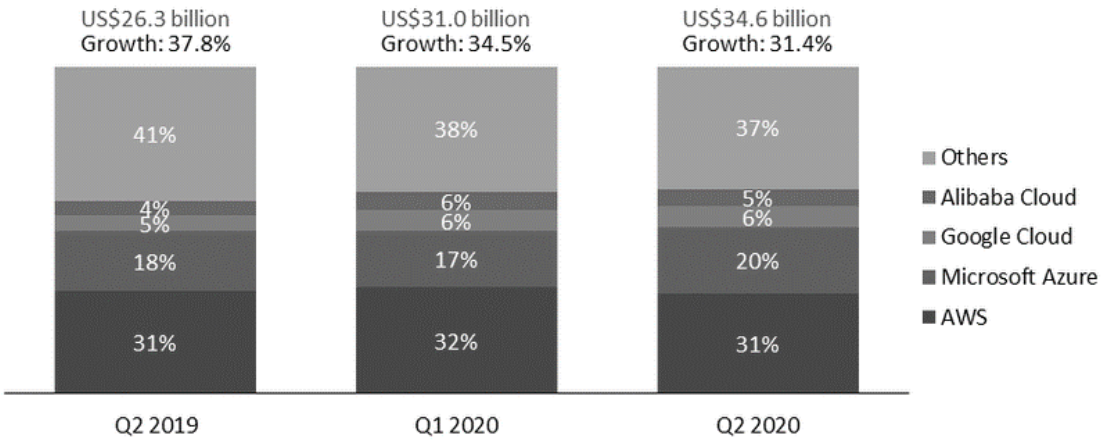
Having established the theoretical justification for the scale of AWS and demonstrated that the cost structures AWS faces are consistent with this theoretical framework, this section will briefly contextualize the discussion of AWS by demonstrating the scale of Amazon in the context of the overall economy as well as the Cloud computing industry in particular. This brief quantification will address the scale of AWS, while the scope of AWS will be addressed in the later section containing the representative case studies.

According to their Q2 2020 earnings report, AWS had $10.808 billion in revenue, yielding $3.357 in profit (Amazon.com, Inc., 2020b). This accounted for 12.1% of total revenue for Amazon.com, Inc., and 57% of its profit. This represents a 29% growth year-over year in revenue and 58% increase in profit. This is difficult to compare with its top competitor, Microsoft, which does not distinguish its cloud revenues or profits from its

other lines of business. Microsoft, for the quarter ending June 30, 2020 reported that their commercial cloud surpassed $50 billion in annual revenue (Microsoft, 2020). Microsoft brands its cloud offerings under the name Azure. The Azure portfolio is portfolio is the segment of the Microsoft business model most comparable to the AWS subsidiary of Amazon.com Inc. It is not clear exactly what portion of Microsoft's $50 billion in commercial cloud revenue can be attributed to Azure, but Microsoft did state that revenue from Azure was 47% higher in this quarter than the same quarter of 2019. Canalys (2020) estimates AWS to comprise 31% of the worldwide cloud infrastructure services spend, followed closely by Microsoft Asure at 20%. Google and Alibaba trail with 6% and 5%, respectively. The top two providers share over half of the market.

Figure 3.3

Worldwide Spend on Cloud Infrastructure



Source: Canalys estimates, July 2020 (Canalys, 2020)

Amazon.com, Inc. employs 876,800 employees, globally (Amazon.com, Inc., 2020c). AWS does not publish its employee count, but a rough order of magnitude can be

ascertained by looking at the number of individuals who self-report working for AWS on LinkedIn. This number is 70,816 (LinkedIn, n.d.-b). This can be compared to the 475,045 self-reporting to be in the employ of the parent company (LinkedIn, n.d.-b).

In 2018, the United States Department of Defense released a final request for proposals for a Joint Enterprise Defense Infrastructure (JEDI) cloud solution (U.S. Department of Defense, 2018). This contract was meant to modernize the information technology infrastructure of the department of defense over the course of 10 years. The body of work represents $10 billion in services over this time (McKinnon, J. D., 2020). The Department of Defense surprised many by awarding the JEDI contract to Microsoft instead of AWS, who was initially considered the front-runner to win the contract. On October 7, 2020, Amazon has appealed to the U.S. Court of Federal Claims who issued an injunction halting work on the contract until a review could be made of the selection process. On July 6, 2021 the Department of Defense canceled the contract entirely and announcing a new contract that could potentially capitalize on interoperability in the cloud (Feiner, 2021). This contract highlights the role of the government as consumer of cloud services, and shows how a single contract can create a 6% swing in market share between competing firms ($1 billion dollars of annual revenue represents approximately 3% of the previously estimated $34.6 billion in global cloud infrastructure spending worldwide in 2020). The result of this contract will considerably impact the competitive dynamics of the cloud computing industry.

This section demonstrates a remaining and significant dominance of AWS in the cloud computing industry. The sheer size of this industry and the position of AWS within it makes AWS a prominent firm in the United States economy. The implications of this

scale are amplified due to the widespread dependence of other firms and industries on AWS. The next section will consider case studies of firms that are dependent on AWS for their computing. The case studies span a variety of industries and demonstrate the ways that specific companies are dependent on AWS and would be harmed in the event of a compromise or failure of AWS.

The Scope of AWS: Representative Case Studies

This section will further establish the dominance of AWS, demonstrating additional network effects, and close coupling with other firms and industries. In order to understand the potential loss in the event that AWS becomes unuseful, this section will establish the scope of users, highlighting the processes which have become dependent on AWS in order to operate. While explaining these dependent business processes, this section will also reveal unique exchanges between customers and AWS, as well as between customers of AWS. This section will begin by examining technology companies providing virtual services and move from virtual outcomes to those related to transportation, manufacturing, financial services, public services and national intelligence.

This section begins with AWS commercial customers that provide digital services or software as a service. These companies are shown to have complete reliance on AWS for profitability as their processes have no physical substitutes which can be carried out apart from the availability of their digital infrastructure. Variations in the service level of AWS are directly translated into variations in the services provided by these companies.

The first type of digital service company considered are those for which AWS provides a critical service to the company's core operations. These are companies that in

the absence of AWS cannot generate sales or provide their core services. Several companies considered rely primarily on subscription revenues. In these instances, it is not entirely clear that an interruption in service will have a direct or immediate impact on revenues in the short run.

Netflix

Netflix is a representative case study that highlights the general relationship between digital service providers and the providers of cloud-based infrastructure. Additionally, the specifics of the relationship between Netflix and AWS demonstrates the unique style of cooperative competition which has become common for customers of AWS. This cooperative competition arises from AWS's dominance in cloud computing occurring in conjunction with Amazon's many ventures in a wide variety of other industries in which its cloud customers also compete. This brief case study examines the core operational processes that make up the value proposition of Netflix. It examines how each process depends on AWS, and the risk management decisions Netflix has made to defend against and mitigate the negative effects of interruptions in AWS services.

Netflix is a producer and provider of streamed digital video content. Netflix provide movies and TV shows to its viewers. The method by which Netflix delivers this content is explained by Adhikari et al. (2012). Since this article, Netflix has completely transitioned all of its datacenters to AWS, excluding their content delivery network servers (CDN's), which are explained below. Outside of that, Netflix's model has not changed significantly, Nair (2017). As described by Nair and Adhikari et al., Netflix produces or secures the rights to digital content, which is stored in AWS (and backed up in the Google Cloud) as a high-quality file that meets cinema standards. This file is

transcoded, which means it is converted into varying file types, each readable by a different playback device. For example, one might playback on a Windows computer using Silverlight, while another will play on an iPad and a third would play on a 4K Television with a Dolby sound system. In addition to these different formats, different qualities of each of these formats are stored in case the streaming speed is not sufficient to watch in higher quality. Finally, these files are broken up into 'chunks' not more than a few seconds in length. Each of these chunks in each of these formats at each of these quality levels are disbursed and stored as close to the customer as possible to maximize response time, using a Content Delivery Network (CDN). Netflix uses third-party CDN's, as well as its own CDN, called Open Connect.

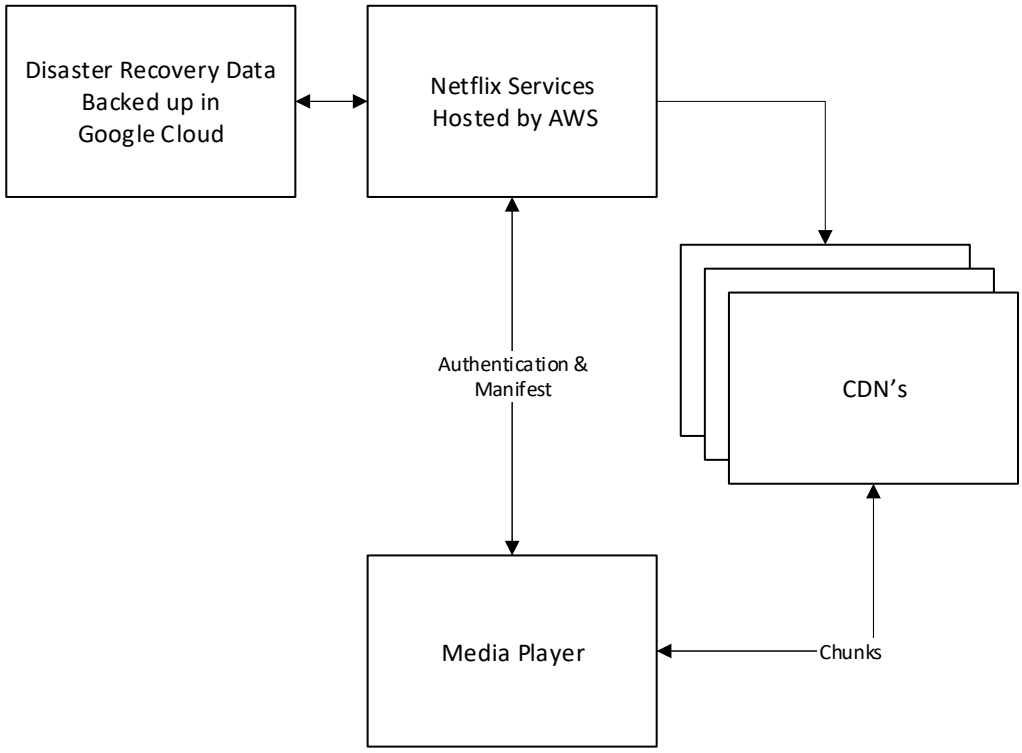
A CDN establishes servers on premise with an internet service provider. CDNs differ from cloud computing in that they serve a narrow need, the storage and provision of content directly to an internet service provider (ISP) to reduce lag-time with the end users. As a rough analogy, a CDN is like a distribution center. A distribution center and a CDN exist to temporarily store items for a speedier delivery to the end-user. Finally, a playback device retrieves the chunks from the CDN and the content can be watched by the customer.

In order to play content, a playback device contacts Netflix services that are hosted by AWS. This is shown in figure 3.4. One service may manage sign-on to authenticate an active, paid subscription. Another service will provide the playback device with a manifest. This manifest tells the device where the chunks of the requested content are stored in the CDN. It also provides the logic by which the player may switch between quality-levels and even CDN locations. Without AWS, account verification and

manifest files cannot be received—so it is unlikely that a user would be able to find and play content; however, once playback has begun it appears to be possible for a complete AWS outage to occur without interrupting content that has already begun to play.

Figure 3.4

Netflix Content Delivery System



Source: Developed by author based on information presented in this work

Netflix reported $4.9 billion in revenue in the second quarter of 2019 (Netflix, 2019). There are 2,208 hours in the second quarter. If revenue were directly derived from content evenly distributed among those hours, that amounts to an estimated $2.2 million in lost value to consumers for every hour of downtime. While the connection between uptime and profitability must be more nuanced than can be represented by such a simple computation, it does provide a rough order of magnitude of value that is at stake. In short, Netflix is incentivized to ensure uninterrupted availability. To this end, Netflix has

invested in three things: multi-region availability and failover, a multi-cloud strategy, and internal system checks.

Netflix pays a premium for multi-region availability and failover from AWS. This is a redundancy technology that was discussed in the previous section. If the data centers in an entire region of AWS (one of 22 globally) goes down, the other regions will split the remaining load. This provides redundancy within AWS and proved valuable on February 27, 2017, when AWS experienced a 4-hour outage in the US-EAST-1 region (Amazon.com, Inc., n.d.-b). Although Netflix had a presence in that region, its redundancy measures with AWS ensured no interruption in content delivery (CBS News, 2017). The second level of redundancy is at least partly supported by Google Cloud according to two sources close to AWS that were quoted in McLaughlin (2018). Google Cloud assists in the business resumption process in the event that AWS is completely compromised. The source goes on to describe Netflix as storing "business-critical data" on the secondary cloud platform. Netflix has not publicly stated whether or not it has redundancy for streaming service in the event of a complete AWS outage. Google cloud could theoretically provide that service, but McLaughlin's sources only imply that the data and configurations crucial to recovering from a disaster are backed up on the cloud. Netflix has made it clear, however, that it intends to continue its extensive partnership with AWS (Krazit, 2018).

Redundancy is only effective to the degree that failover can be achieved seamlessly. Written by former Netflix Cloud Solutions Director, Tseitlin (2013) describes how Netflix reduces uncertainty in its failover procedures, by regularly inducing failure. Netflix has developed a suite of tools that randomly turn off redundant components so

that the company is continuously testing its failover procedures and able to analyze the effectiveness thereof. This suite of tools is dubbed the Simian Army after the original ChaosMonkey tool. This tool is made available to every cloud user as open source software through GitHub (Github, n.d.). The sharing of this tool represents a synergy between AWS and Netflix. Netflix benefits from the use of the tool, and by sharing it, more users are able to ensure proper redundancy of their services on AWS, which makes AWS more useful to its customers.

Despite the synergies that arise between Netflix and AWS, they compete with each other as content providers. As part of its Amazon Prime offerings, Amazon.com, Inc. is a digital provider of original TV shows and movies, produced by its own studio, Amazon Studios, as well as content purchased from other producers. Similarly, Netflix produces its own content and provides third-party content to consumers. So while Netflix is a direct competitor in original and third-party content, it shares AWS infrastructure, side-by-side with Amazon's content.

The Netflix case provides an example of a digital service provider that appears unable to deliver services without AWS. While redundancy reduces the likelihood of needing to rely on an alternate provider, Netflix has not provided any public assurances to its stakeholders that it would still be able to deliver value apart from AWS. The next case expands the analysis from digital entertainment to cloud solutions that impact transportation.

Lyft

Lyft is a ride-sharing application in which riders are paired with drivers through each party's mobile device. Lyft provides background screening and ratings for drivers

and facilitates payments between rider and driver. The Lyft case study demonstrates how applications on mobile devices interact through wireless technology, the internet, and the cloud in order to coordinate real-world, physical activities.

Lyft riders request a ride through Lyft's mobile application. This application uses the mobile device's internet connection by way of cell tower or wireless router to send a message through the internet to the application servers hosted in the AWS cloud. Here, the rider's request for a ride is paired with a driver available to service that route. The cloud-hosted application then notifies the driver by way of the internet and cell tower through the driver's mobile application. Rides are usually summoned by smartphone, and drivers are notified by the same technology. The case of Lyft, Inc. provides insights into the relationship between a transportation company and a cloud provider, as well as insights into the architecture of the supporting cloud infrastructure. This case demonstrates how some methods of transportation are entirely dependent on cloud infrastructure to operate.

Lyft is a cloud-native application, which means that the company never operated its own data center. From the initial launch in 2012 (Lift, Inc. 2019), Lyft has operated entirely in the cloud (Lambert, 2016). Lyft is a company that took advantage of the ease in scalability provided by AWS and has grown in the breadth and depth of its AWS footprint and experience.

On March 1, 2019, Lyft issued its S-1 registration statement detailing its initial public offering of stock (Lift, Inc., 2019). Within this registration statement Lyft establishes itself as a major player in the U.S. ridesharing industry at 39% (p. 2). This same registration statement outlines a deal with AWS to spend a minimum of $80 million

per year with AWS, totaling not less than $300 million over three years. This document establishes the footprint of Lyft in the transportation industry as well as revealing the rough order of magnitude of the contract between Lyft and AWS. Lyft goes on to disclose, “We primarily rely on Amazon Web Services to deliver our offerings to users on our platform, and any disruption of or interference with our use of Amazon Web Services could adversely affect our business, financial condition and results of operations.” This confirms a primary reliance on AWS for operations and suggests that ride-matching would be unavailable in the event of an outage. Furthermore, Lyft provides another example of firms contributing technological tools to the AWS environment, free of charge. The next section will examine the architecture of the Lyft’s infrastructure.

Like Netflix, Lyft shares this experience with other firms in the form of open-source software as explained below. The Lyft infrastructure exists almost entirely in the cloud hosting environment provided by AWS. In order to better monitor their cloud-based infrastructure, Lyft developed a distributed proxy called Envoy. At a presentation at SRECON (Site Reliability Engineers Conference), Klein (2017) explains how Lyft uses Envoy Proxy to create a “service mesh”. As a distributed proxy, Envoy Proxy enhances the load balancing technology used by AWS.

A load balancer makes it possible for multiple servers to respond as a single server instance, as explained in the previous section on technologies. Envoy Proxy transforms the basic AWS load balancer into one with better tracking, troubleshooting, and with additional features. Normally, AWS handles load balancing for its customers behind-the scenes, without the customers having to be aware of this process. This creates effectively a black box (unobservable process) that is hard to troubleshoot, especially for

network issues. Envoy Proxy connects to various services that are hosted in the cloud and connects them together while providing logging, tracing, and other services that ensure resiliency. If any service or network connection happens to go down, the Envoy Proxy clients on other services are able to report a lack of connectivity. This functionality enables the Lyft infrastructure to respond quicker to outages and ensure consistent service to its customers. Envoy gives Lyft yet another layer of redundancy and safeguards in the AWS environment. As shown below, this not only makes a cloud-based business model more resilient for Lyft, other firms are able to capitalize on this technology as well, and also use it in other cloud environments.

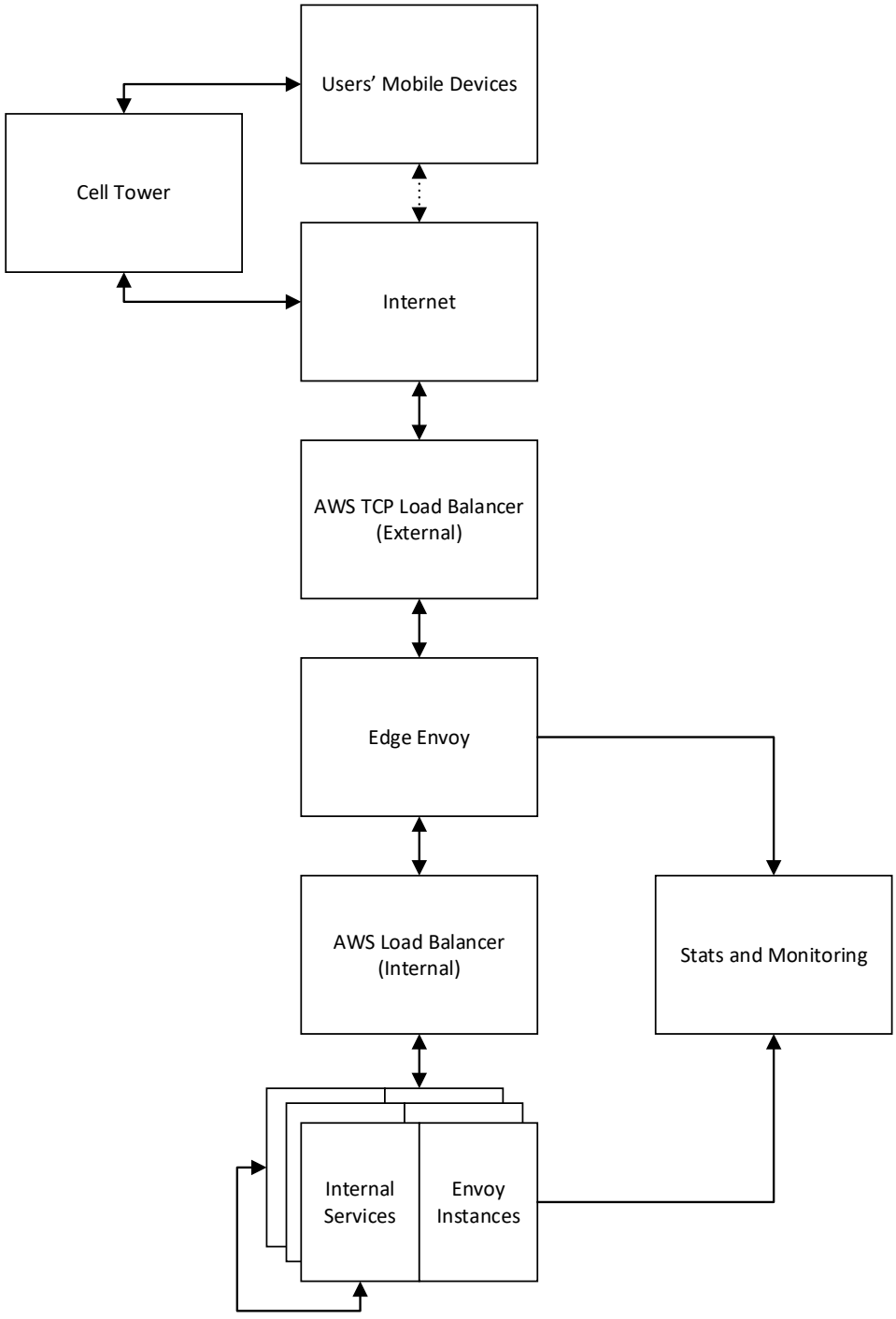
Like Netflix's Chaos Monkey, Envoy Proxy was developed specifically for use in AWS and made available as open source software. However, Envoy Proxy has subsequently been adapted to be used in other cloud environments by a range of companies, including: Airbnb, Booking.com, eBay, F5, Google, IBM, Medium, Microsoft, Netflix, Pinterest, Salesforce, Square, Stripe, Tencent, Twilio, Verizon, and VSCO (Evans, K., 2018). Not only is Envoy Proxy being used in other cloud environments, it is being used by other cloud providers as a building block in their own service mesh offerings. Microsoft built Azure Service Fabric Mesh using Envoy Proxy (Daniel, C, 2018) (Turecek, V., 2018). Google also bases their cloud service mesh offering, Istio, using Envoy Proxy (Yegulalp, S., 2019). In this way, an AWS customer has created a tool, and provided it for free to AWS competitors, who then incorporate it into their paid offerings designed to compete with AWS.

As an alternative load balancer tracking technology, Envoy Proxy effectively competes with the native load balancer used by AWS. This competition is between a

paid, black-box service from AWS, and an open-source solution developed by Lyft. In this way, Lyft is both a customer of AWS and a competitor. In a similar sense, Lyft is also a collaborator-with and conspirator-against AWS in that Envoy Proxy enhances the AWS load-balancing service, but also that of AWS competitors like Google and Microsoft. This multi-dimensional relationship and collaborative competition seems to be especially common-place in the cloud computing industry.

Figure 3.5

Simplified Network Diagram Depicting Lyft's Use of Envoy Proxy



Source: Developed by author based on information presented in this work

While Envoy Proxy is designed to improve the resilience of applications running in AWS, Lyft is still at the mercy of basic AWS functionality in order to maintain its operations. This is confirmed by Lyft's Form S-1 registration statement (ss Filed on March 1, 2019), "In the event that our agreement with AWS is terminated or we add additional cloud infrastructure service providers, we may experience significant costs or downtime in connection with the transfer to, or the addition of, new cloud infrastructure service providers." This is an acknowledgement of the lock-in that occurs when choosing cloud providers. Given enough lead time, it is conceivable that a company like Lyft could migrate away from AWS, but in the event of an abrupt unavailability of AWS, it would incur significant costs and downtime.

Volkswagen

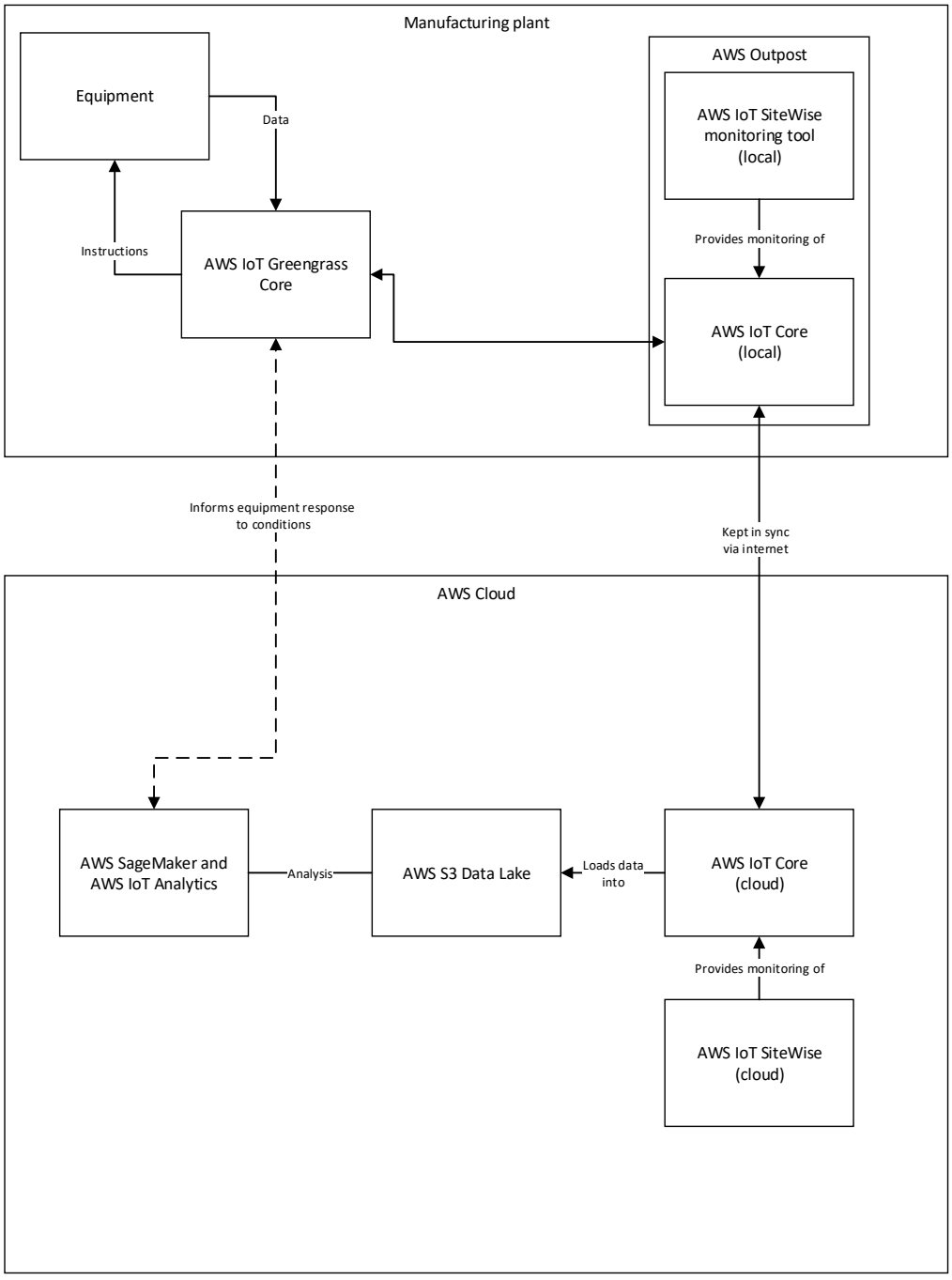
On March 27, 2019, AWS publicized a multi-year agreement to build the Volkswagen Industrial Cloud. The Volkswagen Industrial cloud is a cloud-based digital production platform that will leverage many AWS services, including Internet of things, machine learning, analytics, and basic computing services. The goal is to improve plant efficiency and reduce downtime. This is also expected to aid with plant flexibility while having a positive impact on quality (Amazon Web Services, Inc., 2019h). This cloud solution is expected to bring together operational data from more than 30,000 facilities and 1,500 suppliers and partners. Unlike the previous examples, Volkswagen will make use of AWS Outposts which bring AWS services and operating models onto their premises. This represents a hybrid approach that blends on-premise with cloud solution but all as part of a comprehensive AWS-hosted solution. This example will highlight the breadth of AWS services that are in scope for this joint-venture as well as the way in

which AWS Outposts blends cloud and on-premise solutions with an outcome that is robust to interruptions in the underlying cloud service.

The Volkswagen Industrial cloud is expected to use the suite of AWS IoT offerings, including AWS IoT Greengrass, AWS IoT Core, AWS IoT Analytics and AWS IoT SiteWise. This will be used in conjunction with a data lake that is built on Amazon Simple Storage Service (S3). Amazon SageMaker will allow the development of Machine learning models to improve company processes (Amazon Web Services, Inc., 2019h). These services are provided in a hybrid, on-premise/cloud solution as explained below and diagramed in figure 3.6.

Figure 3.6

Simplified Network Diagram Depicting Volkswagen's Use of Amazon Web Services (AWS), including AWS Outpost and the AWS IOT (Internet of Things) Offerings



Source: Developed by author based on information presented in this work

Volkswagen's actual implementation architecture has not been shared publicly but the following description is based on the services listed in a Volkswagen press release. Anticipating the likely integration of those services, this description provides a high-level configuration of the relevant services. This description shows how those services would form a hybrid on-premise/cloud model.

Each manufacturing plant has their own network. Within this network, devices are then configured to interact with AWS IoT Greengrass Core. Greengrass Core receives information collected from inputs on the devices and can return commands to those devices based on models or algorithms that have been pre-programmed. Greengrass Core is designed to allow plant devices to respond to new data and interact with each other without any dependency on internet connectivity (Amazon Web Services, Inc., 2019e). This ensures that system uptime is not contingent upon internet connectivity or AWS cloud availability.

AWS Greengrass Core will connect with AWS IoT Core, at least intermittently, to share data about and from the connected devices. As depicted in the diagram, IoT Core can be hosted partially on-site to reduce the latency of responses between IoT Core and the devices that are connected to it through AWS IoT Greengrass Core. IoT Core is the primary interface by which devices in geographically disbursed plants can be connected and managed holistically. (Amazon Web Services, Inc., 2019d)

In order to be hosted on-site, the plant would need an instance of AWS Outpost. AWS Outpost is a service by which a collection of servers are hosted on premise, but connected to the local AWS region (Amazon Web Services, Inc., 2019f). This means that time-sensitive analysis can be conducted on the data retrieved from devices and

commands returned to those devices in real-time. The local implementation of AWS IOT Core is able to conduct more complex analysis and incorporate more data than Greengrass Core. By being connected to the cloud through the internet, the AWS Outpost instance allows both the reduced latency of a local solution with the vast computing power of the cloud that can be unleashed for less-time sensitive or more complex computations.

Paired alongside AWS IoT Core is the AWS IoT SiteWise monitoring tool. Like IoT Core, SiteWise can be hosted within AWS Outpost to reduce latency for local monitoring, but also connect to the broader cloud network in order to provide remote-viewing of other locations. SiteWise provides pre-packaged programs that collect, aggregate, and visualize the data produced by connected devices. While much of the IoT infrastructure examined here is geared to machine learning and machine interaction, SiteWise is intended to visualize this information in a way that can be consumed by the human members of management.

AWS Simple Storage Service (S3) is a premier unstructured data storage service. Volkswagen intends to create a data lake on this infrastructure that will support near real-time data storage and retrieval as well as longer term analysis. The analytic services will utilize this data in order to build models that help better manage or coordinate the connected devices.

AWS IoT Analytics and AWS SageMaker are two tools for analyzing the data generated by connected devices. AWS IoT Analytics is designed to help companies glean insights from the unstructured data generated by connected devices (Amazon Web Services, Inc., 2019c). It is specifically designed to make use of IoT data with relatively

simple machine learning and modeling. SageMaker, on the other hand, is a tool designed for developers and data scientists to be able to create more advanced machine learning models and quickly deploy them (Amazon Web Services, Inc., 2019a). The models developed by these analytic tools can then inform the programming that is hosted on the individual devices or on Greengrass Core, and then new data can be generated and tracked to determine the effectiveness of these models in practice.

Volkswagen's IoT implementation in their automobile manufacturing plants is an excellent example of the hybridization of on-premise data centers with cloud infrastructure. By design, the implementation of AWS IoT Greengrass prevents internet or cloud disruptions from spreading to the manufacturing equipment. In this way the Volkswagen plants will be resilient to an AWS outage in a way that Netflix or Lyft are not.

From a security standpoint, however, if Volkswagen's cloud computing resources were compromised, a well-informed actor could disrupt all 30,000 facilities if not detected. This disruption could even extend to suppliers, customers and others who might be on the receiving end of automatically generated orders and communications. Additionally, a major compromise in the security of the hypervisor or software supporting the cloud could also provide the ability to damage or disable the actual plant equipment connected through Volkswagen's IoT implementation. So, while Volkswagen is protected against intermittent outages at AWS, its fate is linked to all other users of AWS in the event of am AWS system compromise.

In summation, the industrial implementation of cloud computing appears to have anticipated the most likely causes of outages and disruptions in services and designed

processes that prevent these digital interruptions from disrupting the operations of the physical plant. Compared to the previous industries considered, the IoT and industrial implementations of cloud computing appear to be the most robust.

Capital One and OakNorth

The health of the financial sector has a major impact on the health of the broader economy, and the health of the financial sector is increasingly dependent on servers and networks. Financial institutions have been conducting business across networks and storing key financial data digitally for decades. Information technology services facilitate the critical functions of financial firms. These information technology services are increasingly hosted in the cloud. The cloud provides the agility that financial companies need in order to keep pace with their competitors (Shevlin, R., 2019). In the section that follows, two companies will be considered as they relate to AWS, one that is in the process of transitioning on-premise data centers to be 100% cloud in the next five years, and a UK banking company that whose core operations have taken place in AWS for the majority of its existence.

Capital One is a financial company that is "all in" on AWS. According to Capital One Vice President of Cloud Strategy, Bernard Golden, "Our plan is to shut down data centers in the somewhat near future—not five years from now; nearer term than that." (Asay, M., 2018). If successful, this would mean that by 2023, the entire computing functions of the 10th largest bank in the United States would be entirely dependent on the cloud (Dixon, A., 2019). While Capital One has been a public advocate of cloud computing in the banking industry, it is not clear to what extent Capital One is still relying on legacy (non-cloud) infrastructure. Because information was not readily

available on the current state, this case study will examine the near-term projected future state, of operating 100% in the cloud, and look at what types of processes would then become dependent on the cloud. These processes can be broadly categorized as customer-interfacing, bank-to-bank, and internal operations. Internal operations are only relevant to this discussion to the degree that they impact external parties, so the focus is given to customer-facing and bank-to-bank operations.

Given the general reliance on digital records and documentation, almost every customer interaction is facilitated by and therefore contingent upon the availability of the digital system. This applies both to in-person interactions at a brick-and-mortar location or online transactions conducted via the internet. Deposits are credited to customer accounts through a digital process that tracks credits to and debits from customer deposits. Account inquiries requesting to know the amount of available funds are generated by the computer system, and so valuation is also contingent upon system availability. Loan creation, and loan extension is also dependent on the computer system which tracks creation and changes to loan agreements. Withdrawals and transfers between same-bank accounts would be impossible as well as any transfers between banks. Paper checks often experience a delay in processing, and so are not contemporaneously dependent on the computing systems of the issuing Bank. This means a check could likely be used to conduct a transaction if the receiving party is not using the check to initiate a digital ACH. Additionally, credit cards that are not directly serviced by the issuing Bank are likely to operate independent of the issuing Bank's information system and cloud platform.

The majority of interactions between Banks takes place digitally. Wire transfers between Banks occur in real-time and take place directly between the computer systems of the individual Banks. Both systems would need to be operational in order to process wired funds as each set of funds are verified in the process. Automated clearing house transfers (ACH's) require files to be submitted to The Clearing House or the Federal Reserve System by the originating financial institution. It is not clear that there is any other option for submitting payments in the event that the originator's computer systems are not functioning. A Bank which is does not have a functioning computer system would not be able to receive funds from the clearing organization.

If Capital One succeeds in making its business completely reliant on cloud computing, then in the event of an outage it is unable to perform any of its usual services to its customers, and would not be available to any other financial institutions for transfers by wire or ACH. Credit cannot be extended and withdrawals are unlikely to occur as well. On the positive side, a run on the Bank does not seem possible during the outage, as transactions are not possible. On the negative side, through virtual Banking, a run on a Bank is not contingent upon the speed of a teller and funds can be digitally removed in seconds or less with the click of a button.

While Capital One is an example of an existing financial institution converting its systems to the cloud, OakNorth is a Bank that has already achieved this milestone in Great Britain and now seeks to expand its influence internationally. In 2015, the Financial Conduct Authority (FCA) published proposed guidance for UK-regulated financial services firms when using cloud-based services (Maughan, A., 2015). This guidance provided the clarity of policy needed for the emergence of the first cloud-based

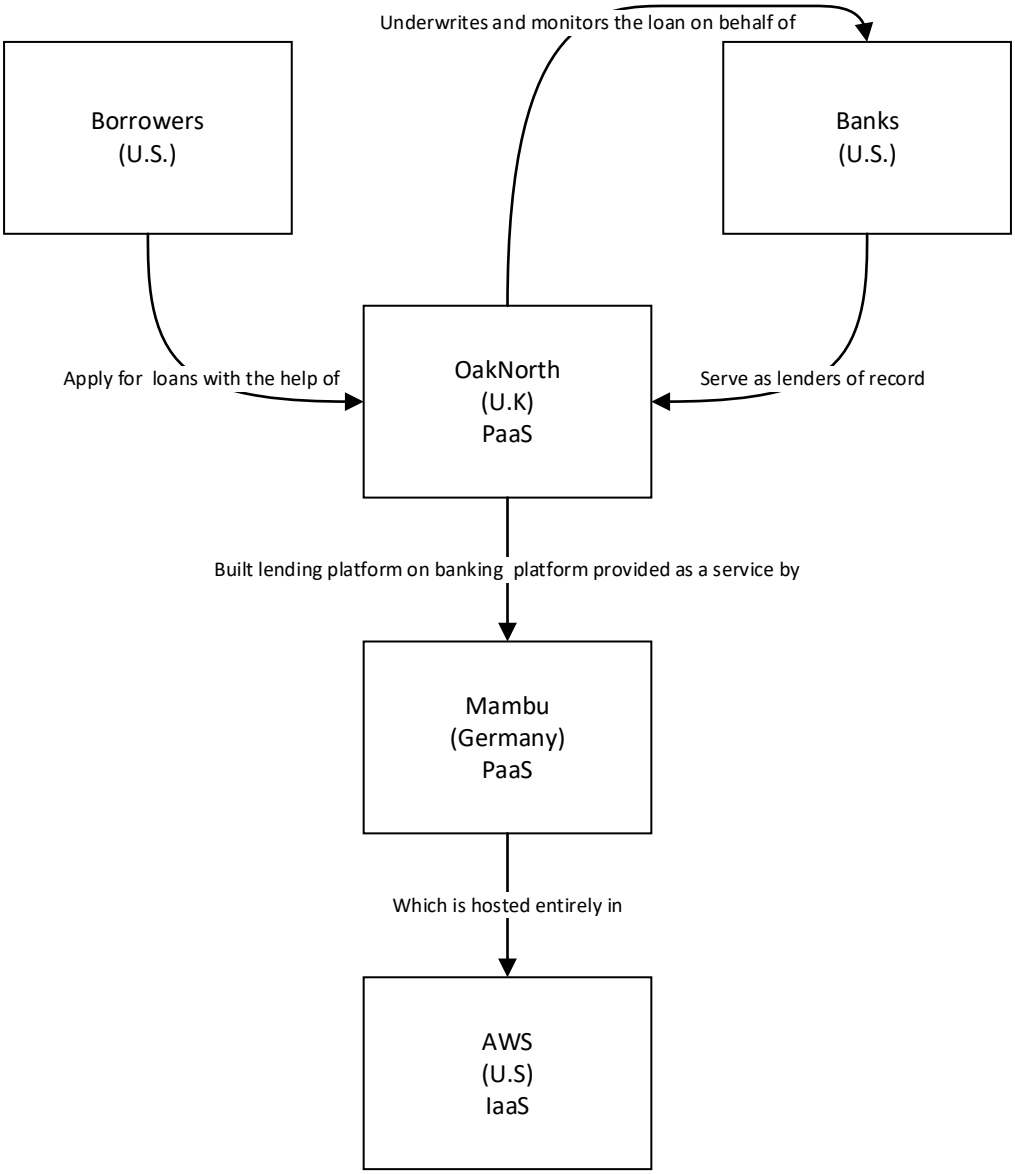
Bank in the UK (Carey, S., 2016). OakNorth received its banking license in March of 2015, and began providing medium-to large loans to entrepreneurs in September of that year (Carey, 2016). In November of that year, FCA published its guidance and OakNorth began transitioning all of its IT services to the cloud. It has since hosted all of its core banking operations in the cloud, utilizing banking software developed by Mambu, a German provider of cloud-based financial software (Carey, 2016). OakNorth is venturing into the US market, but does not have a U.S. banking license (Wintermeyer, L., 2019). Lacking a license, OakNorth is not able to be the lender of record, but plans to originate, underwrite, transact and monitor loans through partnerships with banks that are registered in the U.S.

The chain of dependence represented by the relationship between U.S. customers, U.S. Banks, OakNorth, Mambu, and AWS is show in figure 3.7. In this diagram, a U.K. Bank, running on a German platform, hosted on U.S. infrastructure plays intermediary between lender and borrower in the United States commercial loan market. This global concatenation of industry is not unique, as global companies commonly rely on the products and services of other global companies. What is worth noting from this example is how the platform-as-a-service (PaaS) model places borrower and lender at the mercy of the contemporaneous availability of OakNorth, Mambu and AWS. If any of the upstream service providers makes a meaningful change to their platform, the downstream service providers have to respond to this change in a timely fashion to ensure that the services of each subsequent platform are not impacted by this change. Additionally, each level of dependence may experience a time lag as the changes to one platform impact the platform offered by downstream service providers to customers further downstream. This

division of labor is likely to bring efficiencies, but also adds complexity to the composition of the banking industry.

Figure 3.7

Diagram of OakNorth U.S. Business Model and Significant Partnerships



Source: Author's depiction of information from Wintermeyer (2019)

In the case of OakNorth, an interruption of service at any level does not necessarily have an immediate impact on borrowers, as once the loan is secured, the borrower has very little interest in the monitoring of that loan. The lender, on the other hand, may be more impacted, as they lose their automated connection to the financial status of the borrower. That is to say that an interruption within this chain of dependence is unlikely to have a debilitating impact on the interactions between borrower and lender unless the lender has come to depend on OakNorth for the processing of payments and crediting of accounts. The processing of payments and crediting of accounts is likely part of the core functions of most banks and therefore likely outside of the scope of the services provided by OakNorth in the U.S.

The OakNorth example demonstrates the degrees of separation that arise through the division of labor and specialization of platforms and sub-platforms in the banking industry. In this instance, Banks still play the ceremonial role of lender of record as well as the source of monetary creation, but much of the banking operations take place outside of the immediate purview and control of the banks, themselves. The example of Capital One demonstrates a directional shift, at least for one Bank, toward this hybrid model in which Banking services rely on non-financial and quasi-financial intermediaries for their core business functions. Having demonstrated that the banking industry is increasingly reliant on cloud computing, the focus now moves to government activity. For the purposes of this research, civilian government activity will be considered separate from military activity. Beginning with civilian activity,

U.S. Government

Government users of AWS have more stringent security requirements than most civilian users. The data gathered by Federal government agencies is vast, and a breach could potentially impact every citizen. In many cases this data is sensitive and would cause harm in the wrong hands. The data may also be proprietary to the government agencies and could undermine their credibility or operations. Similarly, if the integrity of the data is compromised, it would disrupt daily operations. Therefore, it is imperative that government agencies preserve the confidentiality, the availability, and the integrity of this data. To this end, the government created the Federal Risk and Authorization Management Program (FedRAMP) for agencies seeking to utilize cloud-based services. FedRAMP provides "a standardized approach to security assessment, authorization and continuous monitoring for cloud products and services" (Federal Risk and Authorization Management Program, n.d.). FedRAMP created and executes a standard process to certify cloud services for use by the Federal Government. There are four security baselines (high, moderate, low, and LI SAAS) ranging from 421 controls on the high end to only 36 controls at the lowest end. Once a cloud service provider receives their authority to operate (ATO) for a given baseline, that service provider may then reuse that baseline ATO with multiple government agency customers. This system frees each government agency of the burden of certifying each cloud service provider on its own and it also allows the cloud service providers to become certified once for a given

baseline and then provide services to all government agencies that fall under that baseline.

Solution providers running on AWS can receive assistance in achieving FedRAMP certification through the Authority to Operate on AWS program. This program is an Amazon Partner Network program in which AWS provides resources that help third-party cloud service providers achieve baseline authorization (Amazon Web Services, Inc., 2019b). These resources streamline the authorization process and help new third-party (non-AWS) cloud service providers to gain access to U.S. government agencies to provide cloud services. With an increasing number of approved resources in that marketplace, AWS can expect increased use of its infrastructure for government work. This also makes it quick and easy for government agencies to be able use any products that have been authorized for use by other agencies. These three stakeholders (AWS, the agency, and the third-party cloud provider) each benefit from the common platform and growing marketplace for authorized services.

From a security perspective, this symbiotic environment provides three advantages. First, by working closely with FedRAMP, AWS is enabled to understand the security needs of federal agencies and the stipulations of FedRAMP. This can contribute to a better execution of those standards both by AWS and by the third-party cloud service providers it assists. Second, by working with a variety of service providers, and by observing multiple ways of complying with those standards, AWS is positioned to be able to determine best practices for cloud solutions and then facilitate the sharing of those practices between cloud service providers that utilize the AWS platform. Third, by working closely with FedRAMP, AWS has the potential to provide constructive feedback

to influence the formulation of FedRAMP standards to keep pace with technological change.

A downside of this symbiotic environment on the security of government cloud applications is that AWS has more incentive to onboard new applications into the environment than it does to ensure the security of those applications. Consequently, rather than advising companies in the spirit of the guidelines AWS may focus instead on meeting the minimum requirements and funnel new service providers through previously identified shortcuts or loopholes. A second potential downside is that when the responsibility for selecting security configurations is centralized and passed down as practices, fewer entities are seriously considering the security settings and their implementation. Therefore, the convergence of infrastructure around common practices coached by AWS may not be true best-practices and may introduce risks that would exist across all applications following these practices.

Concerning competition, the AWS platform and the efficiencies it garners for its stakeholders pose a barrier to entry for competitors of AWS. As the AWS environment provides a clear path to FedRAMP certification and a variety of cloud solutions to government agencies, it may be less desirable for those agencies to do business with competitors of AWS, and those competitors may find it more difficult to find vendors willing to offer FedRAMP-certified services to host on their competing platforms/infrastructure.

Another noteworthy aspect of competition as it relates to the AWS FedRAMP ecosystem is the fact that AWS competes with many of its customers on the services it provides to federal agencies. For example, Rackspace is another provider of

infrastructure as a service (IaaS), which competes with AWS, but in the FedRAMP listing of products, Rackspace is listed as having dependencies with AWS (Federal Risk and Authorization Management Program, 2019). This implies that Rackspace is dependent on a competitor for its ability to provide services to federal agencies. This dependency is more precarious for Rackspace as AWS is streamlining the process for competitors to become FedRAMP certified. As demonstrated in previous examples, much of the interdependencies within the cloud eco-system depend on the magnanimity of AWS in not abusing its market power.

The Central Intelligence Agency (CIA) awarded a $600 million contract to AWS in 2013 to host a private cloud on their behalf (Nix, N., 2018). This private cloud is hosted within the CIA's own data centers and is known as Commercial Cloud Services (C2S). C2S was constructed to provide the services available in the commercial cloud to those organizations which require a highly secure environment, even above FedRAMP High.

This "secret region" is a private availability region that emulates an AWS availability region. This region is made up of three availability zones, in the same fashion that a public AWS availability region is divided into availability zones. Each zone is hosted in three geographically dispersed data centers. This setup is explained to be almost identical to the arrangement within the AWS public cloud, with the exception that this private cloud is air-gapped from the internet and all other data centers. A representative of AWS has claimed that this private cloud is the first air-gapped commercial cloud (Amazon.com, Inc., 2017c). What that means is that the three availability zones are networked together across a private network, sending no information through the internet.

Furthermore, the data centers that make up this network are not connected to the internet so that there is an "air gap" between them and any would-be attackers that may try to use the internet as an avenue of access to the data centers or the systems and data contained therein. John G Edwards, Chief Information Officer of the CIA, describes this system as 'almost' impenetrable at a speech given at the AWS Public Sector Summit in June, 2017. He describes Commercial Cloud Services like this, "We took this battle-tested hardened cloud that's on the outside, that's connected to the internet, dropped it behind our guards, gates, and guns, and we're not connected to the internet" (Amazon.com, Inc., 2017b).

The physical security of the data centers is the responsibility of the CIA, but it is unclear who protects the connections between the data centers. Each data center in an availability zone is connected to the other two, and then the availability zones are connected to each other within the region. Geographic dispersion adds redundancy and resiliency to this distributed system and so these connections are many miles long. While data transmitted between data centers is required to be encrypted, simply damaging the connection could render the private cloud unavailable.

The CIA hosts data in this private cloud and then makes it available to the broader United States intelligence community. This data includes all classification levels: Unclassified, Sensitive, Secret, and Top Secret (Amazon.com, Inc. 2017c). This provides users with a relatively consistent toolset when working with their data across classifications. This toolset is also, ideally, consistent with what is available in other commercial cloud environments so that there is less of a learning curve for new hires working with this data. This also ensures that training and best-practices are transferrable between public and private cloud users.

The example of the CIA provides a unique application of AWS technology. Unlike other customers of AWS, the C2S is not dependent on the broader AWS infrastructure. Being completely air-gapped from the internet and the public cloud means that no outage, internet or network issue taking place in the public cloud would impact this private cloud offering. However, C2S is dependent on Amazon technology in much the same way as the public cloud. This means that C2S is still dependent on:

- AWS employees
- AWS processes
- AWS server and System configuration settings
- AWS network design
- Hardware vulnerabilities potentially of the same variety as the public cloud

With the setup of the private availability zone, AWS employees are needed to care for this private cloud solution. Employees could be one source of unauthorized access to the system. AWS processes would be employed to maintain the cloud and such procedures can be reasonably expected to mimic those used in the public cloud. This may include accountability systems, including the separation of duties among different roles. This systematic separation of duties is designed in such a way as to limit the potential for malicious actors to independently make harmful changes within the system. Any shortcomings of this systematic approach could potentially undermine both public and private cloud solutions hosted by AWS. These accountability systems may also include access management and provisioning, which is the means by which AWS administrators gain access to portions of the system they need in order to carry out their responsibilities. Additionally, AWS security protocols, and especially incident response plans can be

expected to be similar between the public cloud and the private cloud. An understanding of how these processes are executed in the public cloud would likely inform malicious actors as to how they are executed in the private cloud.

AWS server and system configuration settings of both hardware, software, and operating systems can be expected to be mirrored between the public and private cloud. This means that a vulnerability discovered in the public cloud that is related to the configuration settings is likely to be present in the private cloud. In this instance, the air-gap and CIA's physical protection of the data center locations provide a control against this type of exploitation. This means that malicious actors may be aware of a weakness within the private cloud but would need a way to physically access the private network before they would be able to exploit it.

AWS network design is similar to the issue of configurations settings in that knowledge of weaknesses within the public cloud may be used against the C2S private cloud. Similar again, is the idea that access would be required to exploit such an issue. The main difference is that certain network components may be geographically distributed across a region. For example, in order to connect data centers while maintain the air-gap from the internet, a dedicated connection, presumably similar to the Amazon Direct Connect service, must be established between these data centers (Amazon Web Services, Inc., 2018). This is most likely a physical cable that sends encrypted data between data centers. While the data centers which host C2S are described to be walled-in and guarded, the degree of physical protection available along the length of the physical, wired connection could not be confirmed. It is possible that information about

the physical connections among the data centers of the public cloud may make it easier to identify and exploit the connections between the data centers of the private cloud.

Similar to the way that network configurations in the public cloud might inform configuration in the private cloud, the hardware used in the private cloud is presumably similar to that of the public cloud. This similarity means that risks introduced into the public cloud through hardware have likely also been introduced into the private cloud. This could manifest in three main ways: hardware malfunction, hardware vulnerability, and hardware compromise. If hardware used in the public cloud is defective and causes problems, it is likely to undermine the private cloud in a similar fashion. Beyond malfunctioning, hardware may introduce a vulnerability into the system, or might be consistently misconfigured in both the public and private cloud due to similar processes used in both. In this instance, malicious actors may be able to exploit such a vulnerability, but access is likely required in order to take advantage of such a vulnerability in C2S. Finally, hardware may be compromised by a malicious actor, including a foreign government. In this case monitoring or back-door access devices or settings could be embedded within the hardware in the private cloud at the same time it is embedded in the public cloud. In these three ways, hardware similarities between the public and private cloud may serve to either partially or completely couple the fate of the private cloud with that of the public cloud.

The previous five aspects of AWS technology illustrate the degree to which a mutual dependence on such technology links the fates of private and public cloud. Even when the two cloud implementations are completely independent and air-gapped from one another, a reliance on similar technologies can cause both to become vulnerable

simultaneously. So, while a security compromise in the public cloud cannot spread directly to the private cloud, it can potentially be introduced through the same means.

Chapter III Conclusion

This section has established the dominance of AWS, demonstrating how cost structures, network effects, and coopetition make the cloud computing industry unique. This uniqueness often finds Amazon's competitors dependent on AWS. Close coupling of AWS with other firms and industries can allow issues in AWS to spill over into a variety of sectors in the economy. It is easy to see how companies are increasingly reliant on the cloud for their daily operations and thereby imagine the potential loss in the event that AWS becomes unuseful. While this section delved somewhat into the risks faced by stakeholders of AWS as well as highlighting a few of the controls in place to mitigate this risk, the next section will take a structured approach of analyzing the risk of an AWS failure, the extent and the health of the controls in place, as well as how those controls mitigate the likelihood and impact of an AWS failure.

CHAPTER IV

ANALYZING THE RISK OF AN AWS FAILURE

This chapter begins by examining the ways by which AWS might fail. The remainder will discuss the effects of a data breach or prolonged outage. The proprietary nature of the AWS environment limits inter-operability with other cloud providers. Third-party managed cloud service providers are cropping up to bridge the interoperability gap, but the majority of companies using cloud services are not making this investment. One section will detail the potential victims in the event of an AWS failure and how each might be harmed. This will explain how such an event might have a cascading effect on the economy as a whole. The extent of the harm provides motivation for the interventions considered in Chapter V. This chapter considers the path to recovery in the event of an unmediated failure of Amazon Web Services. This analysis finds the existing institutions inadequate to support such a recovery and rectify damages. These inadequacies includes ambiguous legal precedent, insufficient insurance, privacy issues, and a lack of interoperability standards.

The three main aspects of information security are confidentiality, integrity, and security. These three aspects are outlined with respect to the cloud in Tchernykh, Schwiegelsohn, Talbi, & Babenko (2019). Confidentiality of data denotes the restriction of access to only those for whom it was intended by the administrators and/or owners of an application or data set. Integrity refers to the ability to trust the accuracy or completeness of the data or outputs of an application. Availability seeks to guarantee that the correct users are able to access the information system reliably.

One prominent approach to risk management is known as the bowtie method. This method is a visual representation of possible risk events, which are tied on the left to their causes and on the right to their impacts. Also depicted are those risk mitigants and controls which are intended to reduce the risk of the depicted event. The mitigants between the causes and the event exist to reduce the likelihood of the risk event occurring, while the mitigants between the event and the impacts represent impact reduction in the event that the risk event does occur. A secondary level of threats is also identified which might undermine the effectiveness of the first level of controls. Against these threats, secondary levels of controls are mapped to reduce the likelihood of the original control failing through that given method. Further levels of controls are possible, which might act as controls on the controls on the controls, etc.

The bowtie method can be applied to a variety of activities in risk management. For the purposes of this study, the method will be used to identify, organize, and evaluate the effectiveness of the controls that are in place to reduce the risk of certain significant risk events related to cloud computing. This approach will identify two main risk events associated with the operations of AWS and then enumerate and assess the controls that are in place to reduce the likelihood and impact of their occurrences.

Alizadeh & Moshashaei (2015) provide a literature review of the bowtie method, tracing its origins to the "Swiss Cheese model" presented by Reason (1990). In the Swiss cheese model, the barriers between a hazard and actual loss are represented as slices of Swiss cheese, each with holes that represent latent pathways by which accidents may happen or holes created by active failure (Reason, 2006). When the 'holes' of each barrier between a hazard and loss line up, then loss is realized. That is to say that it takes

a contemporaneous failure of each control in succession in order for hazards to become losses. The bowtie method depicts the alignment of controls along a causal pathway, implicitly showing how a risk event requires each control ineffective in order to occur. Alizadeh & Moshashaei (2015) find that the bowtie might have originated in 1979, but the was not known to have been fully integrated into the business practices of an organization until the 1990s, when the Royal Dutch/Shell Group of companies implemented the bowtie approach broadly into their operations. Alizadeh & Moshashaei (2015) find examples of the bowtie analysis being used in a variety of industries beyond petrochemicals, including air travel, ship building, finance, and a variety of international government and regulatory agencies. Alizadeh & Moshashaei (2015, p. 136) conclude that this method is "ideal for structured assessment and communication of risks."

The first step in conducting a bowtie analysis is the identification of a hazard. This hazard, if not handled correctly becomes the source of undesirable events. The two hazards considered in the case of AWS is that of operating a system containing high-value data and high-impact processes, and providing critical services to high-visibility/impact customers. Connected with these two hazards are the primary events which represent a failure to contain or control the hazard. These are malicious compromise, and failure to deliver expected services.

Malicious Compromise

Hazard

The first hazard identified is that of working with high-value data and high-impact processes. Because of the many organizations that store high-value data on AWS and depend on AWS for their business processes, AWS is a substantial target for malicious

actors. The main risk event associated with this hazard is a malicious compromise of AWS information systems. A malicious compromise represents inappropriate access to the information systems that comprise Amazon Web Services. The bowtie for this risk event models the risk from cause to impact, beginning with the means of access to the protected system, and ending with the variety of impacts that closely follow a malicious compromise.

While not captured in the bowtie, the breadth of actors will be summarily addressed. On the more benign end of the spectrum, attackers may simply be activists looking to gain notoriety, publicity or make a point by demonstrating their ability to compromise or undermine the operations of AWS. To the extent that this becomes destructive, activism descends into terrorism and customer data and processes would presumably be impacted. This type of terror-based attack could be state-sponsored cyber-terrorism or an act of cyber-warfare in which a nation state is seeking to advance its political or military agenda through the compromise of a protected system. A breach may also represent corporate espionage, either in an attempt to acquire sensitive intellectual property, or to undermine the operations and strategic activities of a competitor. Finally, a malicious compromise may be motivated by financial gain, through usurping processes, extortion or selling sensitive information. A summary of these actors and their end-goals is depicted in Table 4.1, below.

Table 4.1

Sources and Motivations of Selected Threats to Amazon Web Services and Cloud Computing

Source	Motivation	Service Usurped	Data Theft	Service Disrupted	Compromise Publicized
Nation-State	Political/Military	x	x	x	x
Activist/Terrorist	Ideological/Destructive		x	x	x
Corporate Entity	Competitive Advantage		x	x	x
Thief	Financial	x	x		

Source: *Developed by author based on information presented in this work*

As shown above, a malicious compromise could be carried out by a variety of perpetrators for a variety of reasons. Given that this bowtie analysis is intended to evaluate the effectiveness of the controls in place to mitigate the likelihood and impact of the malicious compromise, the bowtie analysis does not focus on motivation or type of malicious actor, but rather on the means by which malicious actors might carry out an attack.

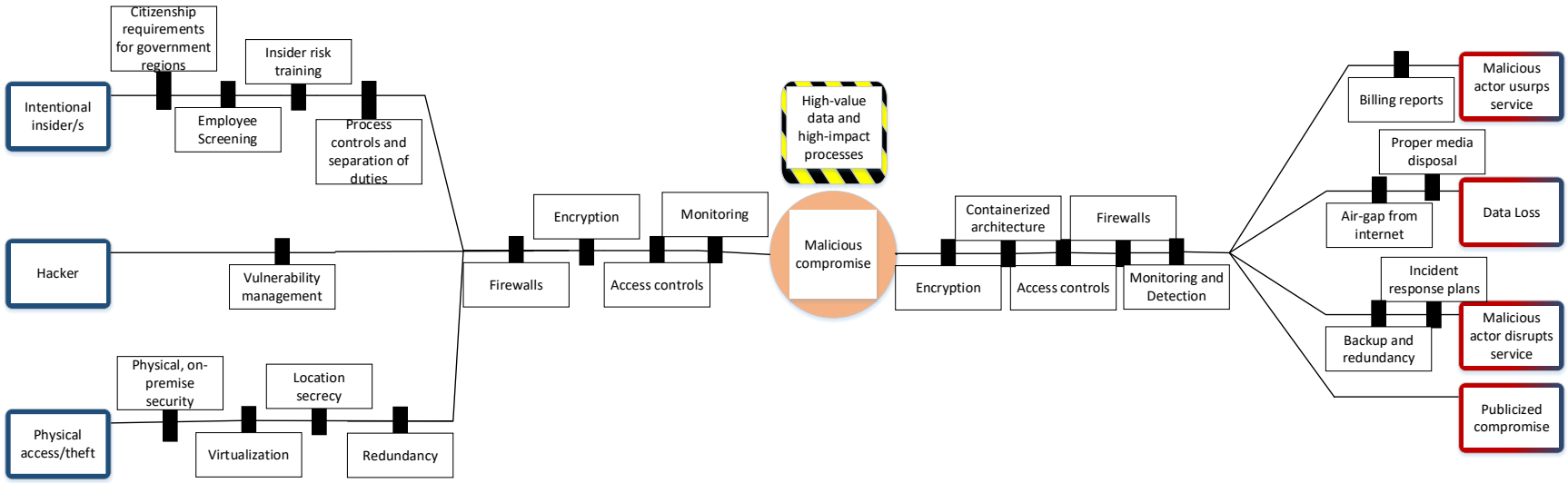
Causes

The three main means of access by which a compromise is achieved are as follows. First, employees, contractors, or partners, here referred to as insiders, may intentionally misuse their knowledge of or legitimate access to an AWS system for malicious purposes. Second, hackers may gain access to the system from the outside, through a vulnerability in the security architecture of the AWS system. Third, a malicious compromise may be achieved by securing, through force or deception, to the physical access data centers and/or the theft of physical storage devices. Protecting AWS from

each of these three sources of risk are a variety of controls, some of which protect against all three types of attack, others of which are specific to each means of compromise. See figure 4.1.

Figure 4.1

Bowtie risk diagram depicting the risk of malicious compromise in the AWS cloud



Source: *Developed by author based on information presented in this work*

Broad Mitigants on Likelihood

The four controls which are intended to prevent all three means are encryption, firewalls, access controls, and monitoring. Encryption is explained previously in the technology section and refers to both encryption in-transit, as well as encryption at-rest. In all three means/causes of compromise, attackers would need to secure a way to decrypt the data they obtain. Intentional insiders that may have access to data centers would need not only access to the servers housing the data, but the encryption keys themselves. This is the same for hackers who may be able to virtually access various servers (virtual or physical) or services, but this access must also accomplish the decryption of the data which they are seeking. Similarly, information that is intercepted between services is encrypted in-transit, which means that a key is still required. Additionally, encryption-in-transit guards against the usurping of services. Malicious messages intended to imitate legitimate messages would need the proper encryption if they are to successfully impersonate messages within the system.

Firewalls are a network security component that determines whether or not to block certain network traffic based on pre-defined rules (Cysco Systems, Inc., n.d.). Firewalls segment servers, services and systems based on their needs to interact with one another. If properly configured, these firewalls ensure that only the intended messages from approved sources are allowed between servers and services. Cysco Systems Inc. also points out that firewalls can be both virtual and physical. In the core infrastructure of AWS, there can be expected to be both physical and virtual firewalls. However, in AWS cloud service offerings, the customer takes responsibility for the virtual firewalls created

within their applications and services. A properly configured firewall can prevent an insider with legitimate access from using that access in unauthorized ways. Firewall protections help to implement the principle of least privilege. This means that access rights of users, accounts, and computing processes are limited to only the minimum levels required to accomplish a given task (BeyondTrust, 2016). If the firewall control is effectively implemented, then hackers that gain access to one portion of the system would be limited from accessing other and more critical processes or sensitive data. Similarly, firewalls may even prevent or increase the difficulty with which someone with physical access to a data center from being able to access other critical components in a distributed application or service.

Access controls also guard against all three means of malicious compromise. Access controls, if properly implemented, prevent insiders from accessing more than is required for their specific roles as explained by least privilege access, above. Additionally access control systems create a record of access that might notify supervisors or security personnel if access is abused. Access controls prevent compromise by a hacker in the same way as for an intentional insider by reducing the amount of access that can be acquired by impersonating a single, or even several insiders. This limits the ways in which hackers may obtain access as well as the degree of access achieved by a single compromise. Access control systems are inclusive of physical access controls, which ensures that only the right people can access the physical infrastructure under the right circumstances.

The final of the four broad controls on malicious compromise is monitoring. Monitoring in the preventative sense is to detect suspicious behavior and attempts at

compromise. Insiders and those with physical access expect to have their activities monitored and disciplined monitoring of the system boundary is likely to detect the presence of a hacker or unauthorized access within the system. Monitoring also acts as a deterrent, by increasing the likelihood of being discovered. This deterrent aspect also applies across all potential causes identified in the bowtie.

These four types of controls, when applied correctly, reduce the likelihood of a malicious compromise. The question concerning these three preventative controls is whether or not they are being applied correctly. For the purposes of this paper, evaluating the success with which AWS implements its portion of these controls is beyond the scope of this research. Rather, this evaluation will take as a given that AWS is indeed delivering best-in-class security as a cloud provider. However, Panetta (2019) asserts that in nearly all cases of breaches in the cloud, "it is the user, not the cloud provider, who fails to manage the controls used to protect an organization's data." This makes sense given how AWS explains its shared responsibility model (Amazon Web Services, Inc., 2019g). Amazon explains that AWS assumes responsibility for the security of the cloud, while the customer assumes responsibility for security in the cloud. A close analogy would be like a hotel handing a guest a key to their room and guaranteeing that the only way into that room is through the door with that key. The guest is then responsible for door access and can rest assured that all other entry points have been addressed. If AWS delivers effective security of the cloud, then the vulnerability of each user of cloud services is determined by the security measures taken by that individual user. The undermining of each of these controls is not systemic unless there is a systematic process by which users configure and maintain their encryption, access management, and monitoring.

There is evidence for systematic vulnerability in firewall configuration. This is evidenced by the data breach of Capital One in July of 2019. According to the indictment, the alleged hacker obtained personal information concerning over 100,000,000 customers of Capital One (United State District Court for the Western District of Washington at Seattle, 2019). The indictment goes on to describe the way by which the alleged hacker identified a firewall misconfiguration with the cloud-hosted servers and exploited it to access the sensitive information. Greene & Harwell (2019) identify the cloud service provider as Amazon Web Services. Furthermore, the indictment explains how the suspect allegedly created and used scanning software which systematically identified organizations with similar misconfigurations. The indictment indicates a possible impact to more than 30 different entities, including Capital One, a regulatory agency, a telecommunications conglomerate, and a public university. While the suspect does not appear to have capitalized on or leaked the sensitive information, this instance is a clear example of how multiple, unrelated organizations all using the same cloud computing service provider can become compromised by a common misconfiguration of a standard firewall setting. In this case, Amazon insists that the cloud itself was not hacked (Greene & Harwell, 2019), and they appear to be correct, however this is still a case of vulnerability systemically stemming from use of a common cloud provider and tools in that cloud ecosystem.

Specific Mitigants on Likelihood

The remaining preventative controls from the left side of the bowtie diagram are specific to each of the identified causes. The first cause listed, intentional insider/s, is primarily controlled by process controls, separation of duties, employee screening, and

insider risk training. Process controls involve the architecture of the hypervisor in conjunction with the containerization of the services that make up AWS. This means that the system is not designed to provide a top-down view of what information is where, or what is contained in the processes that are supported by the infrastructure. The applications and services hosted by AWS are effectively black-box from an infrastructure perspective. There is no indication that an insider would be able to access customer applications or data from the hypervisor or other vantage point within the infrastructure.

Separation of duties ensures that the developer who writes code that goes into the AWS infrastructure cannot single-handedly make changes to and deploy that code without it going through a process that helps ensure no ill-effects. When checking code into a repository, there can be automated checks which evaluates the code to detect and evaluate changes and even deploy the code to test how it works. The repository can also facilitate a peer code review in which one or more software developers examine the code to understand what changes have been made and how it should function in the context of the application or service. These checks ensure that the changes of individual developers do not pass unreviewed into the application or service. In this way the duty to write and edit the code is distinct from the duty to review, approve, and deploy that code.

For sensitive positions throughout Amazon.com, Inc., background checks are to be expected (Klazema, 2018). This ensures that repeat offenders are not given insider information with which they might undermine the security of AWS. For Government regions, citizenship requirements and heightened background screenings are put in place to limit the opportunity for national rivals or enemies to receive insider access to Govcloud and related services (Amazon Web Services, Inc., 2020c).

Finally, AWS may implement insider risk training of their staff to detect and report suspicious behavior. As employees encounter suspicious behavior, even by superiors, they are able to report this kind of activity to the company's ethics hotline (Amazon.com, Inc., 2020a). The health of these controls has not been confirmed by this research and the state of this training is not likely to have a considerable impact on the likelihood of malicious compromise given the other automated and systematic controls that are in place.

The second cause identified is a hacker. The only control identified that uniquely mitigates against a hacker is that of vulnerability management. Vulnerability management refers to the process by which vulnerabilities within the software and potentially hardware configurations are identified and remediated. Identification of vulnerabilities may happen through normal internal processes, such as code reviews, or in conjunction with an external consultant to identify vulnerabilities before they are released into production. Automated code quality reviews will check code for known types of vulnerabilities. Vulnerabilities may also be identified by external parties and listed in the CVE (common vulnerabilities and exposures) which is sponsored by the U.S. Department of Homeland Security Cybersecurity and Infrastructure Security Agency (MITRE Corporation, 2019). The CVE creates a common identifiers and definitions for publicly disclosed vulnerabilities and exposures.

Once a vulnerability has been identified, then the imperative is for the owner of the software to create a version of the code that is not vulnerable to attack. This code is then released as a patch to replace the original code. This new code must still interact with other portions of the code and other services in a way that delivers the original

functionality without the vulnerability. Even with the complexity of AWS code base, the sheer scale and experience of AWS positions them well to identify, respond to, and patch vulnerabilities as they are discovered, as well as catch them in the testing phase before they become public. Given the scale of AWS, the stakes are higher in that more organizations are impacted in the event of an exploited vulnerability, but that same scale provides resources and broadly tried processes to address these vulnerabilities before they cause problems. Additionally, the containerized nature of AWS architecture ensures that if vulnerabilities do arise, they are likely to be contained within the affected service and not allow widespread access or compromise, understanding that some services, such as S3, are near ubiquitous within AWS infrastructure.

The final cause of malicious compromise here considered is that of physical access to servers or theft of data storage devices. AWS goes to great lengths to ensure the physical security of their data centers. The locations of these data centers are not published or broadly known. The facilities themselves are not identified as AWS facilities. On-site security at cloud data centers consists of multi-factor authentication at all access points, closed circuit television cameras, and intrusion detection systems (Amazon Web Services, Inc., 2020d).

Finally, virtualization ensures that even if a physical server were compromised, it is nearly impossible to predict what data or application it contains. Encryption renders physical data extraction essentially useless. A thief or hacker would need intimate and possibly real-time insider information before physical access is likely to yield a compromise of any value. A hacker with that level of insider information might not even need physical access to achieve their agenda, so the threat of physical access or theft is

minimal as concerns data loss or service usurpation. A publicized trespass of a data center may bring about the desired publicity of a malicious actor, but thanks to the redundancy of the cloud environment, it would take a concerted effort across many data centers before service disruption is likely to be achieved through physical compromise.

Impacts

The impacts of a system compromise are aligned along the right side of figure 4.1. The impacts are that the malicious actor usurps service; data loss; malicious actor disrupts service; and a publicized compromise. The first impact is the usurpation of a cloud process by a malicious actor for their own benefit. If this were to occur, the attacker could use a company's cloud-based services for its own advantage or pose as the process-owner. For example, if the service-owner is a bank, then the usurper may be able to conduct transactions and transfer assets between accounts. If the service-owner is a manufacturer or a distributor, the attacker could manipulate inventory levels to hide theft or create waste. A malicious actor may even usurp a service as a form of man-in-the-middle attack, posing as the service-owner in order to acquire data. In the example of a Bank, the usurper may obtain sensitive personal and financial information from users. In the usurpation of services, the attacker uses stealth to avoid detection, but usurpation could take place serially until caught, reconnecting temporarily in each instance.

While the usurpation of a service can lead to data being lost, data loss is treated as a unique consequence of malicious compromise. In the event of a data loss, private data is either made public, held for ransom, or used without the knowledge of the data owner. Usernames and passwords are one type of sensitive data that may allow attackers access to the same system or to other systems if passwords are reused. Social security numbers,

customer activity, and medical information are other potentially valuable types of data that may be targeted. Data loss represents an irreversible breach of privacy. Unlike forms of physical theft, even when data has been found, there is no way to truly confirm that all copies have been addressed.

The intentional disruption of service by a malicious actor is the third impact of compromise, in which an attacker causes the cloud provider or customer to be unable to operate. This may be the intent of the attack as in the case of terrorism, or it may be a side-effect of an attack with another aim. Regardless of intent, the outcome is that operations are interrupted.

The final potential impact of a malicious compromise is the compromise being publicized which could in turn have many consequences for the cloud provider and customers. Publicity may occur as the result of any malicious compromise, but it could also be a sole motivator for an activist or terrorist. In those events, even if there were no other impacts, the compromise of a secure system generates publicity, notoriety, or fear for those who rely on such a system.

Broad Mitigants on Impact

The extent of each of these impacts is mitigated by the controls displayed in figure 4.1. Five controls have a mitigating effect on all four impacts, and five guard against specific impacts. Of the five broad controls, four of them are also mitigants on likelihood. These mitigants, if unsuccessful at preventing the compromise, may still help to contain the compromise.

Encryption helps to prevent the compromise of one system from resulting in a more global compromise. Each encrypted system would require a key in order to make

the access valuable. Access controls ensure that if one user's account is compromised, then access is limited to only to those data and systems to which that user had access. Similarly, if access control systems require escalated privileges, these may be hard to achieve multiple times in succession or for long enough to retrieve the desired data or cause the desired disruption. Firewalls also contribute to the containment of a breach to the individual system or service that was breached. These three are examples of how Perrow's (2008) modularization of a complex system to manage risk. Finally, monitoring and detection will improve response-time and limit the scope of compromise.

The fifth control, which is not a repetition from the prevention side of the diagram is the containerized architecture by which the compromise of one service does not result in the compromise of multiple services. This is partly achieved by firewalls, but the conception of small standalone microservices is increasingly an attribute of many modern approaches to cloud computing and the architecture of applications that are built for the cloud.

Billing reports are a unique feature of cloud computing which may help notify customers of their services being usurped. Cloud platforms have every incentive to detect all activity and cloud customers are incentivized to reconcile their activity with the charges they receive. Illicit activity is likely to contribute costs and be somehow represented within the billing reports of cloud customers. This is a process by which a man-in-the-middle attack may be identified and stopped, thus limiting the extent of the impact.

As a safeguard against data loss, highly sensitive data can be air-gapped from the internet as in the case of the CIA's C2S. This means that if a hacker were to gain physical

access to a data center or secured network in such a way as to access and decrypt valuable data, they would have to find a way to remove the data. With no internet access, they would have to establish their own connection to the internet, to another private network, or use another transmission or storage method in order to remove, save and/or distribute the data.

Proper disposal of media storage devices ensures that data loss cannot occur by scavenging decommissioned hard-drives for sensitive data. AWS shares that their storage devices are disposed of per NIST 800-88 (Amazon Web Services, Inc., 2020d), but does not specify which of the approved methods are utilized. NIST 800-88 asserts that storage devices may be cleared, purged, or destroyed (Kissel, R., Regenscheid, A., Scholl, M., & Stine, K., 2014). AWS does ensure that this decommissioning process happens under AWS control. This means that hard drives are not removed from data centers until they have been fully decommissioned (Amazon Web Services, Inc., 2020d). AWS data center practices are tested by third-party testers to ensure that security measures are established and effective. These external auditors may "perform testing of media disposal, review security camera footage, observe entrances and hallways throughout a data center, test electronic access control devices and examine data center equipment." (Amazon Web Services, Inc., 2020d).

Service disruption, or the extent of service disruption resulting from malicious compromise is specifically controlled by backup and redundancy of services, servers, and infrastructure. Any physical attacks on data centers would have to be successfully replicated many times over before it would result in a meaningful disruption. Likewise, there are several backups of data stored on AWS so if a malicious actor attempted to

delete data, there likely exists a backup, either that the customer has created or that AWS creates in real-time. In addition to backup and redundancy, incident response plans exist to allow AWS to take corrective and protective action in the event of a compromise to limit potential disruptions to other services. These plans include intrusion detection systems that detect unauthorized data center entry.

Bowtie Analysis

The next step in a bowtie analysis is to examine the created bowtie for applicability, gaps and control effectiveness. The bowtie as constructed addresses several key causes and links to several logical impacts. One shortcoming is that the bowtie is not explicit about whether it applies to a compromise of the cloud, or in the cloud. Such a distinction is not crucial to the usefulness of the analysis, it just requires that controls be considered in the light of who owns them, the customer or AWS. For this analysis, the focus will be on AWS-owned controls. The health of these customer controls is outside the scope of this analysis unless the health of customer controls are systematically undermined.

Considering the three primary causes identified, physical access/theft seems to be the least likely, because of the controls in place and the virtual nature of the cloud. Physically proximate servers do not necessarily contain logically proximate virtual storage or virtual processes. Simply accessing the physical servers without compromising the hypervisor is unlikely to yield valuable data or result in a significant outage, if any. Encryption ensures that any data or messages intercepted through physical means is useless in the absence of the ability to decrypt it.

AWS has had no known instances of malicious compromise *of* the cloud due to hacking, and even the systematic compromises *in* the cloud were ultimately limited in scope and impact. The controls against hacking into the cloud appear to be robust and have so far been effective. The likelihood of this particular cause leading to a malicious compromise is difficult to measure and evaluate. Understanding the integrity of the entire AWS boundary is difficult and assessing hacker abilities also quite difficult. The likelihood of compromise by way of intentional insider/s is also difficult to assess given the complexity and opacity of AWS systems.

On the impact side, usurpation seems complicated and risky, and unlikely to be executed broadly across cloud users. Data loss has occurred due to customer controls failing, but not as a result of the security *of* the cloud (Panetta, 2019). Service disruptions so far have not happened as the result of a malicious compromise; however, given that unintentional insiders have caused outages of AWS systems, the most concerning causal chain contained in the bowtie runs from intentional insiders compromising AWS systems and causing a disruption of AWS services. This is the chain that warrants further research into its likelihood and the health of its controls. The next section will undertake to understand the more distant, second-order effects of a failure of AWS to deliver expected services.

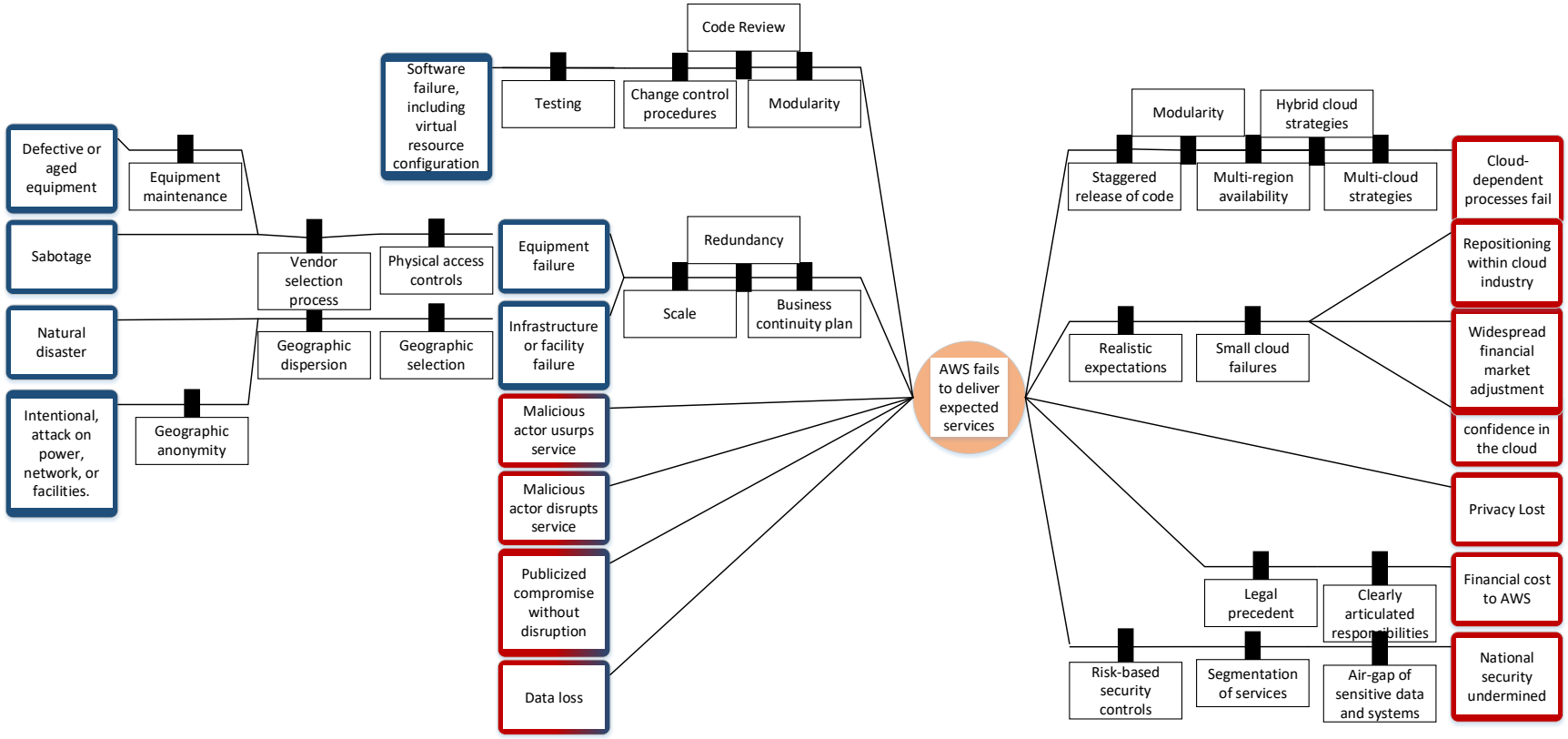
AWS Fails to Deliver Expected Services

The next risk event considered is a failure by AWS to deliver expected services. This bowtie will first examine the potential causes and impacts of service delivery failure. This risk event looks specifically at a failure *of* the cloud, rather than a failure *in* the cloud. The bowtie looking at a Malicious Compromise ended with four impacts which are

here depicted as four unmediated causes of AWS failing to deliver expected services. This means that any malicious compromise of the system represents a failure by AWS to meet the expectations of being safe and secure. This assertion is not based on service level agreements between AWS and its customers but rather the general assumption that seems to have been made by cloud customers that AWS is at least as secure as their own environments and that it highly unlikely that AWS would be compromised. This has borne out to be true so far, in that AWS has seen breaches in the cloud, but never, to public knowledge, a breach *of* the cloud (Panetta, 2019). Being unprecedented, any breach is likely to represent an expectation unmet by its customers. Since each of these causes has received considerable treatment in the previous section the analysis turns to causes which have not been addressed, as shown in figure 4.2.

Figure 4.2

Bowtie risk diagram depicting the risk of AWS failing to deliver expected services



Source: *Developed by author based on information presented in this work*

Causes and Mitigants on Likelihood

Cause 1: Software Failure

The first cause here considered is that of software failure. This cause includes code defects, configuration errors, as well as any other failures introduced through the patching and deployment process. These collective failures have been grouped because of the way by which they are introduced into the AWS environment.

Whenever new code is written or old code is changed, these changes are expected to be vetted and tested through a code review process to ensure no ill-effects. This testing is the first control that ensures that software failure does not lead to service failure. Testing usually begins with a peer code review in which an independent developer, one who did not write the code, reviews the code to identify defects before the code is implemented (Thongtanunam et al., 2015). After peer review, the code is tested in a series of increasingly complex environments. First it is usually run in isolation. Then once the code is determined to be free of internal bugs it has a chance to integrate with the broader body of code with which it interacts. Finally, its broader integration is tested within a production-like environment. Once the code passes all of these tests, it is ready to be deployed into production. Taber (2017) explains how even the release process, itself, can serve as a final test without requiring a wholesale change of the code from old to new.

Change management procedures, the second control against software failure, can include a Blue/Green deployment. A Blue/Green deployment is one in which the blue environment represents the old, proven code, and the green represents the change (Taber,

2017). Two complete environments are created, one blue and one green and they stand ready to process requests simultaneously and interchangibly. The Domain Name Service is then used to gradually funnel traffic from blue to green, ensuring that green is functioning properly before the floodgates of production volume are opened all the way. If green does not perform as expected, the DNS can reroute traffic away from green while the issue is resolved. This ensures that only a few customers will be subjected to the defects that were leaked into the green environment and may even cause them, upon refreshing the page, to be redirected to a working, blue service. In this way defects introduced into the green environment reach only a limited number of customers.

Change management also takes into consideration controls and procedures that keep developers from deploying code straight into production. This piece of change management includes the separation of duties described in the previous section. These procedures outline exactly what kind of testing is required, by whom, and often requires a third-party code review to ensure that no harmful code makes it into the production code-base.

The final mitigating factor that helps prevent software failure from leading to service failure is modularity. Modularity is not a term frequently used by AWS or software development in general, but is a term used by Perrow (2007, 2008) to describe how a complex system is broken down into smaller and potentially redundant parts so that failures do not become global to the system. In software development, the term microservices is often used to convey a close approximation to this concept. Microservices describes the creation of a complex system by stringing together smaller self-contained services that interact with one another to deliver a complex range of

functionality. AWS modularizes their microservices further by dividing the underlying services into cells and then keeping those cells of a manageable size (Amazon Web Services, Inc., 2020e). This partitioning of their services ensures a faster recovery time if services need to be restarted. Smaller cells restart faster than larger cells.

Cause 2: Equipment Failure

The next cause of service failure in this bowtie diagram is equipment failure. This cause is preceded by two root causes, defective equipment and sabotage. Equipment could fail for other reasons, like power surges or a kind of attack, but those external causes will be considered in the context of the cause entitled, infrastructure and facility.

Defective equipment refers to a breakdown in the functionality of the servers, routers, and switches that comprise the computing and networking portion of the data center. The rest of the supporting equipment will be considered as part of the infrastructure. The computing equipment is treated separately for two reasons, first its direct proximity to the sensitive data and critical processes of the cloud, and second, its volume in the data center. A given data center can have tens of thousands of servers, potentially all using identical hardware from a limited number of suppliers.

In the event of defective equipment, the ubiquity of a particular server or server component could mean that if all were installed at about the same time they would be expected to fail around the same timeframe. This would become even riskier if multiple data centers implemented identical technology along a similar timeframe. This is one, albeit unlikely, way that the scale of AWS could result in systemic issues that would not be present in a regime in which independent organizations stand up and scale up a variety of heterogeneous data centers. This risk can be managed by tracking equipment

maintenance schedules to understand their expected lifespan. This would allow AWS to anticipate needed replacements, even if replacement schedules overlap. Monitoring of hardware also make this outlier possibility even less likely, as AWS should be able to see, in real-time, the state of their hardware investment.

Sabotage is another source of equipment failure. This is distinguished from malicious compromise based on the assumption that this sabotage does not necessarily require access to the system. Instead this cause represents a targeted undermining of the hardware of the system. This type of sabotage would take place through predictable or programmatic failure of the equipment with the intent of undermining AWS services. AWS can mitigate against this risk through careful vendor selection.

Sabotage could take place through an undermining of the equipment during installation. In addition to a vendor selection process that controls the type of worker given access to AWS data centers during the installation process, monitoring the installation process can also serve to limit the opportunities vendors and their employees have to attack AWS operations. There is not a lot of public information to describe the processes and security by which new data centers are assembled. It is not clear whether the installation processes create significant opportunities for sabotage. However, once assembled, the security measures required for physical access to the equipment make it unlikely for this type of after-installation sabotage to take place. Furthermore, this type of sabotage would take considerable planning and coordination in order to affect enough equipment at enough data centers to have any noticeable impact on AWS operations.

Even if equipment fails, and fails systematically, the two safeguards that prevent that failure from disrupting AWS services are scale and redundancy. First, redundancy

ensures that data is backed up and computing can fail over to other data centers, availability zones, and potentially regions in the event of widespread equipment failure. This means that anything short of an extremely well-orchestrated and coordinated attack across numerous data centers, zones and regions is unlikely to actually disrupt services. Furthermore, AWS continues to expand regions, availability zones, and data centers, which means that its scale and potential redundancy is only increasing. So, redundancy requires impact across a variety of locations and scale continues to increase that number of locations. Finally, AWS can be expected to have incident response plans that would help it respond quickly to mitigate the impact of equipment failing in quick succession.

Cause 3: Infrastructure or Facility Failure

The final cause also has two root causes that feed into it. Infrastructure or facility failure represents the situation in which data centers or the networks connecting them fail to support the services of AWS. The two root causes of this type of failure are natural disaster, and intentional physical attack. These two root causes could cause infrastructure or facility failure by causing physical damage to a data center, the internet or network connections that provide access to and from the data centers, or the electrical power on which these data centers depend. A natural disaster could take out an availability zone and potentially a region if it were a large enough event. Natural disasters appear to be well guarded against thanks to a wide geographic dispersion and geographic selection. Dispersion ensures that it would take an extremely widespread natural disaster to significantly disrupt service across regions. Selection makes this even less likely by

positioning data centers away from high risk locations and areas prone to natural disasters.

Intentional attacks on power, network, or facilities assumes a coordinated effort directed against AWS, a customer of AWS, or the common infrastructure on which AWS depends. The perpetrator could be a terrorist or foreign government. Geographic anonymity presents a strong defense against this type of attack against AWS infrastructure. Geographic anonymity describes the level of secrecy AWS adopts with regard to its data centers. AWS is secretive about the location of its data centers, and will not divulge information upon request (Burrington, 2016). While this secrecy is likely sufficient to deter small-scale terrorists, this does not seem like an insurmountable barrier to a well-funded terrorist group or a state-sponsored attack. As with a natural disaster, geographic selection places data centers primarily in inland regions. In the United States, one of these is in relatively close proximity to sites of strategic importance in and around Northern Virginia (Burrington, 2016). These sites of importance include the Pentagon, the headquarters of the CIA, the FBI, and Washington D.C. more holistically. Geographic dispersion serves as a significant hurdle to overcome and therefore a deterrent to such an attack.

Infrastructure failure is mitigated by similar attributes as equipment failure. Geographic redundancy and a business continuity plan make it exceedingly difficult for an attacker to disrupt AWS services by undermining the infrastructure, network or facilities. Scale plays into geographic redundancy and is not as extensive as the tens of thousands of servers that comprise a data center, but the sheer number of data centers and

degree to which they are expanding creates a target that requires significant coordination in order to attack.

The power grid is another vector by which AWS may have operations interrupted. AWS is likely to have back-up arrangements in case of power outages, but a widespread and long-lasting disruption to power may have devastating impacts on the ability of AWS to meet the expectations of its customers. This kind of power outage is beyond the ability of AWS to mitigate against. The scale of AWS and cloud computing more generally, reveals an increasing dependence on the power grid. Tellingly, the power grid likely has less redundancy than the AWS facilities. A potential mitigant against the reliance of AWS on the broader power grid is the development of 31 utility-scale wind and solar renewable energy projects, globally (Amazon.com, Inc., 2020b). These projects have been undertaken in part to reach Amazon's goals as part of the Climate Pledge but may have an additional benefit of providing power proximate to AWS facilities that might contribute to resiliency in the event of an outage of the power grid.

Impacts and Mitigants on Impact

Having explored the various causes of and mitigants against AWS failing to deliver expected services, the bowtie analysis examines the impacts of such a failure. The first, and most significant impact is that cloud-dependent processes fail. If AWS fails to deliver the expected services, then the organizations that depend on those services must have some sort of backup for those dependent processes or fail to deliver their own services. As exemplified in the previous case studies, these organizations include digital entertainment, transportation, manufacturing, banking and even national security. Before

examining the fallout of cloud processes failing, the mititgants on this impact are addressed.

One mitigant AWS could make to minimize process failure is to stagger the releases of their code. By performing blue/green deployments where possible (Taber, 2017), AWS might be able to assess the customer impact of their code changes before all customers are impacted. Modularity also mitigates the impact in the event that AWS fails to deliver the expected services by potentially reducing the time it takes to recover. As larger processes are broken down into smaller ones, they are expected to be able to be restarted and come back online sooner. Likewise, modularization may also help to contain code defects within a smaller segment of the cloud infrastructure and limit its impact. These are the main mitigants on impact that are available to AWS. The remaining mitigants to this impact are the responsibility of the cloud customers themselves to establish continuity plans against an outage of AWS. There is an issue of moral hazard in the success of AWS to stave off a significant failure. The more stable it is perceived to be, the less its customers are likely to invest in contingency arrangements.

Multi-region availability is an optional configuration for AWS customers. It costs more to have redundancy between regions, and it adds complexity to the organization's configuration but it provides more reliability. It is unclear how many customers choose not to invest in additional redundancy, but in the February 28, 2017 outage of the Northern Virginia (US-EAST-1) region, it was demonstrated how few organizations had opted for redundancy (Hersher, R., 2017). Amazon, itself, had actually failed to include regional redundancy for its AWS Service Health Dashboard. As a result, AWS was unable to communicate the outage through the channel dedicated to that communication

because that specific channel lacked redundancy (AWS, 2020 Summary) Amazon has since reported to have added redundancy.

Hybrid cloud strategies are a major mitigant against the type of short-term outages that have been typical for AWS. As explained in the Volkswagen case study, hybrid cloud strategies involve on-premise servers used in conjunction with cloud computing so that an organization's local processes are not dependent on cloud connectivity to continue. Hybrid cloud strategies are more expensive to execute and add significant complexity to the logistics of running their information services that are pure cloud-based solutions but they are necessary to ensuring a minimal disruption of cloud-dependent processes.

Finally, multi-cloud strategies involve the use of more than one cloud provider. By having redundancy between two different clouds, this limits the impact in the case of a cloud outage affecting a single cloud provider. By selecting at least two providers and engineering failover processes between them, an organization can mitigate against company-specific risks in the cloud. As explained by Gartner in Panetta (2019), it is expensive and difficult for organizations to take this precaution.

The next three impacts worth considering are repositioning within the cloud industry, widespread financial market adjustment, and lost confidence in the cloud. The first of these three impacts considered is the repositioning within the cloud industry. In the event that AWS fails to deliver expected services, AWS customers may reconsider their use of AWS. They may also reconsider a multi-cloud or hybrid strategy to make their processes more robust to AWS outages. Prospective cloud customers may also view AWS as less safe and opt to use a competitor. The shifting of customers between cloud

service providers would be time consuming and costly for those customers. Those opting for a multi-cloud strategy can expect to more than double their spending on services, utilizing services in two different cloud providers as well as the expenses related to integrating the two platforms. The same impact can be expected for any who would opt for a hybrid on premise strategy, within AWS or between AWS and their own data centers. A hybrid strategy utilizing a company's own data centers deeply undermines the efficiency and simplicity gained by utilizing cloud service.

The impact of lost confidence in the cloud more broadly would have wide-reaching impacts on the cloud industry as well as the many industries that have come to depend on cloud computing services. In order to result in lost confidence in the cloud, the extent and duration of an AWS failure would have to be substantial. Such an outage is without precedent. In this scenario, companies that have never managed their own data centers may consider investing in them, while others that have transitioned away from owning and operating their own data centers might consider re-establishing their on-premise computing capabilities.

Given the under-utilization of servers which is typical for on-premise computing, a return to on-premise computing will cause a significant increase in the demand for hardware, software, facilities, and the personnel necessary to establish and run them. The environmental impact and energy usage would be significant. Based on historic responses to AWS outages, it would take an outage that was not only widespread and long-lasting, but one which reveals an insurmountable deficiency in the cloud model more generally, such that technological adjustments are not readily able to remedy the issue. One potential source of this kind of outage is one related to the lack of verifiability in the

cloud computing space which will be addressed in detail later in this section. A widespread loss of confidence in the cloud could occur as a reduction in trust and expectations, rather than a wholesale shift away from the technology.

Widespread financial market adjustment will inevitably occur if every company that depends on the cloud decides to pursue other computing solutions for its cloud-dependent business. If the cloud model itself is brought into question, then even the stock value of competitors could be impacted by an AWS outage. In this case, the value of AWS stock is expected to decline, along with the stock-value of its competitors based on the breadth and duration of the outage. Some competitor valuations may improve, if their operating model is significantly immune to the concerns raised by the outage. If the outage is caused by a terrorist attack or cyber attack by a foreign government, then the market impact is likely to be even more severe and more prolonged. Just as the airline industry was affected after the September 11 attack in New York, the cloud computing industry, being the vector of the attack, can be expected to contract for a time. This hit would create uncertainty for those dependent on the cloud as well.

Finally, if companies choose to adjust their cloud strategies to incorporate hybrid and multi-cloud strategies, then the IT expenditure by those companies can be expected to increase significantly. If this increase in spend is widespread, the companies may reduce discretionary spending on other services or potentially cut back on future ventures. It is possible for this to lead to bearish expectations and potentially a widespread recession. This type of outcome could unfold in a manner similar to the .com crash of the early 2000s.

Potentially mitigating against these impacts are two mitigating factors: realistic expectations, and small cloud failures. Realistic expectations ensure that customers and stakeholders are not surprised by outages. AWS offers no real assurances, or guaranteed service levels. While AWS touts a strong record for reliable service, it is not even contractually required to maintain that record. Given noncommittal approach of AWS, it can be inferred that the company is aware of its inability to fend off outages and has intentionally avoided committing otherwise. Despite this lack of commitment from AWS, few companies have adopted a multi-cloud approach, according to Gartner's acknowledgement, cited in Weise (2017). It can be inferred that many AWS customers consider AWS reliable enough to single-threadedly handle key strategic services. This may imply a mismatch between the expectations of AWS and its customers, but that would be difficult to conclusively prove. If, however, AWS were successful at instilling realistic expectations in their customers and, as relevant, the stakeholders of their customers, then a failure to deliver services is less likely to cause a significant repositioning in the cloud industry, loss of confidence in the cloud, or widespread financial market adjustment. If the stock market is aware of the realistic potential for failure, then that failure should be priced into stocks in such a way as to reduce, not eliminate, the market impact of an AWS failure. The stock marker may become aware of the potential for failure through small, infrequent failures of AWS systems. In this way small failures serve to align the expectations of the customers and stakeholders of AWS with the constant possibility for complex systems to fail (Perrow, 2011).

The next impact of an AWS failure is that of privacy lost. In the event that AWS fails to deliver the security expected by customers of the cloud, and unauthorized users

come into possession of data hosted by AWS, then the privacy of the data owners cannot be completely restored. Additional impacts, including identity theft, and extortion, may have remedies, financial and otherwise, that can help minimize the impact to victims. However, no remedy can actually restore or ensure the privacy of the data once leaked. Privacy loss is also a means by which other impacts are realized. Privacy loss could cause significant repositioning within the cloud industry, loss of confidence in the cloud, significant financial market adjustments, lawsuits and financial losses to AWS, and potentially an impact on national security.

In the event that AWS fails to deliver the expected services, it could potentially face financial costs and penalties. Financial costs could come in the form of repair costs, financial damages, legal fees, fines, lost customers, lost revenue, over-compensated controls, lost economies of scale, and stock devaluation limiting access to capital. The first four types of financial costs relate most closely to the failure itself. They represent the costs incurred to restore service and address damages to customers. Given the overall lack of legal precedent concerning issues related to cloud computing, there is some uncertainty as to how a significant outage may be handled from a legal perspective. Additionally, AWS provides no guarantees beyond a best-effort restoration of service.

The loss of customers and revenue are secondary effects of an AWS failure and would likely occur with a delay, and most likely a significant delay. For an organization to cease using AWS, it would have to be able to revert to an on-premise solution or lift and shift their infrastructure to another cloud provider. This could take months and potentially years for some companies, especially those who depend on services that are proprietary or exclusive to AWS. Over-compensated controls refer to the potential for an

extreme outlier event to result in heightened controls, the cost of which could outweigh the expected benefits. This type of over-correction could be motivated by an effort to restore confidence in AWS in order to retain customers and/or reassure investors. Any decrease in customer base or the growth rate of the customer base is likely to have an impact on the economies of scale that enables AWS to offer competitive services at a competitive price. If the customer base or projected customer base changes materially, then AWS may need to adjust its operating model to scale down to optimize the new normal.

Finally, an inability to reassure investors could lead to a significant decrease in the company's valuation which could impact its access to additional capital. If the company sees an increase in costs and a decrease in revenues related to a crisis, the accompanying hit to the stock price may limit the company's access to capital when it needs it most. This myriad of financial costs are contained within AWS although the ripple effects on AWS employment and potentially Amazon.com, Inc. employment could have significant impacts on the local economies in which they operate. Furthermore, financial impacts to AWS are likely to have a spillover effect on the cloud industry and the broader financial market. Ultimately, the uncertainty of the legal impact of a major cloud outage becomes uncertainty around the financial impact on AWS and Amazon.com Inc. and this uncertainty is likely to have real-world implications, including an impact on the broader financial markets.

The two controls on these financial impacts are legal precedent, and clearly articulated responsibilities. Legal precedent refers to court cases between cloud providers and cloud users. As these court cases occur, they help AWS and other cloud providers to

understand how damages might be awarded in the case of a cloud provider failing to provide the expected service, including the loss of privacy. While privacy loss cases are happening more frequently, there is still more experience needed to understand the extent of liability that a cloud provider like AWS might be exposed to in the case of a failure.

The second control on the aforementioned financial impacts is clearly articulated responsibilities. On their website, AWS provides Service Level Agreements (SLA's) for each of its services (Amazon Web Services, Inc., 2020b). These documents outline the available uptime or successful response rate that customers can expect from AWS for each service provided. These SLAs outline the financial remuneration each customer is entitled to in the event that AWS fails to achieve the agreed-upon service level. For each of the 110 AWS services, listed as of February 29, 2020, each of the SLAs promise only service credits based upon the monthly uptime percentage. See table 4.2 for an example from the current SLA (as of February 29, 2020) for Amazon Inspector. These service credits can offset charges owed to AWS for services, but are only paid out as a refund at the discretion of AWS. This implies that the risk of financial loss to AWS in the event of an outage only applies to future revenues or litigation that awards damages exceeding those outlined in the SLA.

Table 4.2

Excerpt from AWS Service Level Agreement for Amazon Inspector (retrieved February 29, 2020), Detailing Service Credits Owed as Determined by Monthly Uptime Percentage

Monthly Uptime Percentage	Service Credit Percentage
Less than 99.9% but equal to or greater than 99.0%	10%
Less than 99.0% but equal to or greater than 95.0%	25%
Less than 95.0%	100%

(Amazon Web Services, Inc., 2020a)

For customers to receive their credit for AWS downtime, the customer needs to document the specific dates, times, and services that have failed to live up to the SLA. Part of this documentation is providing the request logs for a given service. Unless a customer is using a hybrid or multi-cloud strategy, these request logs are stored in an AWS S3 bucket. This S3 bucket is stored in the cloud and the storage of these logs is charged like any other data. This process lacks verifiability from a customer perspective. The customer must retrieve data from an AWS service to support a claim against that same service. While there is no evidence to suspect that these logs have ever been tampered with, the customer still has no real assurances that the logs in question have not been altered. Newer technologies, including the use of blockchain technologies, may help create a verifiable, immutable record of logs. Until then, customers are at the mercy of AWS services to inform the customer when AWS services are not functioning as expected. A similar catch-22 occurred with the S3 outage in 2017, during which the AWS

dashboard that informed the customer of outages in real-time, displayed no outage because of an outage in the S3 storage space. Failures in these types of tightly-coupled, complex processes can be expected to recur as the wide breadth of AWS services continues to increase.

The final impact considered for this event is the undermining of national security. This impact refers primarily to the services AWS provides to the CIA through C2S and to the broader U.S. Government through GovCloud. Those government processes which depend on AWS services are vulnerable to attacks on AWS by enemies of the state. There is also a sense in which the smooth functioning of the U.S. economy contributes to the security of the United States and any significant impact on the broader economy could weaken the security of the United States from a strategic perspective. The financial impacts articulated previously demonstrate that an outage could have significant economic impact resulting in a reduced comparative standing of the United States on the world stage. Further risk to national security consists of the loss of sensitive government information, critical process disruption, and broader economic damage culminating in the undermining of the broader economy.

The impacts on national security through the economy are primarily mitigated by the controls on the failure of cloud-dependent processes and the financial impacts that were addressed first in this section. Guarding specifically against national security are those controls which govern the government-specific regions of AWS, C2S and Govcloud. Risk-based security controls describes the different tiers of security and their application to different classification of data and processes in Govcloud. This ensures that

the most sensitive information and processes are protected more heavily than those that are less sensitive.

Next, the segmentation of services between the public AWS cloud and the government regions requires that attackers specifically target government resources in order to undermine government processes and also ensures that the appropriate preventative measures can be concentrated on those data centers, availability zones and regions that support sensitive government processes.

Finally, for the most sensitive government intelligence, held in C2S, an air-gap of the system from the internet ensures that attackers would have to compromise the physical security of the CIA or an equivalent agency in order to even access the information. The air-gap makes it especially difficult for a weakness in the AWS infrastructure to undermine the security of the United States.

Having laid out the potential causes, and impacts of an AWS failure, and analyzed the controls in place to reduce the likelihood and impact of that event, the next step in the bowtie analysis is to evaluate the causal chains to identify where controls are missing or insufficient. In the case of AWS failing to deliver expected services, the scale and redundancy within AWS appear to be effective at preventing a service disruption caused by equipment or infrastructure failure. In the case of infrastructure or facility failure, geographic selection and dispersion also appear to be effective controls against natural disaster and small-scale attacks on infrastructure. What remains relatively unmitigated is a large-scale attack on the infrastructure on which AWS relies for its data centers. However, given the geographic dispersion of AWS data centers it would require significant coordination and resources to ensure a widespread outage through this means.

Foreign-government sponsored attacks seem most capable of coordinating such a large-scale attack and that only within the context of war. Additionally, it seems unlikely that consolidation in the cloud computing industry materially adds to the losses or undermines system resilience in the case of a widespread debilitation of the underlying infrastructure such as internet or electricity. Such a wide-spread attack would inherently cause damage of similar and greater magnitudes to other companies and industries; disruption through a lack of cloud computing resources would be just one of many debilitations. The implications of cloud consolidation in the event of widespread infrastructure outage in the context of warfare would only be one small and likely insignificant component of the wide sweeping disruption.

Software failure, while inevitable, has historically resulted in only a few, short, outages. While these outages have had considerable impacts on the affected organizations, there have not been widespread financial market adjustment, major repositioning in the cloud industry, a discernable loss of confidence in the cloud, lost privacy, significant financial hardship for AWS or an undermining of national security.

This leaves only those causes emanating from a malicious compromise. Of the three causes of malicious compromise, intentional insiders and hackers appear the most likely threat. Intentional insiders would have a unique knowledge of the system and points of weakness. It seems that service disruption is the most significant outcome that an insider would be likely to achieve, due to the number of AWS access controls, monitoring, encryption and firewalls. The ongoing usurpation of service seems unlikely, and the controls against data loss may be sufficient to deter attempts by insiders to divulge sensitive information.

Service disruption could have an immediate and significant impact on AWS services and the organizations which depend on them. One significant source insider risk appears to be from the threat of strike or sabotage by AWS workers. If AWS workers choose to cause an outage or to stand-down in a crisis, it would be very difficult for AWS to replace existing staff with resources who have the skills and experience to step in and trouble-shoot an unfamiliar system. A service disruption caused by strike or sabotage seems capable of effecting all of the outlined outcomes, except perhaps the undermining of national security. While the intelligence community depends on AWS to house its information, a temporary lack of staff members to service those systems seems insufficient to undermine national security. A prolonged lapse in service could have a more significant impact.

A hacker seems to be the only other plausible source of disruption to AWS services. Of the four impact-causes bridging the two bowties together, a compromise with no ill-effect is not expected to yield impacts to the broader economy, and so the remaining impact-paths worth further consideration are service usurpation, service disruption and data loss. Long-term usurpation of the service seems a tall order given the firewalls, encryption and monitoring of AWS services. A temporary usurpation of the service with the intent of damaging real-world equipment could be possible, especially given the growth of internet of things and cloud implementation in manufacturing contexts, such as in Volkswagen. In an attack like this, the hacker would take control of the cloud processes with the intent of disrupting physical processes and causing damage. This approach would appear similar to the Stuxnet attack on the Natanz Iranian uranium enrichment facility. In this attack a sophisticated piece of malware was used to damage

over a thousand centrifuges a the Natanz facility partly by sending commands to change rotation speeds, as well as open and close valves to cause chronic failure of the equipment (Lindsay, 2013). This same approach could be used to cause catastrophic failure of equipment by sending similarly destructive commands. Given the possibility for a compromise to quickly cause irreversible real-world damage, usurpation is a significant cause.

In contrast to the sending of fraudulent messages causing real-world damage, service interruption targets the cloud-based processes themselves with the intent of disruption. If hackers do succeed in causing a widespread service disruption, the system would have to be compromised in such a way as not to be easily recoverable. Given the length and impact of unintentional outages, it could be presumed that an intentional outage would last longer and have a wider impact. It is not clear how recoverable the cloud infrastructure would be to a targeted attack, and it is also not clear how easy it would be to attack in a way that prevent fast recovery. Based on recent unintentional outages, recovery times may be quick enough to stave of significant repercussions.

Hackers accessing sensitive data is a likely scenario, but recent data loss events have demonstrated that companies are resilient to these types of events. Even secret government documents have been leaked in recent years without catastrophic implications. The sheer volume of data housed by AWS would make it a prime target for this type of attack, but the sheer volume of data also creates a difficulty in locating and extracting specific data.

Chapter IV Conclusion

Applying the bowtie method to the risk posed to and by AWS yields a clear articulation of the potential causes and impacts of two significant and related risk events. Furthermore, it highlights instances of systemic risk, in which the size of AWS can be expected to have both direct and inverse relationships to different aspects of risk. Through further research and analysis it could be assessed which effect dominates and whether or not the scale of AWS represents an asset or liability to the broader economy.

Insider risk remains the most plausible cause of a malicious compromise, and that with the plausible impact of a service disruption. This risk is expected to increase with scale, as more insiders are needed to support a widening breadth of services. Furthermore, the longer the cloud infrastructure exists the more insiders there are with intimate knowledge of how it functions. Insider risk may prove difficult to study in depth, due to issues of opacity and a lack of verifiability. It is incumbent on AWS to reduce the likelihood and impact of insider risk in this context, and would behoove the general public to be able to ensure that these controls are carried out effectively.

Any compromise would necessarily lead to AWS failing to deliver the expected services which could have wide-reaching impact across the economy. This could cause widespread outages of cloud-based processes, impacting the performance of a wide variety of firms in a variety of industries. The cloud industry itself could face significant and costly repositioning, as companies may opt to lift and shift away from AWS or adopt a more costly multi-cloud strategy. If confidence in the cloud is lost, then the economy would face considerable financial and environmental costs of shifting to hybrid or on-premise data centers. This kind of repositioning could have dramatic effects on financial

markets. AWS has survived many accidental outages, it is unclear how AWS would be affected as a going concern in the event that that an outage is intentional and/or lasts longer than previous accidental outages.

Government and military dependence on AWS means that an undermining of AWS poses a threat to national security. Furthermore, this tie makes AWS a target for enemies of the state. In this way, Government embrace of cloud computing (particularly implementations of AWS and its close competitors) increases the size of the cloud, contributes to consolidation in the cloud computing industry, and links the fates of private and public enterprise. This raises the stakes for a cloud outage, increasing both the likelihood an impact of a malicious compromise leading to a failure to deliver cloud services. By aligning enemies of the state with others who would profit from cloud disruption, likelihood increases as more malicious actors can concentrate on one platform. That one platform supports both public and private sectors.

Beyond the general conclusions yielded from this analysis, this qualitative analysis organizes the risk landscape into discrete topics for research, articulated through causes, controls, and impacts. Malicious compromise is traced back to three key causes, each of which warrant further study. Insider risk, itself, would require an interdisciplinary approach to evaluation, as it touches on issues of technology, organizational structure, corporate governance, sociology, criminology, geopolitics. Issues of privacy from insiders, as well as the verifiability of privacy warrants further research.

The vulnerability of hypervisors to hacking is an area that warrants additional research given that all cloud infrastructure is dependent on some form of hypervisor. Given the wide breadth of applications, just on AWS, much rides on hypervisor security.

Issues of privacy should also be considered. Is there back-door access of any kind build into AWS hypervisors, and how could insiders, governments, or hackers exploit this kind of access.

From an impact perspective, the usurpation of cloud services has not made headlines, especially not through a compromise *of* the cloud. A potential area of study would be to examine whether or not billing reports have been used to identify compromises *in* the cloud. This question has both economic, information security, and business implications.

The financial implications of a cloud outage could be assessed in light of the growing breadth of AWS. The role of AWS and Amazon.com, Inc in the broader economy should be considered and the risk that AWS might pose to Amazon.com, both in a parent-child company context, as well as the fact that Amazon.com runs on AWS infrastructure.

Other areas of research could be derived from this bowtie analysis. There are a variety of controls that in this context were taken for granted but could be verified and further assessed as to their effectiveness. Furthermore, real incidents will inform the analysis as they are compared with the bowtie to assess the performance of these controls. Time will reveal much, but it is in the best interest of the general public and the broader economy to reduce the likelihood of a significant outage of the cloud. The best way to mitigate the breadth of impact is to limit the breadth of any single cloud component.

CHAPTER V

A NORMATIVE SYSTEMS ANALYSIS OF AWS

Amazon Web Services has become a critical computing infrastructure within the United States and, increasingly, world-wide. The natural monopolistic tendencies established in Chapter 3 have caused AWS to act as a public utility of sorts. The AWS-hosted clouds contain private and government information unlike any other. AWS is a government and military contractor on a considerable scale. Having become critical infrastructure, AWS has also become too big to fail as a critical service provider. The number of firms dependent on AWS and the degree of dependence appears only to be increasing. Among these dependent organizations, as was shown in Chapter 3, are competitors of AWS and Amazon.com, Inc. AWS as a standalone entity is a firm unlike any other, but when considered in the context of its parent company, the largest online retailer in the World, it becomes clear that there is no coherent framework for understanding much less regulating it and prevent it from introducing major systemic risks into the broader economic system.

Hayden (2008) contends that finding the gaps, discontinuity, disharmony, and conflicts within normative systems is of paramount concern, given the fragility of the modern world. This chapter will apply Hayden's framework as a normative systems analysis of instituted processes method (hereafter, NSAIM), in this case aimed at organizing the normative duties and obligations concerning Amazon Web Services.

Normative System Dynamics of AWS

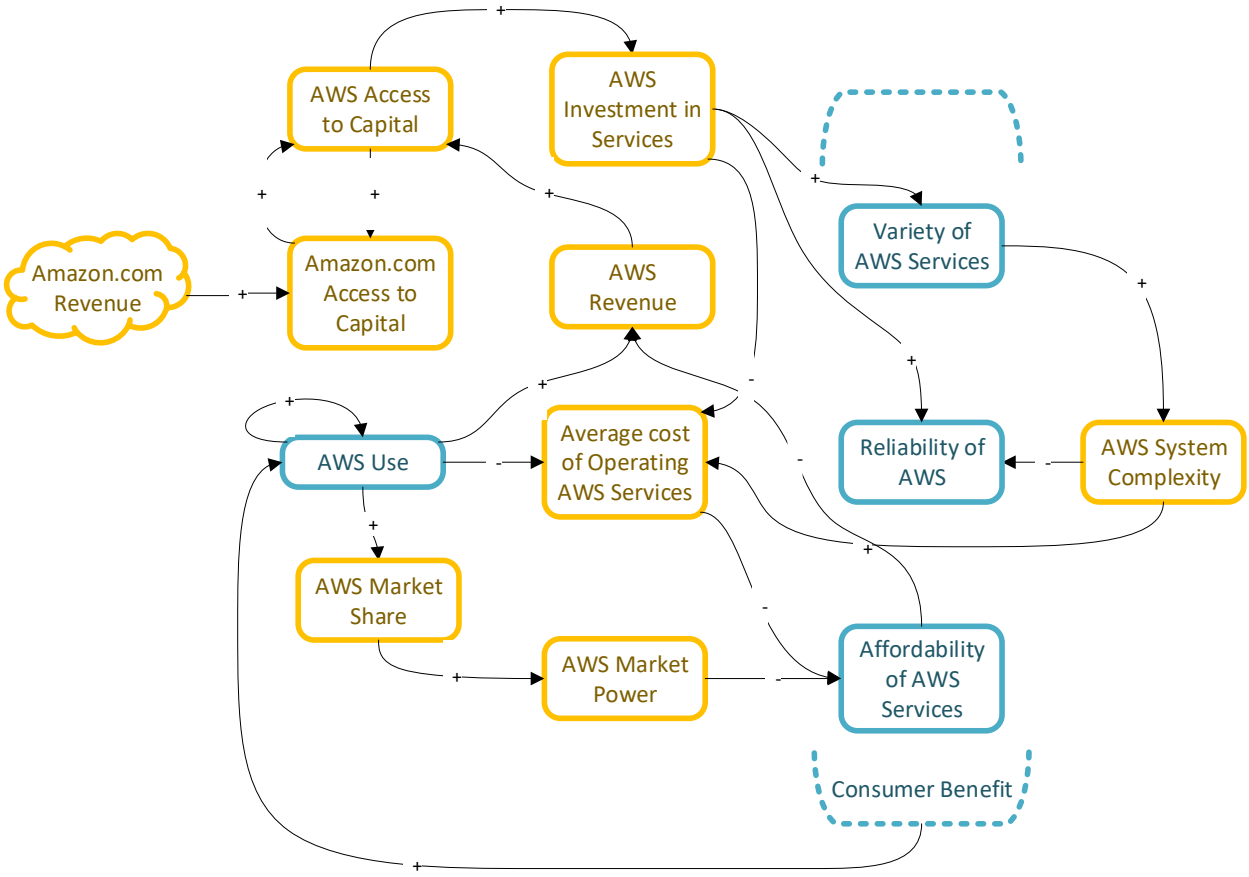
This section begins with an explanation of the NSAIM to depicting a system's dynamics. This approach is used to diagram and explain the high-level dynamics of the

business system in which AWS operates. This business system will depict how the use of AWS leads to increased investment in AWS services, which increases consumer benefit and further reinforces the use of AWS. In this system, AWS use is limited by increasing complexity. The next section will outline the ways that AWS usage contributes to increased oversight, regulation, and public influence on AWS through the authorizing institutions. The following section will overlay the normative beliefs guiding those authorizing institutions. The fifth and final section provides a framework for understanding how intervention and influence by the authorizing institutions can be expected to influence the business dynamics of AWS in the cloud computing industry

The analysis in this chapter thereby provides a disciplined approach to analyzing complexity. The diagrammatic approach adopted here closely follows the tradition of Jay Forrester as laid out by Sterman (2000). In this approach, causal relationships between variables are drawn, connected with arrows indicating either a positive or negative relationship (varying directly, or inversely, respectively). These relationships assume *ceteris paribus* concerning the other variables in the system, depicted or not. Key insights come from establishing feedback loops. These loops depict the effect of a variable upon itself. The effects will either be balancing effects, where an increase in the variable limits future increases or leads to decline; or the effects will be reinforcing, where an increase in a variable leads to additional increases in that variable. Both loops may be present and change in dominance as the system changes over time.

Figure 5.1

Busines Dynamics of Amazon Web Services (AWS), Depicting the Reinforcing Feedback Loops Through Which AWS Use Leads to More AWS Use.



Source: *Developed by author based on information presented in this work*

Figure 5.1 depicts a system diagram for AWS use, complexity, and access to capital. To explain this system, each variable will be explained in succession in the context of an increase in AWS use, since this is the dominating pattern seen so far in AWS history. Beginning with the *AWS Use* variable, we see this variable linked in a

reinforcing circular route with itself. This immediate, reinforcing feedback with itself is due to path dependence and network effects which acknowledges the path-dependent nature of AWS services as a platform. There is an entire other system that could be drawn in detail to justify and explain this platform path-dependence, but for a succinct justification, consider the case of Amazon.com, Inc. making use of AWS as its cloud provider. Having built its information systems on AWS it is likely to continue using that framework for additions to its information systems. Also captured in this feedback use is a QWERTY-style path dependence of workers being trained to use AWS in one setting and being more likely to reuse it in other settings. Conversely, if Amazon or other significant users begin to decrease their use of AWS, that is a decision that is likely to be repeated by other users.

Increased AWS Use also leads to increased AWS revenue, assuming of course that AWS charges for use. Increased revenue leads to an increase in the ability of AWS to access capital. This relationship takes place through two means: profit and financing. To the extent that AWS costs are less than it charges (and reported earnings have shown this to be the case), increased revenue will increase profits and directly provide capital for further investment. Even apart from profitability, cash flows can be borrowed against, or attract additional equity financing by providing information about the viability of a business activity.

AWS's access to capital can be a source of access to capital for Amazon.com Inc. If that access to capital funds investment that results in increased Amazon.com revenues, then there exists a potential reinforcing loop between AWS's access to capital and Amazon.com revenues, as depicted by the dashed line. The reinforcing nature of this loop

is less important for the present analysis than the ability for Amazon.com, Inc. to feed into the AWS system by providing the capital through which AWS can invest in its services. The converse also introduces a reinforcing risk to the system, in the event that Amazon.com needs or decides to withhold or withdraw capital from AWS.

If AWS has access to capital either through retained earnings or financing and investing activities, then it can make additional investment in its services. This investment can be expected to increase the variety of services provided by AWS, increase the reliability of AWS systems, and/or reduce the average cost of operating AWS services. A decrease in the average cost of AWS Services should increase the affordability of AWS services. Variety, reliability, and affordability are combined to represent the consumer benefit derived by AWS investment in its services. Each of these three variables increase with increased investment in services and they can be expected to individually or collectively encourage additional use of AWS. This closes the second reinforcing loop on AWS use.

AWS investment in services also leads to a balancing loop. This loop follows an investment in services through increased variety of AWS services to AWS complexity (in figure 5.1). As AWS increases the variety of its services, their various interactions increase the complexity of the AWS service offering. As the systems become more complex, reliability decreases, and average operating costs increase for AWS. This increase in costs decreases affordability. In this way, system complexity reduces both the reliability and affordability of AWS. It is not clear whether these decreases offset the consumer benefit of increased variety. Complexity, however, can be expected to increase exponentially with variety to the extent that the variety of services offered interact with

each other. This type of exponential relationship between variety and complexity makes it likely that at some point the system encounters diminished marginal benefit from variety.

Revisiting the variable of average cost of operating AWS services, this variable is inversely related to AWS access to capital. As the average cost of operating AWS services decreases, profitability and therefore access to capital can be expected to increase further. This is a second reinforcing aspect of investing in AWS services. Additionally, as AWS use increases, the presence of fixed costs means that average costs continue to decline with use. This accounts for the negative relationship between use and average cost.

The final feedback loops considered flow from AWS use to market share, with both increasing simultaneously. AWS market power can also be expected to increase with market share. With market power, AWS can reduce the affordability of AWS services by commanding high prices. Reduced affordability leads to increased revenue through the command of high prices. While these double-negatives seem cumbersome in the written description, they provide the convenience of representing affordability, reliability, and variety as consumer benefits with the same, positive, relationship to AWS use.

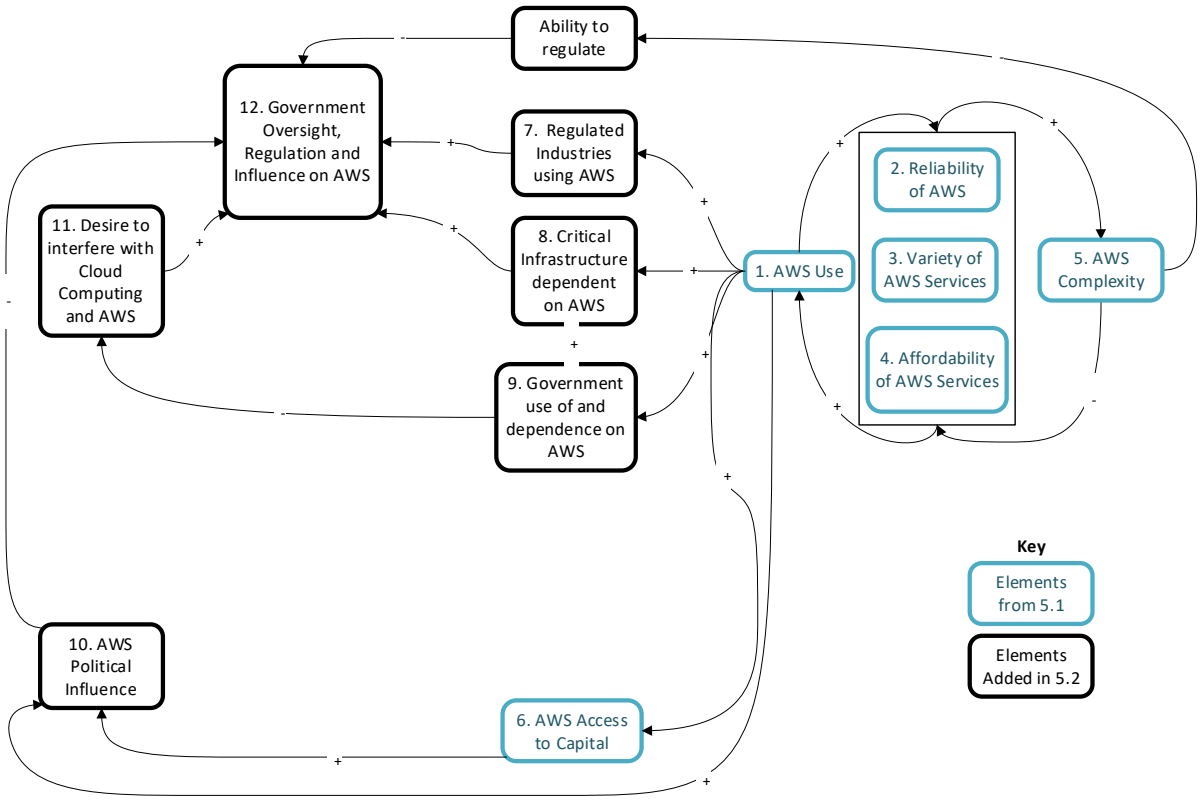
AWS market power, therefore, has both a potential reinforcing and balancing effect on AWS use. The reinforcing effect occurs by providing capital (through increased profits) for increased investment in services. The decreased profitability has a balancing effect of decreasing affordability and discouraging use.

Institutional Context

This section contextualizes AWS use in the context of critical infrastructure, and a government which acts is both a user and authorizer of AWS. Figure 5.2 is a systems diagram that depicts this contextualization. Within Figure 5.2, the key elements of Figure 5.1 have been condensed for simplicity, with a few key elements shown on the right side of Figure 5.2. This analysis begins by positively associating AWS use (5.2.1) with use of AWS by regulated industries (5.2.7), critical infrastructure (5.2.8), and the government's own use of AWS (5.2.9).

Figure 5.2

Systems Diagram of the Business System of Amazon Web Services (AWS), Contextualized by Relationships with Authorizing Institutions



Source: *Developed by author based on information presented in this work*

When regulated industries use AWS for regulated activities, this opens AWS to oversight by that industry's regulators. An example of this was reported in by Hoffman, Mattioli and Tracy (2019). In this case, the Federal Reserve Bank of Richmond conducted a formal examination of an AWS facility in Virginia. As more companies in regulated industries use AWS, the amount and degree of oversight can be expected to increase (discussed in Hoffman, et al (2019).

When companies that comprise the critical infrastructure of the United States use AWS it means that AWS itself is critical infrastructure. The Department of Homeland Security (DHS) is charged with ensuring the resiliency of critical infrastructure and it therefore has a responsibility to understand this dependence of critical infrastructure. DHS is therefore responsible to ensure the security of AWS information technology systems as well as their ability to rapidly recover from all hazards. However, to the degree that critical infrastructure (5.2.8) exists within regulated industries (5.2.7), DHS is expected to defer to that industry's regulators, which in the case of the Information Technology Sector is DHS. In other words, in developing its obligatory understanding of AWS's relationship to critical infrastructure, DHS relies on the IT sector's regulator, which is also DHS—there is no "other regulator for DHS to go to for help. Importantly, DHS, is a *user* of AWS cloud services precisely because it lacks the expertise in cloud hosting to host its own data.

Government use of AWS is also expected to grow with the use of AWS by all other industries. As the government increases its use of AWS, it is also expected to increase its dependence on AWS for crucial functions. As the government becomes increasingly dependent on AWS it may find itself less desirous of interfering with the functioning of AWS. For example, limiting mergers and acquisitions also limits the variety of services available to the government through AWS. More drastic action, such as a breakup or dissolution could severely complicated government processes that have become dependent on one or more AWS services.

The next variable related to AWS acknowledges the link between the financial capital of AWS and the political power of AWS. This political power primarily takes two

forms, political donations and lobbying. The Amazon.com PAC spent over $1.9 million on the 2020 election cycle in 2019-2020 (Center for Responsive Politics, 2020b). In 2019 Amazon.com, Inc. spent almost $17 million on lobbying (Center for Responsive Politics, 2020a). While these figures do not at face value suggest an inordinate influence on the democratic process, they do represent a means by which AWS use and revenues may lead to an increased influence on elected officials. Additionally, Amazon.com, Inc. exerts influence as an employer, purchaser, and tax-payer in its various jurisdictions. To the extent that Amazon.com has political power, that power contributes to the political power of AWS. This power could potentially be used to influence the manner and degree to which AWS receives oversight, regulation, and influence from the government.

Recent events have also highlighted a little-examined aspect of political power held by AWS. Related to the 2020 election, AWS hosted several significant functions, including campaign websites, over registration databases, election result storage, and more (Brooks, 2020). AWS is also host to social media sites which play a significant role in facilitating political debate and information dissemination. The potential influence of AWS became apparent when AWS decided to withdraw services from Parler on January 11, 2021, shortly after protesters attacked the U.S. Capitol building (Shead, 2021). Parler has been associated with supporters of Donald Trump, and was determined by AWS to be in breach of AWS terms of service because Parler failed to address what AWS described as incitements to violence that occurred on the Parler platform while hosted on AWS. Parler has since sued Amazon.com Inc for its decision to withdraw services, accusing Amazon of “political animus” (Shead, 2021).

The final pathway from AWS use to government oversight, regulation, and influence on AWS comes through flows from the variety of AWS services to AWS complexity. The complexity of the AWS information systems poses challenges to regulators. As this complexity increases, regulatory bodies will find it increasingly difficult to find experienced staff members who can learn the intricacies of the AWS system well enough to carry out their regulatory responsibilities. If governing bodies are not confident in their ability to oversee and understand the operations of AWS they will be even more hesitant to intervene with remedies even if they have reason to believe that AWS activities are not in the public interest.

Normative Beliefs in the Business System

Having modeled, at a high level, the business dynamics of AWS in the context of authorizing institutions. This section examines the receipts and deliveries between this system and the normative criteria that influence its behavior. As changes occur within the system, Hayden (2009) explains that the system provides social information back to the beliefs. This section will incorporate this social feedback into the system diagram for AWS, Figure 5.2.

AWS is influenced by many social beliefs. These beliefs influence the decisions AWS makes, as well as the decisions of firms, government agencies and the public who are stakeholders in the activities of AWS. These beliefs give rise to the normative criteria against which AWS is evaluated. These beliefs predate cloud computing and are continuously receiving social information from this institutional situation. By laying out the normative beliefs, this analysis will identify the relevant conflicts between these beliefs in the context of AWS and the cloud computing industry.

To incorporate social feedback into a system diagram, there needs to be a clear articulation of how beliefs are affected by actions and events. Social information can either challenge or reinforce a belief (Hayden, 2009). A belief is challenged by the presence, persistence, or proliferation of actions and evidence dissonant with the belief (Hayden, 2009). This dissonance can be created by the actions of authorizing institutions or processing institutions. If authorizing institutions fail to enforce the criteria of the belief, this inaction challenges the belief. If processing institutions fail to consistently conform to the demands of a norm, or the criteria imposed by the authorizing institutions, then this nonconformance challenges the social belief, the underlying cultural values, or both. Beliefs are challenged by actions contrary to the belief.

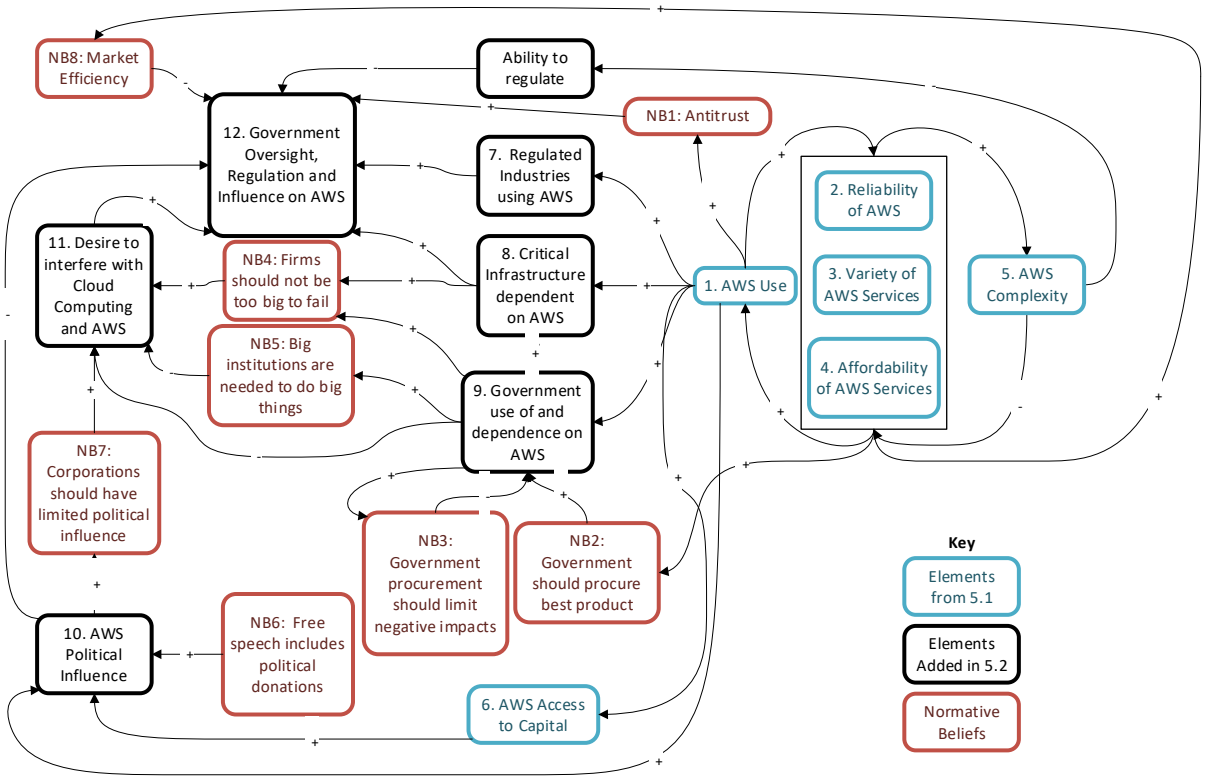
When authorizing and processing institutions act in ways consistent with a belief, this conformance adds to the inertia of the belief and reinforces that belief as well as its connection to underlying values. This reinforcement can also take place when a belief is challenged, but then bolstered through increased criteria imposed by the authorizing institutions onto processing institutions. The prevalence and enforcement of criteria reinforce the undergirding beliefs and values.

Figure 5.3 overlays beliefs onto the system diagram connecting AWS use with oversight and regulation, showing how at a high level each belief has the potential to receive and/or deliver at each step. As described above, conformance and non-conformance are the main ways by which beliefs receive input. In the context of this diagram, and consistent with Hayden (2009) the beliefs make deliveries into this system through authorizing institutions. While the specific institutions have been omitted from

Figure 5.3. The textual description here presented will address more specific institutions where relevant.

Figure 5.3

The Relationship between Beliefs and the Oversight, Regulation, and Influence of Authorizing Institutions on Amazon Web Services (AWS)



Source: *Developed by author based on information presented in this work*

NB1: Antitrust

Beginning, as before, with AWS use, the belief closest in system proximity is the belief of antitrust. This belief communicates that the public should be protected from abuse by large and powerful firms. Antitrust (NB1) has the following subcriteria:

- Nb1-1: Size invites regulation
- Nb1-2: Competition should be preserved
- Nb1-3: Size can be justified by consumer benefit

Subcriterion nb1-1 states that size inherently invites regulation. In order to protect its citizens from powerful firms, a state has the right to examine the business practices to ensure that the rights and capabilities of its citizens are not impinged by the firm. The first chapter of Paul (2020) provides a brief history of antitrust from a fairness perspective, demonstrating that even the early proponents of the economic argument acknowledged that the foundational regulations concerning antitrust were bound together with the "laws of fair conduct" (p. 6).

The argument of fairness need not be dealt with distinctly from the argument of efficiency in the context of this normative analysis, as both could potentially motivate the enforcement of the rules and regulations concerning antitrust. The unfairness modifier expands the potential application of antitrust laws beyond a response to a measurable economic abuse, but this is consistent with the legal framework explained in the following section that provides leeway for qualitative assessment of abuses in their

incipiency. The following section will briefly examine the history of Antitrust and justify the three subnorms presented above.

The economic argument against monopoly is based on a relatively straight-forward logic presented in entry-level economic textbooks, that monopolists are able to charge higher prices than they would be inclined to in the face of competition and this causes the quantity demanded and supplied to "fall below the social optimum" (Mankiw, 2015, p. 312). This elementary economic insight provides some of the economic justification for antitrust to exist as an exception to the norm of limited government intervention.

The subcriterion of nb1-1 requires the government to protect the public from monopolistic abuse. In response to the threat of unfairness or inefficiency caused by monopolistic behavior, the government can justify interference with the operations of monopolistic firms to preserve competition and to promote fairness and general economic welfare. This occurs on a discretionary basis. The Federal government derives this discretion from the Commerce Clause of the Constitution. This specific subcriterion was articulated into regulation in 1890 when congress passed the Sherman Anti-Trust Act. This act prohibits contracts and combinations, in restraint of trade.

In order to broaden the power of the federal government to protect competition from firms that would seek to reduce it, the Sherman Act was followed in 1914 by the Clayton Act and the Federal Trade Commission Act. The Clayton Act primarily outlawed price discrimination if it served to lessen competition. Three remaining clauses in the Sherman Act clarified potential applications of its scope. It forbade exclusive dealings and forced bundling of products whenever such actions lessen competition. It provided

for federal intervention in any proposed mergers and acquisitions, completion of which would "substantially" lessen competition. Finally, it barred any person from serving as a director for two or more competitors.

The Federal Trade Commission Act of 1914 stated that it created and empowered the Commission to do the following:

a) Prevent unfair methods of competition and unfair or deceptive acts or practices in or affecting commerce
b) Seek monetary redress and other relief for conduct injurious to consumers
c) Prescribe trade regulation rules defining with specificity acts or practices that are unfair or deceptive and establishing requirements designed to prevent such acts or practices
d) Conduct investigations relating to the organization business, practices, and management of entities engaged in commerce and
e) Make reports and legislative recommendations to congress. (Federal Trade Commission Act of 1914)

By this act, the FTC was created as an authorizing institution which would help to define the criteria, rules and regulations governing big business. The FTC was to replace the Bureau of Corporations, with all employees and functions being transferred to the FTC. The role of the Commission continues to evolve and this process is best captured by the court cases and amendments where specific clarifications and changes to the rules regulations and requirements take place.

The next significant act of Congress with regard to antitrust laws was the Robinson—Patman Act of 1936, which amended the Clayton Act by specifically

prohibiting price discrimination that "substantially" lessens completion or tends to create monopolies. In order to prove this price discrimination there must be two sales of a commodity that were made under very similar conditions but at different prices for two parties. This criteria narrows the scope of activity that is addressed by this act and some criticize its ineffectiveness (Blair, R. & DePasquale, C., 2014). Nonetheless, this legislation has been applied to a variety of situations including FTC v. Morton Salt Co (1948). In this case the FTC found that the Morton Salt Company had discriminated in price and issued a cease and desist order to stop the exclusive volume discounts provided to the largest retailers of its products. Cloud computing companies do provide volume discounts, but the criteria for fair discrimination in price have not been applied within the context of cloud computing.

The Cellar—Kefauver Act of 1950 prohibited the assets of one business from being purchased by another if this resulted in reduced competition. This provided the potential for the federal government to prevent vertical integration within an industry as well as to limit the scope of a company spanning multiple industries. This helped to fill a loophole by which businesses were expanding their control and influence simply by holding stock in other companies (Kaufman, A., & Englander, E. J., 1993). Amazon.com, Inc. has a long list of acquisitions, even the acquisition of competitors in the same industry. This acquisition activity falls under the purview of FTC, based on the Cellar—Kefauver Act.

While antitrust laws are generated by the legislative branch of government, they are interpreted by the courts and executed by the Department of Justice and, after 1914, the Federal Trade Commission. It is in the application of the laws that the criteria become

clear and social norms are established about the ways in which businesses are permitted to grow and the activities in which they are permitted to engage.

Hovenkamp (2010) makes a case that the FTC was created to expand the scope of antitrust beyond the written words of the Sherman and Clayton acts to be able to address actual unfairness and dishonesty in their "incipiency". This incipiency refers to the wording of the Clayton act in which actions are prohibited "where the effect *may be* to substantially lessen competition" [emphasis added]. The phrase "may be" would allow for regulators to take action in a case where no damages have occurred and where no monopoly power exists yet. Hovenkamp cites FTC v. Sperry & Hutchinson Co. (1972) in which the justices suggest that the reach of the FTC extends into a "penumbra" of activities that fall into unfair or dishonest practice but are not explicitly stated by the law. This provides leeway and flexibility in how the FTC regulates businesses. What this leeway and flexibility means for the cloud computing industry is that there are broad options from a legal perspective for addressing issues of scale and scope in the cloud computing industry. The decision to act, however, will likely need to take into account other precedents, current norms in other industries, and well as social and performance indicators that take into account the outcomes and potential outcomes in the cloud computing space.

Having grounded the belief of antitrust and its sub-beliefs in the applicable legal history and development in the United States, the belief will now be addressed within the context of the systems diagram, Figure 5.3. In this diagram, AWS use serves as a proxy for size which was previously established as the primary trigger for the application of antitrust action. This means that as AWS use increases, its size creates cognitive

dissonance with the antitrust belief. As the size of AWS increases, so does the responsibility of the FTC to investigate the activities of AWS. The responsibility of the FTC to intervene in those activities also increases as the scale and scope of AWS creates opportunities for abuses and perceived abuses. In this way, as AWS use increases, so do the impositions of the antitrust belief, which lead to increased oversight, regulation, and influence on AWS.

NB2: Government should procure best product

The government must purchase goods and services in order to carry out its responsibilities on behalf of its citizens. The two norms which impose demands on the procurement process are: NB2, government should procure best product; and NB3, government procurement should limit negative impacts. In the United States, the Office of Federal Procurement Policy (OFPP) shapes the policies and practices of how agencies carry out their requisite purchases (The White House, 2021). The OFPP issues a variety of guidelines and memoranda which aim to improve the acquisition process of United States agencies. These guidelines direct the procurement process toward quality, cost savings and efficiency (The White House, 2011) (White House, 2004) (General Services Administration, 2021). In this way the Government expresses NB2.

NB3: Government procurement should limit negative impacts

Federal Acquisition Regulations (FAR) is issued by the Government Services Administration, National Aeronautics and Space Administration, and the Department of Defense. It lays out a consolidated set of guidelines for government purchasing. Within the FAR, little attention is given to undesirable market and industry impacts of government procurement. One exception within the FAR is an exception for contracts set

aside or targeted for small or minority owned businesses. The opposite consideration, of withholding contracts from firms that are too large or whose award of the contract may further cement their dominance in an industry is not addressed. The provisions for small and minority-owned businesses is an example of NB3 balancing NB2. It demonstrates that there are other concerns besides quality, cost, and efficiency that agents of the Federal Government consider when purchasing. While protections for small and minority-owned businesses are considered there are not documented guidelines concerning large corporations and the impact of government purchasing on industrial organization.

NB4: Firms should not be too big to fail

As more components of the nation's critical infrastructure become dependent on AWS, the more problematic AWS outages become. This dependence could force the government to intervene in the event of a failure. The phrase *Too Big to Fail* became popularized in the 1980s and became a prominent theme in the aftermath of the 2008 global financial crisis (Dash, 2009). It refers to institutions whose demise could have significant adverse effects on the broader economy. The government may be inclined or induced to interfere with a disorderly dissolution of the troubled institution. In the case of AWS, the justification for intervention increases further to the extent that the government itself is reliant on AWS. For technical failures, intervention is not a major consideration, but from a financial perspective, AWS could face moral hazard in the way it conducts its business and especially the financial risks that it takes. So as dependency on AWS increases, the perception of AWS being too big to fail increases and this would create a desire on the part of some to interfere with cloud computing and AWS. The financial

strength of AWS and Amazon.com Inc. mitigates the perception of being too big to fail, as the likelihood of failure seems low. Given an interest coverage ratio of 20.98 at the end of 2020 (Reuters, 2021), Amazon.com Inc uses leverage, but not to the point that there is concern for its ability to cover its financial obligations. Also, as of December 31, 2020, Amazon reports having more than $84 billion in cash and short-term investments; in this context, increasing leverage likely has more to do with the effects of its capital structure on its own cost of capital rather than materially increasing Amazon's default risk.

NB5: Big institutions are needed to do big things

Standing in opposition to the idea that firms should not be too big to fail is the belief that big institutions are necessary to do big things. This belief is articulated by Galbraith *in The New Industrial State* (2015). In a different way, it is also in the introduction to Drucker (1962) as an "acceptance of big business by the American public as the tool best suited for most of the economic jobs of our society…" Hagel, Brown, & Davison (2009) offer a third perspective that comes to a similar conclusion, arguing that large firms help to scale long-term trust-based relationships which facilitate the sharing and utilization of tacit knowledge in a knowledge-based economy. Overall, this belief articulates that there are valid reasons to allow organizations to grow to a large size.

In Figure 5.3 government use of and dependence on AWS reinforces the idea that large and complex institutions are able to efficiently deliver reliable solutions on the scale needed by the federal government. Government dependence on AWS becomes representative of the value created by large corporations. Because of the economies of scale and efficiencies enabled by a common platform for cloud hosting and development,

both private and public sectors benefit from this size and the network effects that accompany it.

To the extent that people believe that large institutions serve a critical role in the economy and provide value, there is less pressure to interfere with industries that tend to concentrate power in just a few firms. The longer AWS operates at this size and scale and as a good partner to with the government, the more comfortable the American public is with the idea that largeness is necessary. While there may be some interest in evaluating the health of such an integral component of the economy, there is less likely to be an adversarial relationship between the public or public sector and AWS.

NB6: Free Speech Includes Political Donations

This chapter previously outlined the connections between AWS use and AWS political influence. This influence is partially sustained by a belief that free speech includes political donations. This belief is clearly encapsulated in the United States Supreme Court decision on Buckley v. Valeo, which states that "A restriction on the amount of money a person or group can spend on political communication during a campaign necessarily reduces the quantity of expression by restricting the number of issues discussed, the depth of their exploration and the size of the audience reached." (Supreme Court Of The United States, 1975, p.19) The opinion elaborates by articulating how "virtually every means of communicating" requires expenditure. Extrapolating from that, it is clear to see that increasing the ability of an organization and its stakeholders to spend necessarily increases their ability to communicate. The more AWS makes, the more its voice and preferences can be heard. This is true in principle even if this power or

voice is not currently being exercised. To the extent that this belief holds sway, the connection between AWS access to capital and AWS influence remains in tact.

NB7: Corporations should have limited political influence

The idea that corporations should have limited political influence is also clearly articulated in the appellee's argument in Buckely v. Valeo (United States Supreme Court, 1975, p. 25). As articulated in the opinion of the court, the appellee's position is aimed at three governmental interests. They are first, "…the prevention of corruption span and the appearance of corruption spawned by the real or imagined coercive influence of large financial contributions on candidates' positions and on their actions if elected to office." Second, to "mute the voices of affluent persons and groups in the election process and thereby to equalize the relative ability of all citizens to affect the outcome of elections." And third, to "act as a brake on the skyrocketing cost of political campaigns and thereby serve to open the political system more widely to candidates without access to sources of large amounts of money." The opinion goes on to clarify that vested interest and quid pro quo arrangements undermine representative democracy.

The United Nations' 1948 Universal Declaration of Human Rights, Article 21, states that "The will of the people shall be the basis of the authority of government." Ignoring that corporations are legally considered people in some contexts, it is clear that corporate interests will not always align with the will of the (real) people. As corporations are given voice in political decision-making, that voice will align with those people who share a vested interest with the corporations but may be misaligned with other stakeholders of the business.

As corporate interests are perceived to be represented in government actions, or in figure 5.3, as AWS political influence becomes apparent, it will be at odds with the belief that corporations should have limited political influence. This dissonance will reinforce the belief and thereby increase the desire to interfere with cloud computing and AWS. This creates a balancing dynamic that can come into play to limit the political power of AWS as that political power increases to a visible or otherwise undesirable level.

NB8: Market efficiency

Market efficiency denotes the expectation for markets to tend toward efficient outcomes through the discipline of creative destruction. This logic predicts that firms which deliver superior value to their stakeholders will survive, while firms that do not will fail. This form of economic selection is akin to the process of natural selection. An extension of this norm is the expectation that the breadth of business processes within a firm are also subject to a similar mechanism of selection. If certain business processes fail to lead to increased profitability, then they can be expected to be eliminated which would therefore leave only the most efficient processes. Efficiency in and of itself is an inherently normative term.

Market efficiency is often traced at least as far back as Adam Smith's *The Wealth of Nations (2013)*, originally published in 1776. Smith explains market efficiency in terms of specialization and the division of labor. The concept of creative destruction originates in the Marxian critique of capitalism (Reinert, H. & Reinert, E. S., 2006), but is popularly associated with Schumpeter. Schumpeter famously developed this concept in his 1942 work *Capitalism, Socialism, and Democracy* (2010). Schumpeter explains how capitalism involves both the accumulation and destruction of wealth through competition

and the struggle for survival. The efficient market hypothesis became popularized by Fama in the 1960's, but was effectively formulated in 1900 by Bachelier (Dimson, E., & Mussavian, M., 1998). This hypothesis states that financial markets price in all available information. It represents a logical conclusion of the market efficiency norm, which imbues markets with the power to eliminate waste and falsehood through the pursuit of profit.

This belief matters to an analysis of the cloud computing industry because it provides an instrumental and therefore ethical basis for the pursuit of profit. If profit is consistent with efficiency, then this provides an instrumental reason to avoid interference with cloud providers. Cloud computing providers, to the extent that they are operating in a competitive environment, are expected to create and adopt efficient processes for service delivery. Failure to operate efficiently is expected to yield sub-optimal returns and potential elimination from the industry. The pursuit of profit is assumed to proxy the pursuit of efficiency in the absence of monopoly or other market inefficiency.

In Figure 5.3, NB8 directly influences government oversight, regulation, and influence on AWS. By presupposing the supremacy of market forces over government interventions, this belief will temper the desire to interfere with cloud computing and AWS. This belief is influenced by the system primarily through the consumer benefits of AWS. To the extent that AWS delivers a reliable, variety of affordable services that meet the needs of its customers, it will be perceived to be acting efficiently and thereby reinforcing the belief that for-profit firms tend toward efficiency and social benefit.

Associated with this major norm are two key subnorms, the first is the belief that innovation is rewarded and therefore encouraged by competitive markets. This is a

common theme in neoclassical economics. Baumol (2002), building on Schumpeter, argues in the first part of his book, *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism* that innovation is an inevitable outcome of an oligopolistic capitalist free market system. This theme is further developed to align with the title that free markets promote innovation. The phrase free markets may be associated with many meanings. Baumol here uses it to describe competition of at least an oligopolistic nature. This includes an underlying rule of law that includes enforceability of contracts and property rights. Within those property rights are assumed the rights to intellectual property that can then be sold, traded, or rented. This segues to the second subnorm.

Intellectual property rights confer a temporary competitive advantage to a market participant based on discovering or inventing a patentable product or process. When a firm stands to secure this competitive advantage, it is expected to be more likely to innovate and develop novel products and processes. In theory, a non-protected invention will yield very short-term supernormal profits until other firms are able to replicate the invention. Mennell (2019) explains how intellectual property rights can be used to generate market dominance through network effects. Mennell shows that by creating systems with proprietary interfaces, network effects may be created and preserved. Mennell provides many examples of modern technology companies, including Amazon.com, Inc., that benefit from this dynamic.

Policy Considerations

Returning to Hayden (1998), the normative criteria driving the complex interactions of AWS, regulators, policymakers, courts, and—frankly—the rest of the economy, there are innumerable and obvious "gaps, discontinuities, disharmonies, and

conflicts" that are of "paramount concern." Figure 5.1 and 5.3 allow us to consider various policy implications within the context of the business system and normative beliefs. This section will propose five policy options for influencing this complex system and then examine possible systemic effects. These are:

1. Divestiture
2. Merger prevention
3. Treat as Utility
4. Strategic Purchasing
5. Oversight

Divestiture

The first options considered would be to dissolve Amazon.com, Inc. into smaller companies much like the breakup of the Bell System. Such a divestiture would originate with the FTC and be carried out by the Department of Justice. The obvious perforation exists between AWS and its parent company Amazon.com, Inc. By severing this link, the synergistic flow of capital between AWS and its parent company would cease. This can be visualized in figure 5.1 by simply severing the connection between Amazon.com Access to Capital and AWS Access to capital. In this simple divestiture model, a source of capital has been removed from the system as well as a potential drain on capital from the system. The effect this has on the stability of the system overall depends on whether it is more likely that the parent company draws capital from or invests in AWS. In either scenario, the parent company and the subsidiary provide diversification and therefore a degree of stability to one another. If financial insolvency is one potential cause by which

AWS fails to deliver expected services, then this policy choice leads to a less stable system.

By severing the ties between AWS and its parent company, it would be more likely for Amazon.com, Inc. to pursue an alternate vendor of cloud services, thus reducing the AWS market share and revenue base. Given the difficulty of switching cloud platforms this seems unlikely in the near-term. The more likely impact on AWS market share and revenue would come from an increase in use from retail competitors of Amazon.com who boycott the subsidiary because of the parent company. Spinning off AWS as a separate company seems to be the most likely starting place for a divestiture at this point. Other potential divestitures could involve separating the infrastructure as a service portion of the company from software as a service portion of the company. The issue here is that the lines tend to blur and there are many other competitors who are vertically integrated in this way.

Merger Prevention

A less disruptive approach to address the size and scope of AWS would be to prevent future mergers and acquisitions related to AWS. This process would take place as part of the FTC's standard review of significant mergers and acquisitions. Prevention is a much simpler process than divestiture, and while it does not address the current scale of AWS, it can reduce the rate at which AWS is growing. AWS has been able to incorporate the technologies of its competitors into its suite of tools. This can be seen in its acquisition of Cloud9 IDE (Janakiram, 2016). In this acquisition, AWS both eliminated a competitor in the development-as-a-service platform space as well as filling out its suite

of services for cloud development with what Janakiram (2016) describes as "a robust development environment that's tightly integrated with its platform."

Eliminating future mergers can be expected to increase competition for cloud service providers in the short run. What is not clear is to what extent such acquisitions encourage companies to invest in cloud computing services that might be acquired by AWS. Additionally, the purchase of competitors may be less disruptive to the target companies and their employees than continuing to compete with AWS and potentially being put out of business by a larger company with a greater ability to invest capital in its services. The lasting effects of these types of limitations is difficult to assess. Figure 5.3 shows us that AWS market share would grow more slowly at first, while the company would retain more capital which it could invest in its services, or reinvest in its parent company. The variety of AWS services would also be held to a lower value in the short-run, as it is likely to take longer to develop these types of tools versus acquiring them. Such a limitation reduces the consumer benefit provided by AWS until such tools can be added to the AWS stack. This also reduces the network effects and other efficiencies gained from a robust and integrated environment.

Not only is merger prevention an answer to the size and scope of AWS, it also addresses issues of stability as well. As previously mentioned the temporary suppression of the variety of AWS services will necessarily also suppress the complexity of AWS. One could also argue that in-house developed applications are less likely to fail than those that have been purchased and integrated, but there would be strong assumptions made in such an argument concerning the quality of internally versus externally developed code. The speed at which the code is developed may impact quality, if AWS is

rushed in its desire to fill out a particular service offering. Additionally, it is unclear whether being developed in-house really provides advantages from an integration and stability standpoint. If however, the lower variety of services in the short-run persists and is not made up for elsewhere, then the complexity at each future time period has been reduced and each time period should be more stable than it would be with mergers allowed. This suppression of variety may also suppress AWS use, and this will reduce the impact of an outage at all future times as well.

Treat as a Utility

Resuming a theme from the introduction of this dissertation, it is conceivable that computing could be considered a utility and regulated as such. Of the policy options here considered this is the most disruptive to the business system as it would involve a near insurmountable degree of oversight and regulation to implement. Rather than examine the effects on the system, it may be simpler to outline the reasons why computing does not lend to the utility model.

First, utilities tend to provide largely undifferentiated commodities to consumers. In the case of cloud computing, the services proffered are highly differentiated. At every level of service from infrastructure as a service to software as a service, the differentiation and customization is considerable.

Second, utilities, as a common carrier, have historically provided services to the broad public. In the case of cloud computing, customers may include the public, but there is not necessarily a need for every or even most citizens to have direct access to cloud computing. Most individual users will access their cloud computing resources through at least one intermediary. For example, users who backup iPhone information to the iCloud

interface with Apple, but Apple is uses AWS and other cloud providers behind the scenes.

Based on the two reasons above, it is hard to imagine that cloud computing would be a good candidate for the utility model. Furthermore, additional consolidation would be required in order to operate cloud computing as a utility and this would significantly increase the systemic risk created by AWS. At least in the current arrangement there are other competitors operating systems in tandem—that is to say there is an alternative to switch to, even if at high cost of time, money and effort. The closest historic example of a utility in the information and networking space is The Bell System. Temin & Galambos (1987) explain some of the difficulties caused by both the operation and dissolution of the Bell System. Two difficulties in operating the Bell System are germane to this analysis. First, that of pricing the services; and second, that of regulating the market. The complexity of pricing both long-distance and regional services created difficulties in establishing fair and useful prices for the services that were interdependent on each other. From this perspective, the variety and interdependence of services within cloud computing would make regulated pricing very difficult to execute. Furthermore, the introduction of competitors into certain aspects of the Bell System's business model made it very difficult for the highly regulated firm to compete with new and less regulated entrants. Eliminating competitors in cloud computing would be both difficult and come at the cost of innovation.

Strategic Purchasing

The government is a significant purchaser of cloud computing services. As discussed in the introductory chapter, the JEDI contract alone was able to cause a 6%

swing in market share between two suppliers of cloud infrastructure. The previous section of Chapter V explained the conflicting beliefs related to government purchasing and noted that monopolistic concerns are not addressed in federal guidelines for government purchasing. Given the reinforcing nature of the cloud computing industry, path dependence and network effects, government purchasing is able to make a significant and lasting impact on the cloud computing industry. While the government has a direct interest in obtaining the best products and services, it must also consider the impact to the nature of competition in the industries in which it purchases. In cases where services are similar, purchasing from a firm with a smaller market share, though not otherwise considered a small business, may be a way of preserving competition. It may also be a way of spreading out the risk of a systemically important service.

Chapter V Conclusion

Hayden (2008) articulates the optional nature of Polanyi (1957) and Commons (1924) versions of prohibitions, obligations, and permissions, stating, "some actors have the status and power to be indifferent toward a normative rule." (p.106) This indifference can be seen in how AWS continues to grow and maintain significant market share in juxtaposition with NB1: Antitrust, and NB4: Firms should not be too big to fail. This indifference is not limited to AWS, but rather, it extends to many large firms. Figure 5.3 describes how these beliefs are influenced by several systemic factors, including the consumer benefits of AWS, AWS complexity, and countervailing normative beliefs.

This chapter demonstrates how the tension between AWS and the outlined beliefs will continue, as many of the systemic effects are reinforcing with the growth of AWS. There are several balancing effects in which feedback from the system brings cognitive

dissonance with the normative beliefs which would influence the system. The primary balancing effect, within the business system, relates complexity of AWS systems. Cloud computing is an inherently complex service to provide. The opportunities for failure stem both from the tight coupling of the system as well as the complex interdependencies which grow exponentially as new services are added. It is unclear if or when this effect of complexity will begin to significantly reduce the stability of the system and the value it provides consumers. What is clear is that if the business system or the technological system becomes destabilized, it would be helpful to have other cloud providers available to use, interoperability between those cloud providers, and critical infrastructure spread out between different cloud providers.

CHAPTER VI

THE SOCIAL FABRIC MATRIX

Setting up a Social Fabric Matrix of Amazon Web Services

The purpose of this chapter is to apply the social fabric matrix approach (SFM-A) to detail the operations Amazon Web Services within the cloud computing industry and the broader economy. The intent is to outline the institutional framework in which AWS does business. By contextualizing the operations of AWS within the broader social norms, institutions, and technologies, one can better understand which institutions and beliefs contribute to the increasing scale and scope of AWS, and which institutions and beliefs can be expected to constrain its future growth and influence. The SFM-A will elucidate the interrelationships within the industry and with industries and entities that provide inputs to or receive inputs from the cloud computing industry. The SFM-A in this case will forego an analysis of ecology and attitudes and focus on the social norms, the institutions, and the technologies which are relevant to cloud computing.

Figure 6.1 details a social fabric matrix (SFM) of the cloud computing industry. Each of the vertical components deliver inputs to the receiving components listed horizontally. A "1" in a cell of the matrix denotes a delivery from the delivering component to the corresponding receiving component. A cursory glance reveals that the interconnections are more numerous than can be exhaustively analyzed within this study. Rather, each component will be described based on its significance to the research question, and the individual deliveries will be detailed, only to the extent that they contribute relevant information to this inquiry.

Figure 6.1

Social Fabric Matric Approach to the Cloud Computing Industry

Delivering Components \ Receiving Components			Social Norms								Institutions												Technology							
			Antitrust	Government should procure best product	Government procurement should limit negative impact	Firms should not be too big to fail	Big Institutions are needed to do big things	Free speech includes political donations	Corporations should have limited political influence	Market efficiency	US President	US Congress	The US Supreme Court	Federal Trade Commission	Department of Homeland Security	Federal Communications Commission	Banking Regulators	Amazon Inc (non-AWS)	Amazon Web Services	AWS Competitors	AWS Customers	Contributors to AWS ecosystem	On-premise model	Hypervisor	Multi-tenancy	Redundancy systems	Encryption	Internet Infrastructure	Information Security	Malware
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Social Norms	Antitrust	1									1	1	1	1			1	1	1	1		1								
	Government should procure best product	2										1			1							1								
	Governement procurement should limit negative impacts	2										1																		
	Firms should not be too big to fail	3									1	1	1				1	1	1	1										
	Big Institutions are needed to do big things	3									1	1	1	1	1			1	1			1								
	Free speech includes political donations	4									1	1	1					1	1											
	Corporations should have limited political influence	4									1	1	1	1				1	1											
	Market efficiency	5									1	1	1	1	1							1								
Institutions	US President	5	1	1	1	1	1	1	1	1		1	1	1	1	1	1													
	US Congress	6	1	1	1	1	1	1	1	1			1	1	1	1	1													
	The US Supreme Court	6	1				1	1	1																					
	Federal Trade Commission	7	1			1	1			1							1	1	1	1	1	1					1			
	Department of Homeland Security	7		1	1	1												1	1								1	1	1	
	Federal Communications Commission	8																	1								1			
	Banking Regulators	8				1	1			1				1					1	1	1					1	1		1	1
	Amazon Inc (non-AWS)	9	1	1	1	1	1	1	1	1									1											
	Amazon Web Services	9	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1
	AWS Competitors	10	1	1	1	1	1	1	1	1	1	1							1		1	1	1	1	1	1	1	1	1	1
	AWS Customers	10										1							1	1		1								
	Contributors to AWS ecosystem	11																	1	1	1									
Technology	On-premise model	11																	1	1										
	Hypervisor	12																	1	1										
	Multi-tenancy	12																	1	1										
	Redundancy systems	13																	1	1										
	Encryption	13																	1	1										
	Internet Infrastructure	14																	1	1										
	Information Security	14																	1	1										
	Malware	15																	1	1										

Source: *Developed by author based on information presented in this work*

Beliefs

Beliefs are essential to evaluating the policy considerations related to the cloud computing industry. These normative beliefs are denoted as NB*i* (where *i* = 1,2,3, and so on). For the consideration of this study, nine relevant beliefs have been selected:

NB1: Antitrust

NB2: Government should procure best product

NB3: Government procurement should limit negative impacts

NB4: Firms should not be too big to fail

NB5: Big Institutions are needed to do big things

NB6: Free speech includes political donations

NB7: Corporations should have limited political influence

NB8: Market efficiency

As discussed in the previous Chapter, these 8 beliefs are in mutual tension with one another, especially as they influence government oversight, regulation, and influence on AWS

Institutions

The institutions listed in the SFM include law makers, public purse-holders, regulators, firms, and other stakeholders. These institutions can be usefully grouped into authorizing and processing institutions, consistent with Fullwiler (2009) and as set forth in Hayden (2009). Even though AWS does business internationally, the present study will focus on those authorizing institutions which govern activity within the United States.

This is done for sake of simplicity and given that AWS is headquartered in the United States.

The authorizing institutions can further be subdivided into higher and lower authorizing institutions. The authorizing institutions in the United States relevant to this discussion receive their authority through Congress and the President. This authority is further defined by the court system, including the Supreme Court. The lower authorizing institutions include the Federal Trade Commission, the Department of Homeland Security, the Federal Communications Commission and banking regulators. The deliveries between these institutions and AWS will be established below in the section concerning rules, regulations, and requirements.

In the bottom right corner of Figure 6.1 is a highlighted region depicting the deliveries, between AWS, its competitors, its customers, and others who contribute to the AWS ecosystem by creating software that enhances its usefulness. The symmetry of the diagram shows that there are mutual deliveries back and forth between these institutions. It should also be noted that there is overlap in the identities of these institutions, as established in Chapter III: some entities, like Netflix, acting as customer, competitor, and a contributor to the AWS environment by creating and sharing the software, ChaosMonkey, to improve resiliency of cloud implementations. This region also contains within it, regulators that also happen to be AWS customers. This means that authorizing institutions are also benefactors of AWS, its competitors, and its customers. The overlap is difficult to fully convey using the SFM, but it is present, nonetheless.

Technologies

There are seven technologies which comprise the fourth section. These technologies enable the business practices of AWS, its customers, competitors, and the institutions that authorize its activities. The first technology is the incumbent, on-premise data center model (T1).

On-Premise Model (T1)

This model became the primary mode of computing after mainframe technology and remained dominant until the advent of cloud computing. In the on-premise data center model, a company maintains a servers in racks in a data center that it is responsible for operating. Company employees conduct or oversee all of the activities required to deliver data center services to the organization. Admittedly it is common for some or all of these activities to be outsourced to another company which provides these services. However, this is distinguished from cloud computing in that these activities take place in a data center controlled by and dedicated to the organization. This technology may never be eliminated entirely by cloud computing but it serves as the current platform for many organizations shifting to the cloud, as well as the likely fallback option if the cloud cannot meet the demands of its users.

Hypervisor (T2)

Hypervisor technology allows for the virtualization of servers. Hypervisors are a common component in both on premise and cloud data centers. Before virtual machines, a single server instance required its own, physical server: a single computer in a rack, comprised primarily of disk, memory, and processor with a single operating system. This was a one-to-one relationship: one single instance with one dedicated operating system,

CPU, memory, and storage. Virtualization allows multiple operating systems to run on one machine, or for multiple machines to pool their resources to run any number of operating systems. A many-to-many relationship allows the disk, memory, and processor of a server to be divided and shared across multiple virtual machines based on their need. In this model excess capacity on one server can be made available to the virtual machines that need it most.

Multi-tenancy (T3)

Multitenancy is often achieved using virtual machines; however, virtualization is not strictly required for multi-tenancy. Multi-tenancy allows multiple users to use the same instance of a virtual or physical server. Each user can only access the resources assigned to them and do not expect to be affected by the activity of other users. Multi-tenancy complicates the legal environment of computing because the established laws did not anticipate multi-tenancy.

Redundancy Systems (T4)

Redundancy is common in computing and other industries. The premise is to have extra, un- or under-utilized resources available in the event of failure. On premise data centers may effectively have a second, or third clone of their data available in the event of failure, with the ability to failover piecewise or holistically. They may also spread out their data center operations geographically so that the backups are in a different location. In the cloud, redundancy is often achieved between data centers and even regions—redundancy could even be achieved between two different cloud providers. The extent of redundancy contributes both to the cost and, ideally, the stability of any system and this is

true of cloud computing. The phrase 'ideally' is used here because of the potential for iatrogenic effects (Taleb, 2012) caused by increased complexity (Perrow, 2011).

Encryption (T5)

Data may be encrypted in storage (encryption at rest) and/or encrypted in transit upon exiting a sub-system. In either case, the data is encoded, and made unreadable without a key. While there are ways of breaking encryption, most are both too costly and time consuming to be practical. Encryption ensures that only the proper recipient, or key-holder will be able to decipher the information disclosed. This means that simply intercepting data from storage or in transit does no good unless it can be intercepted from the system that has a key. Thus, the physical burglary of a data center is less problematic than acquiring digital access credentials.

Internet Infrastructure (T6)

Internet infrastructure refers to the system of software, hardware, and organizations which host and service the internet. Because cloud infrastructure tends to be geographically disbursed from its users, the internet often (though not always, as in the case of air-gapped proprietary clouds), provides the means by which users interact with their cloud resources. This means that any interruption of internet services is most likely accompanied by a similar disruption of cloud services. This coupling of the two infrastructures significantly increases the effects of an internet outage or a widespread degradation of performance.

Information Security (T7) & Malware (T8)

The final two technologies are opposing and complementary. They are the body of information security systems and the malware and methods designed to exploit them.

In this case information security systems include both practices and their technical implementation; they are the systems that prevent unauthorized access to information systems. Conversely, malware technology and predatory methods are those that seek to obtain unauthorized access or cause an undesirable outcome. These two technologies are grouped because of the tight coupling of interactions: a new security feature prompts new methods of circumventing it and new nefarious methods require new ways of thwarting them. They are kept distinct because each gives and receives different inputs to and from the various institutions.

A Normative Systems Analysis of Amazon Web Services

Having constructed the SFM, the next step is to establish the deliveries between the beliefs, institutions, and technologies listed above and use them to describe the normative criteria guiding the institutional interactions of AWS. This approach provides a structured way by which to investigate the interaction of normative criteria and institutional action. Consistent with the approach taken in Fullwiler (2009) this SFM-A will detail the criteria, rules, regulations, and requirements of the authoritative and processing institutions. This will determine the "ends toward which the system is dynamically evolving" (Fullwiler, 2009). This approach incorporates primary sources of a legal or regulatory nature, as well as artifacts such as speeches, policies, press releases, or other narratives which document the normative characteristics of the system. As

Fullwiler (2009) explains, this supplies the evidence for how we know that we know how a system works.

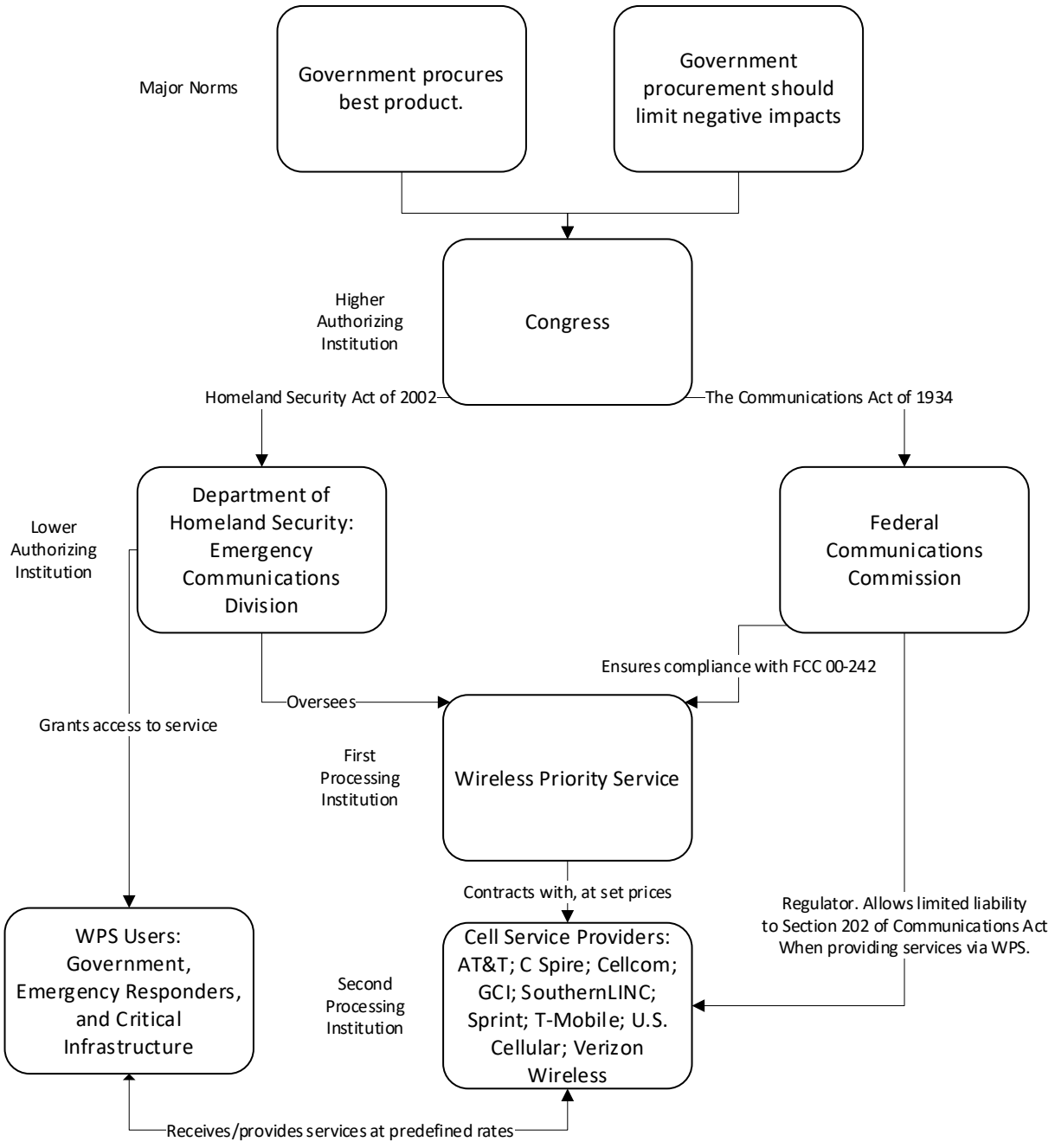
Normative Beliefs, Authorizing Institutions, and Processing Institutions

The eight normative beliefs identified in the previous chapter provide the broad normative context in which the authorizing institutions act to determine the behavior of the system. Within that context, these normative beliefs direct behavior of the higher authorizing institutions, lower authorizing institutions, and processing institutions. The obligations of the authorizing institutions may be obfuscated by complex relationships with the other actors. In the case of cloud computing, authorizing institutions are both consumers and regulators of cloud services. This creates a more complex web of relationships, but there is plenty of precedent of authorizing institutions navigating dependency and authority. One specific example is explored below:

The dual role of government as consumer and regulator can be seen in action in the Government Emergency Telecommunications Service (GETS) and its companion service, the Wireless Priority Service (WPS). GETS provides national security, emergency preparedness, and critical infrastructure personnel a high probability of call completion on a landline during an emergency, when communications networks are congested. WPS provides the same priority for calls made through a wireless network. WPS is an apt articulation of the major norm of government protection and the sub-criteria of smooth functioning and self-protection in the context of critical communications services.

Figure 6.2

Articulation of major norms into rules, regulations, and requirements for Wireless Priority Service



Source: *Developed by author based on information presented in this work*

In Figure 6.2, authority works through the higher authorizing institution of Congress to create the lower authorizing institutions. The Department of Homeland Security was created by the Homeland Security Act of 2002. Through this act, several other agencies and functions were combined under the leadership of DHS. One of the functional components of DHS is the Cybersecurity and Infrastructure Security Agency (CISA, not pictured). The Emergency Communications Division of CISA contains within it Priority Telecommunications Services, which includes WPS. WPS coordinates with mobile service providers to allow for the prioritization of national security and emergency preparedness calls in the event of call-congestion and/or emergency.

In conjunction with this function of DHS, the Federal Communications Commission is also authorized by congress to regulate the communications industry as a form of inter-state commerce. The FCC works in conjunction with DHS to ensure that the WPS operates in compliance with FCC 00-242. It also offers limited liability to the operators of WPS as it relates to Section 202 of the Communications Act. Section 202 prevents common carriers from offering unfair preference to any customers. The FCC does not force mobile providers to participate in WPS, but it does provide the guidelines, standardization, and legal protection that ensures that emergency access is uniformly offered across the country.

The WPS provides an established model for the prioritization of critical communication infrastructure based on Government-determined needs. This prioritization process is facilitated by DHS and will undoubtedly inform the Next Generation Network Priority Services (NGN-PS). NGN-PS will provide priority communications through internet protocols and other more advanced communications networks. Government

priority in cloud computing could follow this precedent from the mobile phone industry. Regardless of priority access, the government will need to navigate its role as both consumer and regulator of this new and growing sector of the economy.

Rules, Regulations and Requirements

This section will introduce and further explain how the normative beliefs are articulated to the cloud computing industry through rules, regulations and requirements issued by the authorizing institutions and carried out by the processing institutions. While some of these institutions have been introduced in previous sections and chapters, others will be addressed as they relate to AWS.

Incorporation

As established in the first chapter, AWS is a subsidiary of Amazon.com Inc. Amazon.com Inc. is a corporation organized under the General Corporation Law of the State of Delaware (Amazon.com, Inc., 1996). Incorporation establishes a charter and bylaws, which explain how shareholders empower board members to act on their behalf to oversee the operations of the corporation. Amazon.com, Inc. is governed by a board of 10 directors. These directors are elected by in accordance with its charter and bylaws (Amazon, 2016). These directors appoint the executive team which is responsible for establishing policies, executing the strategy, and establishing the practices and procedures that govern the organization. Being incorporated in Delaware means that corporate law disputes must be arbitrated by the state's specialized Chancery Court (Daines, R, 2001).

These disputes would be related to corporate governance and the way in which the Board of Directors represent the shareholders in the oversight of the corporation.

Incorporation as a publicly traded company brings with it the oversight and accountability provided by the Securities and Exchange Commission (SEC). The SEC is endowed by congress with the mission to "protect investors, maintain fair, orderly, and efficient markets; and facilitate capital formation." (Securities and Exchange Commission, 2013) This has been established through the Securities Act of 1933, the Securities Exchange Act of 1934, the Trust Indenture Act of 1939, the Investment Company Act of 1940, the Investment Advisers Act of 1940, the Sarbanes-Oxley Act of 2002, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, and the Jumpstart our Business Startups Act of 2012 (Securities and Exchange Commission, 2013). NASDAQ, as the listing exchange, makes a few demands on Amazon.com, Inc., including minimum revenue amounts, stock price constraints, and quantity of shares publicly held. Being publicly traded means that AWS is beholden to Amazon.com, Inc. shareholders and is responsible for generating financial returns consistent with its bylaws and investment principles.

Federal Trade Commission

Owing to the size of AWS, its operations and acquisitions fall within the scope of antitrust laws. This authorizes the Federal Trade Commission (FTC) to investigate and prevent any anticompetitive behavior in its incipiency. This is consistent with the outlined history and potential applicability of anti-trust laws in the previous section. The FTC imposes a variety of rules that are potentially applicable to Amazon.com, Inc., broadly, but none that deal directly with Cloud computing. The Children's Online

Privacy Protection Rule ("COPPA") would apply to customers of AWS who use AWS to store personal information about children under 13 years of age (Federal Trade Commission, 2013). Similarly, the Safeguards Rule requires financial institutions to have measures in place to keep customer information secure by implementing an information security program. This rule applies to affiliates and service providers of financial institutions who may store or process sensitive information. This implicates AWS. In the enforcement of this and other rules, AWS may be called upon to provide evidence. Beyond these rules, there are also several consumer protection rules related to advertising and general sales practices that define the broader scope of doing business in the United States, but which have not been significantly formative for the organization of the cloud computing industry.

Department of Homeland Security

The Under Secretary of Information Analysis and Infrastructure Protection within the Department of Homeland Security is responsible:

> To develop a comprehensive national plan for securing the key resources and critical infrastructure of the United States, including power production, generation, and distribution systems, information technology and telecommunications systems (including satellites), electronic financial and property record storage and transmission systems, emergency preparedness communications systems, and the physical and technological assets that support such systems. (Homeland Security Act of 2002, Title II, SEC. 201, d, 5)

This means that DHS has a responsibility to work with AWS to ensure the security of their information technology systems to the extent that AWS is hosting data and

processes deemed critical infrastructure. Critical infrastructure is defined in section 1016(e) of the USA Patriot Act of 2001 (42 U.S.C. 5195c(e)) as "systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters." (The White House, 2013). While the Homeland Security Act of 2002 imparts the responsibility for DHS to work with AWS, AWS cooperation can be compelled by the Defense Production Act of 1950.

The Defense Production Act of 1950

The Defense Production Act of 1950 contains three titles relevant to the purposes of this analysis. Title I: Priorities and Allocations; Title III: Expansion of Productive Capacity and Supply; and Title VII: General Provisions. Title I empowers the president to requires persons, corporations and businesses to prioritize and accept contracts from the government in order to promote the national defense (Congressional Research Service, 2020) (The Defense Production Act of 1950, 2009). Title I has been utilized in times of crisis or disaster as well. Title I could be leveraged to prioritize availability of cloud computing for Government use, or for critical infrastructure and disaster response.

Title III of the Defense Production act of 1950 provides for the expansion of productive capacity and supply. This is accomplished through "financial incentives to develop, maintain, modernize, restore, and expand the production capacity of domestic sources for critical components critical technology items, materials, and industrial resources essential for the execution of the national strategy for the United States." (Congressional Research Service, 2020 p. 9) These financial incentives can take the form

of loan guarantees, direct loans, purchases, purchase commitments, and installation of equipment in private facilities. There are a variety of provisions and restrictions around the loan guarantees and direct loans. For direct purchases and purchase commitments, the President is authorized to make purchases up to $50 million without authorization from Congress, and that limit can be waived if the shortfall of the industrial base would severely impair the national defense. If cloud computing capabilities are determined to be critical to the national strategy or response to a crisis, then there are a variety of tools available to policy-makers to prop-up, arrange, or direct the cloud computing industry to be able to provide those services that are critical to national infrastructure.

The Critical Infrastructure Act of 2002

Furthermore, key, non-public information related to critical infrastructure is called critical infrastructure information and is governed by the Critical Infrastructure Information Act of 2002. This act facilitates the sharing of critical infrastructure information among and between the private and public sectors (Department of Homeland Security, n.d.). If AWS is determined to control or house critical infrastructure, then the metadata related to those assets or processes almost certainly becomes critical infrastructure information. Furthermore, it is possible that other providers of critical infrastructure store their information using AWS storage services and this could further impel AWS to collaborate with the relevant government agencies to share this information as needed with government agencies and other providers of critical infrastructure.

Federal Communications Commission

In 2018 AWS announced plans to expand its services to provide communications with satellites (Barr, 2018). Hosted on an Amazon EC2 instance (virtual, cloud based computer), AWS Ground processes analog signals received from satellites into digital data streams that integrate smoothly into AWS hosted computing resources. AWS Ground provides shared time on large-scale antenna stations in the same way that it provides shared time on large-scale computing infrastructure. This new business line receives operating requirements from the Federal Communications Commission (FCC) with regard to the signals used for satellite communication. The FCC "regulates interstate and international communications by radio, television, wire, satellite, and cable…" (Federal Communications Commission, n.d.). The FCC is authorized by Congress through the Communications Act of 1934 in which the Federal Radio Commission was replaced by the Federal Communications Commission. Amazon can operate listening services without the burden of FCC requirements. Listening is not a regulated activity, rather it is the transmission of communications that falls under FCC requirements. The responsibility for registering with the FCC rests with the owner of the transmitting satellites. Two-way communication with satellites would cause the activities of AWS Ground to come under greater oversight by the FCC, but based on the 2018 announcement, the main focus of this project is to provide the opportunity to simplify and expedite the receipt and processing of data collected by satellites.

Taxation

AWS is also subject to taxation by the IRS and state governments. Tax laws may influence a variety of decisions by AWS and others in the cloud computing industry. The

next section provides a cursory overview on the incentives created by taxation by type. The major types of taxation relevant to cloud computing are earnings tax, employment tax, excise, property, and sales tax.

Earnings taxes are those levied on corporations based on the profits earned in a given period. Being incorporated in Delaware, Amazon.com, Inc. is subject to federal earnings taxes, but Delaware state taxes only for "federal taxable income allocated and apportioned to Delaware based on an equally weighted three-factor method of apportionment. The factors are property, wages and sales in Delaware as a ratio of property, wages and sales everywhere (Chapter 19, Title 30, Delaware Code)." (Delaware Division of Revenue, n.d.). Delaware also has a gross receipts tax imposed on goods and services sold within the state, as well as an annual business license requirement. For the portion of profits earned outside of Delaware, Amazon.com, Inc. faces a variety of state income tax rates and apportionment methodologies based on property, wages and sales. The differences in apportionment methodologies can be seen in Appendix A (Institute on Taxation and Economic Policy, 2012). Such disparities between states can be expected to influence where AWS and other cloud providers locate their property and workforce.

Personal income tax variations by state and municipality may also influence the location of employment for cloud computing service providers because of the implications for hiring. The presence of additional earnings taxes reduces the amount of take-home pay an employee receives. All else equal, this could disadvantage an employer in a region with higher income taxes because employees may demand a higher paycheck in order to achieve a take-home pay commensurate with a non-taxed region.

The second type of tax is employment tax. At the Federal level, employment taxes have two components, FUTA and FICA. FUTA was established by the Federal Unemployment Tax Act. It represents approximately 6% of the employees eligible wages, is paid by the employer, and can be offset by contributions to state-level unemployment taxes. State level unemployment taxes or unemployment insurance differs from state to state. While companies can offset FUTA expenses with state unemployment insurance expenses, many states have unemployment insurance rates which exceed FUTA. This makes some states and municipalities more desirable than others from an employment tax perspective.

FICA stands for the Federal Insurance Contributions Act and amounts to 7.65% of gross wages paid by employee and employer for a total of 15.3%. This covers the Social Security tax and the Medicare tax (Social Security Administration, 2017). FICA is evenly applied across the states and any business implication would not be specific to the cloud computing industry.

Excise taxes are taxes paid when purchases are made on a specific good. A common example is a tax levied per gallon of gasoline sold. At the federal level, there are no excise taxes directly relevant to cloud computing (Internal Revenue Service, 2018). The communications excise tax is only relevant to local telephone service and teletypewriter exchange services. Excise taxes could be a tool for influencing the types of services and methods of delivery within the cloud computing industry.

Sales tax is a major consideration for Amazon.com, Inc., with certain sales tax provisions being commonly referred to as an "Amazon Tax" (Baugh, B., Ben-David, I., & Park, H., 2014). These provisions are a reaction to gaps in tax laws and enforcement

that allowed Amazon to sell products without collecting state or local sales tax. While retail sales tax implications may make headlines, AWS is navigating the taxation of virtual assets and services. For the most part, cloud services are taxed by the region in which they are used. For AWS users, this is the billing address on file. If multiple users at multiple sites make use of services provided by AWS, the tax obligation is split among those locations of operations. In many states and municipalities, clarifying language has been adopted into the tax laws to ensure that cloud computing is not exempt from taxes. For example, Washington State Legislature adopted ESHB 2075 in 2009 which has added language to clarify that sales/use tax applies to all digital products and the types of services provided by AWS and other cloud providers (Washington State Department of Revenue, 2010). Sales and use taxes differ between states and are more likely to impact the geographical arrangement of users of cloud services, rather than the providers.

US Department of Labor

AWS is subject to the relevant labor laws through the United States Department of Labor as well as state-specific employment laws in the states in which it operates. These laws, in addition to the employment taxes already discussed, add to the institutions which give shape to the relationship between AWS, its employees and its management. The other institutions are the corporate model, profit sharing, at-will hiring, and collective employee action.

Corporate Governance

The corporate model for AWS has been broadly outlined above and the following section will focus on the downstream implications on employees. As established, the

shareholders of Amazon.com, Inc. elect the board of directors, who then appoint the official staff. The Board may also delegate the authority to appoint officers to the existing officers or sub-committees of the board. Andrew R. Jassy is the current Chief Executive Officer of Amazon Web Services. On behalf of the board, Jassy oversees and directs the managers who run the AWS subsidiary according to the vision, guidelines, and bylaws established by the board of directors and the shareholders, and under the direction of the CEO of Amazon.com, Inc., Jeff Bezos. Jassy has a team of managers (see Appendix B & C) which oversee the operations of AWS. This management team hires staff and other managers who perform the daily operations and expansion of AWS. Amazon.com, Inc. does not disclose what portion of its 647,500 employees work on AWS, however as of September 27, 2020, on LinkedIn, there are over 70,000 people that self-report working for the subsidiary (LinkedIn, n.d.-b).

The organizational structure of AWS follows a corporate hierarchy. Within this hierarchy, insubordination is grounds for termination. Insubordination is defined as "an employee's intentional refusal to obey an employer's lawful and reasonable orders." (Society for Human Resource Management, 2018). Within this context, the default response of an employee to the requests of a superior are to comply. It then become incumbent upon the recipient of an order to assess its reasonableness and legality, at the risk of losing their job. In the new and complex legal environment of cloud computing this may not always be clear-cut, especially to employees unfamiliar with the law or new to cloud computing. Within this context, malicious actors in positions of leadership may be able to influence employees to perform actions that undermine the confidentiality, integrity, or availability of systems or data. Corporate policies as executed through the

AWS management team represent the first line of defense against insider risk. Internal compliance, human resources, and risk management teams represent the second line of defense. The Legal and Audit functions act as the third line of defense and provide the Amazon Ethics Line. The Amazon Ethics Line and EthicsPoint website allow anonymous submission of questionable practices (EthicsPoint, Inc., 2020). In this way, the collective scruples of the management team can be called into question at the scruples of individual employees. Employees making reports to their management team or this hotline are assured against retaliation for reports made in good faith. (Amazon.com, Inc., 2020a).

Another institution that could amplify insider risk is the practice of profit sharing. While profit-sharing is intended to align incentives between management and shareholders, early theorists posited that this might lead managers to target short-term gains at the expense of long-run profitability (Narayanan, M., 1985). This same dynamic could extend to the employee level as concerns security, where gaps in procedure or vulnerabilities are overlooked in order to protect the value of employee stock ownership or returns from profit-sharing, or simply to preserve their job.

One of the features of corporate hierarchy in many states is at-will hiring. This means that employees can be fired or leave with no notice and for any legal reason. The legality of a reason is determined by state and national statutes, but generally are related to the discrimination of a protected class. Lack of contracts can lead to shorter tenures or increased employee turnover. The destabilizing effects this would have would be to cause more proprietary and security-related information to leak to other firms or to malicious actors. At-will work arrangement may put communities at risk of losing employment in

the event that AWS chooses to move an entire data center or other facility as Amazon.com, Inc. has done in the past (Greene, J., 2014).

Organized Labor

Unions and collective action by employees are another institution which may affect the operations of AWS. Amazon has a history of resisting the unionization of its workforce (Kopytoff, V., 2014) (Schoolov, K., 2019). The active resistance to unionization implies at least some tendency of Amazon workers toward collective action. This collective action could have impacts even without forming a union. Nascent collective action has been seen at other large technology and cloud computing companies, but has not yet led to a widespread disruption in service.

In 2018, Google began working with the United States Pentagon on the development of artificial intelligence (Shane, S. & Wakabayashi, D., 2018). The program was called Maven, and it employed artificial intelligence to process video images. This technology could be applied to improve the accuracy and/or lethality of military drone strikes. The application of artificial intelligence, in the context of military applications generated a lot of media attention and caught the attention of Google's workforce (Wakabayashi, D. & Shane, S., 2018). When employees became aware of the potential application of their company's contract, over three thousand employees signed a letter stating that "Google should not be in the business of war" (Shane, S. & Wakabayashi, D., 2018).

A similar event occurred in 2019 when Microsoft workers protested the company's work with the United States Army to bring augmented reality to the battlefield (Lecher, C. 2019). On Sept 20, 2019 workers spanning Google, Amazon, and

Microsoft participated in a global walk-out, protesting their companies' partnerships with the oil industry (Newcomb, A., 2019). These examples highlight the possibility for employees to collectively organize to protest and resist corporate decisions. This represents a risk to government agencies that contract with large tech companies, especially those unaccustomed to working with the government. The risk is that such protests could result in disrupted operations and compromised projects.

Collective action is always a threat to systems of power and serves as a check on government power, even where strikes are prohibited by law. There are two reasons that this issue is especially poignant in the cloud computing industry, and for AWS in particular. First, the size, scale, and scope of Amazon operations amplifies the effect of a company-specific strike far beyond just cloud computing. Even within the cloud computing division, AWS, the span of impact has been demonstrated to be across industries. Bringing these varied activities under one parent company means that widespread impact can be achieved without having to coordinate collective action across multiple companies. Second, cloud computing achieves scale through automation, and this automation amplifies the impact of each employee's contribution. This means that the actions or inaction of fewer employees has a much broader impact since fewer it takes fewer employees to run a cloud computing ecosystem at scale.

Insurance

The next institution affecting AWS is the insurance industry. Consumer insurance agencies are impacted by the introduction of smart-home devices which can relay risk-relevant information to home owners as well as insurance companies. Consumers can now connect home security systems, cameras, smoke detectors and other safety devices

to their home networks and potentially to the internet and cloud providers. Amazon's partnership in this space with Travelers insurance has created concern for other insurers (Grzadkowska, A., 2018). Ultimately many of these safety devices interface with or store data in the cloud. It is possible that a disaster capable of impacting cloud availability would also undermine insurance processes built around devices that depend on cloud availability to function properly. An additional interaction with the domestic insurance industry is Amazon.com, Inc.'s partnership with JPMorgan and Berkshire Hathaway on a joint insurance plan venture known as Haven. Haven was started to improve the quality, service and cost of healthcare for employees (Thorne, J., 2019).

Insurance companies are also increasingly using cloud computing services (Crisanto, J., Donaldson, C., Ocampo, D., & Prenio, J., 2018). This subjects AWS processes to the scrutiny of insurance regulators like the NAIC. These regulators have an obligation to evaluate whether data handling practices and information security requirements imposed on the insurance companies are carried out by their third-party contractors and cloud-providers. This is similar to the way that bank regulators have an obligation to ensure that banking regulations are carried out by third-party providers. The extent to which regulators are examining cloud service providers is unclear. It is also unclear the extent to which regulators are able to verify the actual processes of cloud computing providers, given the issues previously identified with verifiability. Donaldson et al (2018) also highlight that there is a learning curve for regulators examining these institutions given the complexity of multi-tenant cloud computing services. Donaldson et al (2018) also note that further research is required to evaluate the risks of consolidation within the cloud computing industry.

Amazon.com, Inc. has also entered into India's insurance market (Olano, G. 2018). The degree to which each of these ventures leverages AWS capabilities or introduce new rules and regulations on those services is unclear. These examples reveal a complicated dynamic between the parent company and the insurance industry.

Chapter VI Conclusion

The above examples articulate the wide and varied deliveries between AWS and the social fabric within which it operates. Some regulatory relationships exist in a manner consistent with many other industries, while others are complicated by other norms and rules are self-imposed as part of the corporate charter or partially selected through the jurisdictions in which Amazon.com, Inc. is incorporated, employs, or conducts business. This returns the discussion to the original question pursued by the normative systems analysis: what are the ends toward which this system is dynamically evolving? This system is consolidating around a few very large global companies providing increasingly robust, though homogenous systems, on which an increasing portion of the economy depends.

One aspect of this evolution involves the receipts and deliveries within the highlighted region of figure 6.1. When authorizing institutions become dependent on the processing institutions they are responsible for overseeing, then the normative criteria they issue should not be expected to be free of vested interest. Furthermore, this dependency could create conditions in which AWS or another provider of crucial services is able to ignore the normative criteria imposed on them by the normative beliefs or the rules, regulations, and requirements. While not the first time that an authorizing institution is dependent on a processing institution for its own processes, this instance is

unique from other supplier/customer relationships. The government is a significant user of goods and services provided by private-sector companies, however, in the case of cloud computing, the services received are continuous, rather than intermittent, and they are difficult to understand and verify how they work. The cloud method of software provision requires complete trust and dependence on the supplier. In an on-premise model, government users have a modicum of control over their data centers and configurations, while all control in the cloud is merely virtual.

Market consolidation is predicted by the cost structures of cloud computing and compounded by the network effect. Increasing economies of scale have been observed in the costs and revenues of AWS. The market has seen consolidation around the few largest global cloud providers. The case studies have also shown how existing customers improve the eco-system of a cloud provider by sharing their tools and experience with other users.

The robustness of the cloud computing infrastructure seems to only improve as time and scale increase. AWS has an increasing number of data centers in an increasing number of locations, which diversifies the system against a range of attack vectors, both natural, criminal, and even militaristic. This diversification is offset, only partially, by the homogeneity of the technologies employed. As the industry consolidates around a few large players, these players scale their technologies through replicating the same server and data center configurations across similar network configurations, built on increasingly standardized hardware and running large-scale standardized software. With time, this software has become more modularized which adds to its resiliency, but does

not offset the massive scale that is achieved by scaling basic cloud computing technologies.

The increasing dependence on cloud computing shows no real signs of slowing. More and more industries are finding ways of safely implementing cloud technologies for their software, storage, and computing needs. Cloud computing has also created space for entirely cloud-based technology companies to operate. Brick-and-mortar companies are finding ways to build hybrid cloud strategies, including some that leverage AWS configurations for their on-premise data centers, as has been seen with the Volkswagen case study. By expanding the myriad of services offered by AWS or developed within the AWS eco-system, cloud computing is being made available to a wider and wider base of users. While beyond the scope of this study, the expansion of high-speed internet access in both wireless and wired connections, further facilitates the implementations of cloud computing solutions in contexts that would otherwise be unfeasible. This expansion of the internet, and internet of things, is expected to expand the dependence of the broader economy on cloud computing.

These ends toward which the system is evolving continues to challenge regulators. Regulators in a variety of sectors are now having to consider the implications of cloud computing on the existing laws, regulations, and practices. Technological efficiencies speed data back and forth across borders of all types, even storing them seamlessly and redundantly, limited in geographical dispersion primarily by regulatory requirements. Rules and regulations influence the cloud industry primarily through the conduit of the customers and industries now dependent on cloud technology, as there are much fewer guidelines pertaining directly to cloud technologies, themselves. More often

the cloud computing industry is impacted to the degree that cloud users in other industries influence their cloud providers.

The Federal Trade Commission has the responsibility and the power to influence the evolution of the cloud computing industry, itself. The FTC is empowered to investigate AWS or any of its competitors concerning scale, scope, practices and especially acquisitions. However, the major norm of market efficiency appears to dominate the major norm of the undesirability of monopoly. The FTC has not intervened concerning the scale of AWS or its acquisitions of suppliers and competitors. Specifically, the subnorm that government protects the public from monopolistic abuse is qualified by the term abuse, which is not observable in the current state of the cloud computing industry. As established in the previous section, abuse is not a prerequisite for investigation or preventative action by the FTC, but recent norms concerning how the FTC enforces antitrust laws indicate that until specific abuses are observed, the FTC is unlikely to interfere with or influence the formation of this industry.

CHAPTER VII

CONCLUSION AND DISCUSSION

Chapter I introduces the potential danger posed by Amazon Web Services as both critical infrastructure and a complex system. Chapter II examines the relevant literature for studies on cloud computing, complexity and risk. This revealed that complex systems with tight coupling are most prone to catastrophic failure and that AWS has experienced intermittent failures with news-worthy, though relatively minor impacts. Chapter III explores the scale and scope of AWS, finding that cost structures, network effects and coopetition have positioned AWS as a dominant leader in the cloud computing industry. The brief case studies demonstrate the uniqueness of this industry in the ways that competitors of Amazon.com, Inc are also entirely dependent on AWS for their primary operations. The case studies suggest a significant and increasing reliance on AWS, both in the breadth of industries and firms using AWS as well as the extent to which their critical processes are tied to AWS. Chapter IV explores the specific risks of malicious compromise and an overall failure to deliver service. This chapter found that insider risk is the most plausible cause of both malicious compromise and a failure to deliver; it is also difficult to assess and control this risk completely. This risk and others are expected to increase with the scale and scope of AWS. The bowtie framework in Chapter IV is useful for organizing areas of further research and risk assessment. Finally, Chapter V provides a normative systems analysis examining how normative beliefs interact with the business system in which AWS operates. While the systemic growth of AWS has many reinforcing feedback loops, there are also balancing loops that influence the system through complexity as well as tension with existing beliefs about the proper size and

scope of business. Government purchasing was highlighted as an influential policy tool that shapes the cloud computing industry. Enhancing the criteria for government purchasing can significantly alter the nature of the competition in the cloud computing industry. This was the policy tool that introduced the least disruption to the system, while encouraging more, smaller cloud providers across which the systemic risk is diversified. Finally, using the Social Fabric Matrix, the unique interdepencies among actors within the cloud computing industry are made plain. This analysis highlights the widespread coopetition among the actors, as well as barriers to effective regulation. No regulator is positioned to safeguard the public interest from developments within this industry and due to the dependence of the Federal Government on cloud providers, there is a vested interest in significant aspects of the status quo.

Discussion

The cloud computing industry and AWS continues to grow in scope, scale, complexity, and importance. Academics, policymakers, practitioners, and the general public all have an interest in understanding the complex systems that interact with this new technology and the firms that wield it. This discussion will focus on each of these stakeholders and what this dissertation contributes to the understanding.

This interdisciplinary study brings together the academic disciplines of computer science, economics, business administration, and entrepreneurship. Of these, computer science has some of the most compelling and time-sensitive issues to address. Of these, verifiability and interoperability stand out as the most pressing. Verifiability refers to the ability to verify that a black-box system like AWS is delivering the services it has promised in the way that it has promised. This technological limitation undermines trust

in the cloud. Given the increasing involvement of AWS in critical societal processes like elections there is an increased need to verify accuracy, privacy, and reliability of data and processes hosted in the cloud.

Interoperability in the case of cloud computing is the ability for processes hosted by one cloud provider to interact with and more importantly fail-over to another cloud provider. Brazeal (2020), describing a new service from AWS called EKS Anywhere, explains how this functionality could open the door to multi-cloud solutions in which AWS products are able to be run on infrastructure provided by Microsoft Azure or others. By opening the door to interoperability in this way, AWS is yet again creating the possibility for coopetition between themselves and a major competitor. The efficiency of this type of arrangement requires further study, as well as the reliability of the arrangement. If successful, interoperability could drastically change the risk landscape for cloud computing. As interoperability becomes easier and cheaper, firms will have lower-cost options for redundancy between clouds. This could potentially halve the types of risks to society outlined in Chapter IV in a way that was previously cost-prohibitive. Further study is required in order to confirm that this interoperability results in full redundancy; confirming that if the AWS cloud is down the EKS servers hosted on Azure are able to independently maintain services during that outage. If not, then resiliency could be reduced by half if an outage in *either* cloud could cause a service to fail.

The economist has much to learn from this industry and its major players. From a microeconomic perspective, the pricing of the services AWS provides would inform our understanding of market power and might anticipate future developments within the industry. At the macro-level, fluctuations in the reliability of AWS can be expected to

have an increasing impact on the broader impact and measuring this impact would inform the risk landscape described in Chapter IV. The macro-economic impact of cloud computing would be worth studying. This technology is increasingly making on-premise solutions obsolete. This obsolescence has an impact on the skills demanded by the information technology sector, it also changes the composition and arrangement of those jobs across firms in the economy. For example, an IT admin working for a corporation in an on-premise datacenter has more agency, flexibility, and very different responsibilities from an IT admin working in an Amazon-run data center. The scale of AWS and other cloud providers brings specialization that cannot be achieved at a smaller scale, drastically changing the nature of the jobs done in support of that system. These efficiencies necessarily reduce the need for labor, but also concentrates corporate earnings in a lucrative field on fewer and fewer employees. The examination of scale and scope in Chapter III also demonstrates how difficult it is, based on publicly available information, to estimate market share and the true scope and scale of AWS or any of its competitors. This represents an opportunity for researchers to determine how best to study market composition in cloud computing.

Political economists will consider the many policy choices that influence the normative business system in which AWS operates. In Chapter V, several of the four policy choices offered warrant additional research as to feasibility. The model of the system in Chapter V, represented in Figures 5.1, 5.2, and 5.3, can be tested and evaluated as well as built upon based on available data and events. The issue of complexity should be tracked to understand if the stability of the system is truly eroded by the steadily increasing suite of services provided by AWS. There are many other issues of political

economy that intersect AWS and cloud computing, but the tools developed in this study can be adapted for other contexts.

Business administration scholars as well as practitioners can glean from chapter IV that in the current environment, cloud computing offers a service with significant controls against catastrophic outages. The same analysis, however reveals that there are some risks that continue to grow with scale. This is an environment in which there is a sort of coordination failure among firms. While it is advantageous for any one firm to use a cloud service provider, it can prove problematic if all firms choose the same provider. If the technological limitations allow firms to adopt a multi-cloud approach in a cost effective way, this can be combined with a hybrid approach that achieves levels of resiliency that would previously have been extremely costly.

Coopetition, as developed in Chapter III, introduces the potential for novel business models that take advantage of the symbiotic relationships offered by AWS and others in the cloud computing space. The dynamics between competitors and the platform on which they are hosted and compete with will make for interesting case studies and challenge many of the dominant frameworks in game theory. Monitoring firms dependent on AWS to determine what long-term effects this power dynamic has on the smaller competitors.

The field of entrepreneurship would also be benefitted by considering the coopetition of AWS. This changes the risk landscape for new firm formation. As cloud services continue to be built out, many of he services hosted on AWS compete directly with services owned by AWS. This leads to two potential exit scenarios with vastly different payoffs to shareholders: AWS outcompetes the new firm and the firm is unable

to capture a sufficient share of the market, or the firm is bought out by AWS and incorporated into their service offering. Additionally, the impact of the AWS cloud environment for new firm creation, mergers and acquisitions, and new firm failure rates would give insight into how the dominance of AWS has impacted competitors for cloud services.

Returning to the research question considered: is there systemic risk in the consolidation of the cloud computing industry, particularly in the scale and scope of AWS? Yes, but the landscape changes so quickly that it is difficult to assess with any certainty. Complexity only increases with time, but this risk is at least partially offset by the modularity inherent in the design of the cloud. If the complexity of the AWS ecosystem increases geometrically with new services, then even the inherent modularity would not be enough to mitigate the risk of failure in the interactions between n services with (n-1) other services. Additionally, insider risk increases as scale forces a larger staff to oversee operations which have a wider impact if they fail or are sabotaged. AWS has extreme redundancy and a very strong track record of availability. If, through interoperability, this track record can be bolstered by Microsoft Azure or another competitor, then cloud computing may prove to be the most reliable computing system ever devised, and future studies would need to investigate whether interoperability as it is implemented is actually delivering the resiliency predicted. Amazon.com Inc.'s history of coopetition gives some encouragement that this road will be explored. What is not clear is whether interoperability will open the door to new competitors or continue to consolidate computing services in the few top firms. AWS may be too big to fail, but it seems unlikely to do so any time soon.

APPENDIX

A. State Apportionment Formulas

Table A.1.1

State Apportionment Formulas, 2012

STATE	Apportionment Formula	STATE	Apportionment Formula
AL	Double Weighted Sales	**MT**	Equal Weighted Formula
AK	Equal Weighted Formula	**NE**	Single Sales
AZ	80% Sales, 10% Payroll, 10% Property	**NV**	NO CIT
AR	Double Weighted Sales	**NH**	Double Weighted Sales
CA	Double Weighted Sales/Optional Single Sales	**NJ**	Double Weighted Sales
CO	Single Sales	**NM**	Equal Weighted Formula
CT	Double Weighted Sales	**NY**	Single Sales
DE	Equal Weighted Formula	**NC**	Double Weighted Sales
DC	Equal Weighted Formula	**ND**	Equal Weighted Formula
FL	Double Weighted Sales	**OH**	Triple Weighted Sales
GA	Single Sales	**OK**	Equal Weighted Formula
HI	Equal Weighted Formula	**OR**	Single Sales
ID	Double Weighted Sales	**PA**	Single Sales
IL	Single Sales	**RI**	Double Weighted Sales
IN	Single Sales	**SC**	Double Weighted Sales
IA	Single Sales	**SD**	NO CIT
KS	Equal Weighted Formula	**TN**	Double Weighted Sales
KY	Double Weighted Sales	**TX**	NO CIT
LA	Single Sales	**UT**	Double Weighted Sales
ME	Single Sales	**VT**	Double Weighted Sales
MD	Single Sales	**VA**	Double Weighted Sales
MA	Double Weighted Sales	**WA**	NO CIT
MI	Single Sales	**WV**	Double Weighted Sales
MN	90% Sales, 5% Payroll, 5% Property	**WI**	Single Sales
MS	Single Sales	**WY**	NO CIT
MO	Single Sales		

Source: Institute on Taxation and Economic Policy, 2012

B. The Senior Management Team over Amazon.com, Inc.

- Jeff Wilke, CEO of worldwide consumer
- Andy Jassy, CEO of Amazon Web Services
- Jeff Blackburn, SVP, business and corporate development
- Dave Limp, SVP, Amazon devices, digital management
- Brian Olsavsky, SVP & CFO
- David Zapolsky, SVP & general counsel
- Beth Galetti, SVP, human resources
- Jay Carney, SVP, corporate affairs
- Wei Gao, VP Technical Advisor to CEO

Note: As reported in Kim (2019)

C. The Senior Management Team over AWS

- Adam Bosworth, VP, AWS New Products
- Alex Yung, VP, AWS Sales China
- Ariel Kelman, VP, AWS WW Marketing
- Babak Parviz, VP, Grand Challenge
- Bing Gordon, Strategic Advisor
- Charlie Bell, SVP Utility Computing Services
- Doug Yeum, Director, Technical Advisor, AWS
- Emmett Shear, CEO, Office of CEO
- James Hamilton, VP/Distinguished Engineer, AWS
- Michael Frazzini, VP, Game Services and Studios
- Mike Clayville, VP, WW Sales & BD, AWS Commercial Sales
- Peter DeSantis, VP, AWS Global Infrastructure
- Stephen Schmidt, Chief Info Sec Officer, AWS Security
- Teresa Carlson, VP, Worldwide Pubic Sector Sales
- Werner Vogels, VP and CTO

Note: As reported in Kim (2019)

REFERENCES

Adhikari, V. K., Guo, Y., Hao, F., Varvello, M., Hilt, V., Steiner, M., & Zhang, Z. L. (2012). Unreeling Netflix: Understanding and improving multi-CDN movie delivery. In *2012 Proceedings IEEE INFOCOM, INFOCOM 2012* (pp. 1620-1628). [6195531] (Proceedings - IEEE INFOCOM). https://doi.org/10.1109/INFCOM.2012.6195531

Alizadeh, S. S., & Moshashaei, P. (2015). The bowtie method in safety management system: A literature review. *Scientific Journal of Review*, *4*(9), 133-138.

Amazon.com, Inc. (n.d.-a). *Global cloud infrastructure | regions & availability zones | AWS.* Retrieved April 13, 2019, from https://aws.amazon.com/about-aws/global-infrastructure/.

Amazon.com, Inc. (1996). *Certificate of incorporation*. Retrieved August 3, 2019, from https://ir.aboutamazon.com/corporate-governance/documents-charters/certificate-incorporation .

Amazon.com Inc. (2016). *Amended and restated bylaws of Amazon.com, Inc.* Retrieved August 3, 2019, from https://ir.aboutamazon.com/static-files/46190e86-758b-4975-bca4-3a9d56b594c9.

Amazon.com Inc. (2017a). 2017 *Amazon annual report*. Retrieved August 3, 2019, from https://ir.aboutamazon.com/static-files/917130c5-e6bf-4790-a7bc-cc43ac7fb30a.

Amazon.com Inc. (2017b, June 21). Customer Keynote: John Edwards, Central Intelligence Agency. AWS Public Sector Summit 2017. Retrieved December 3, 2019, from https://youtu.be/DEc6kVAXSs8.

Amazon.com Inc. (2017c, November 20). Announcing the new AWS Secret Region. *AWS Government, Education, & Nonprofits Blog*. Retrieved on December 3, 2019, from https://aws.amazon.com/blogs/publicsector/announcing-the-new-aws-secret-region/.

Amazon.com Inc. (2018). 2018 *Amazon annual report.* Retrieved August 3, 2019, from https://ir.aboutamazon.com/static-files/ce3b13a9-4bf1-4388-89a0-e4bd4abd07b8.

Amazon.com, Inc. (2020a). *Code of business conduct and ethics*. Retrieved on May 2, 2020, from https://ir.aboutamazon.com/corporate-governance/documents-and-charters/code-of-business-conduct-and-ethics/default.aspx.

Amazon.com, Inc. (2020b, May 21). *Amazon announces five new utility-scale solar projects to power global operations in China, Australia, and the U.S*. Press Center. Retrieved October 25, 2020, from https://press.aboutamazon.com/news-releases/news-release-details/amazon-announces-five-new-utility-scale-solar-projects-power.

Amazon.com, Inc. (2020c, July 30). *Amazon.com announces second quarter results.* Retrieved September 27, 2020, from https://s2.q4cdn.com/299287126/files/doc_financials/2020/q2/Q2-2020-Amazon-Earnings-Release.pdf.

Amazon Web Service, Inc. (2018). *Amazon Web Services: overview of security processes.* Retrieved on December 14, 2019, from https://d0.awsstatic.com/whitepapers/Security/AWS_Security_Whitepaper.pdf.

Amazon Web Services, Inc. (2019a). *Amazon SageMaker: Machine learning for every developer and data scientist.* Retrieved on December 30, 2019, from https://aws.amazon.com/sagemaker/.

Amazon Web Services, Inc. (2019b). *Authority to operate on AWS: Accelerating security and compliance certifications and authorizations.* AWS Partner Network. Retrieved on December 28, 2019, from https://aws.amazon.com/partners/ato/.

Amazon Web Services, Inc. (2019c). *AWS IoT Analytics: Analytics for IoT devices*. Retrieved on December 30, 2019, from https://aws.amazon.com/iot-analytics/

Amazon Web Services, Inc. (2019d). *AWS IoT Core: Easily and securely connect devices to the cloud. Reliably scale to billions of devices and trillions of messages.* Retrieved on December 30, 2019, from https://aws.amazon.com/iot-core/.

Amazon Web Services, Inc. (2019e). *AWS IoT Greengrass: Bring local compute, messaging, data management, sync, and ML interface capabilities to edge devices.* Retrieved on December 30, 2019, from https://aws.amazon.com/greengrass/.

Amazon Web Services, Inc. (2019f). *AWS Outpost: Run AWS infrastructure and services on premises for a truly consistent hybrid experience*. Retrieved on December 30, 2019, from https://aws.amazon.com/outposts/.

Amazon Web Services, Inc. (2019g). *Shared Responsibility Model.* Retrieved on January 25, 2020, from https://aws.amazon.com/compliance/shared-responsibility-model/.

Amazon Web Services, Inc. (2019h, March 27). *Volkswagen and AWS join forces to transform automotive manufacturing*. Business Wire. Retrieved on December 29, 2019, from https://www.businesswire.com/news/home/20190327005235/en/.

Amazon Web Services, Inc. (2020a). *Amazon Inspector service level agreement.* Retrieved on February 29, 2020. From https://aws.amazon.com/inspector/sla/.

Amazon Web Services, Inc. (2020b). *AWS service level agreements (SLAs).* Retrieved on February 29, 2020, from https://aws.amazon.com/legal/service-level-agreements/.

Amazon Web Services, Inc. (2020c). *Introduction to the AWS GovCloud (US) regions.* Retrieved on October 18, 2020, from https://aws.amazon.com/govcloud-us/.

Amazon Web Services, Inc. (2020d). *Our controls.* Retrieved on January 31, 2020, from https://aws.amazon.com/compliance/data-center/controls/.

Amazon Web Services, Inc. (2020e). *Summary of the Amazon S3 service disruption in the Northern Virginia (US-EAST-1) Region*. Retrieved February 3, 2020, from https://aws.amazon.com/message/41926/.

Ambrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). *A view of cloud computing. Communications of the ACM, 53*(4), 50-58.

Asay, M. (2018, December 7). *Capital One's 'all-in' cloud strategy is much more than a tech decision.* Tech Republic. Retrieved October 14, 2019, from https://www.techrepublic.com/article/capital-ones-all-in-cloud-strategy-is-much-more-than-a-tech-decision/.

Barr, J. (2018). AWS Ground Station – Ingest and process data from orbiting satellites. *AWS News Blog*. Retrieved on August 10, 2019, from https://aws.amazon.com/blogs/aws/aws-ground-station-ingest-and-process-data-from-orbiting-satellites/.

Bates, S., Bowers, J., Greenstein, S., Weinstock, J., & Zittrain, J. (2018). *Evidence of decreasing internet entropy: The lack of redundancy in DNS resolution by major websites and services* (No. w24317). National Bureau of Economic Research.

Baugh, B., Ben-David, I., & Park, H. (2014). *The "Amazon tax": Empirical evidence from Amazon and main street retailers*. Cambridge, MA: National Bureau of Economic Research.

Baumol, W. J. (2002). *The free-market innovation machine: Analyzing the growth miracle of capitalism.* Princeton University Press.

BeyondTrust. (2016, November 17). What is least privilege & why do you need it? *BeyondTrust Blog*. Retrieved on October 14, 2020, from https://www.beyondtrust.com/blog/entry/what-is-least-privilege#:~:text=Least%20privilege%20is%20the%20concept,to%20bypass%20certain%20security%20restraints.

Blair, R., & DePasquale, C. (2014). Antitrust's least glorious hour: The Robinson-Patman Act. *The Journal of Law & Economics, 57*(S3), S201-S215. doi:10.1086/675783

Brazeal, F. (2020). AWS just went multi-cloud… and it's only the beginning. *A Cloud Guru*. Retrieved on March 20, 2021, from https://acloudguru.com/blog/business/aws-just-went-multi-cloud-and-its-only-the-beginning

Brooks, J. (2020, October 1). *Groups urge Amazon to disclose any election data breaches.* CBSNews.com. Retrieved on March 6, 2021, from: https://www.cbsnews.com/news/election-data-breach-security-privacy-firms-demand-transparency-amazon-aws/.

Burrington, I. (2016, January 8). Why Amazon's data centers are hidden in spy country. *The Atlantic*. Retrieved on February 5, 2020, from https://www.theatlantic.com/technology/archive/2016/01/amazon-web-services-data-center/423147/.

Cancila, M., Toombs, D., Waite, A., & Khnaser, E. (2016, October 13) *2017 Planning Guide for Cloud Computing.* Retrieved from Gartner database.

Carey, S. (2016, May 27). Entrepreneur-focused OakNorth becomes first UK bank to move core systems to the cloud with AWS. *Computerworld*. Retrieved on December 20, 2019, from https://www.computerworld.com/article/3427129/entrepreneur-focused-oaknorth-becomes-first-uk-bank-to-move-core-system-to-the-cloud-with-aws.html.

Cavallo, A. (2018). *More Amazon effects: Online competition and pricing behaviors* (No. w25138). National Bureau of Economic Research.

CBS News. (2017, February 28). Amazon Web Services outage causes widespread internet problems. *CBS News.* Retrieved on September 14, 2019, from https://www.cbsnews.com/news/amazon-web-services-cloud-outage-internet-crashes/.

Celler-Kefauver Anti-Merger Act of 1950. Law Library of Congress. Public Law 81-899, 81st Congress, H.R. 2734. Accessed on June 20, 2019, from https://fraser.stlouisfed.org/title/5841.

Center for Responsive Politics. (2020a, October 23). Client profile: Amazon.com. *OpenSecrets.org.* Retrieved December 29, 2020, from https://www.opensecrets.org/federal-lobbying/clients/summary?cycle=2019&id=D000023883.

Center for Responsive Politics. (2020b, December 29). Amazon.com PAC summary data, 2019-2020. *OpenSecrets.org*. Retrieved December 29, 2020, from https://www.opensecrets.org/political-action-committees-pacs/C00360354/summary/2020.

Chen, Y., Paxson, V., & Katz, R. H. (2010). What's new about cloud computing security. *University of California, Berkeley Report No. UCB/EECS-2010-5 January*, *20*(2010), 2010-5.

Cheng, T. (2012). A Cloudy Forecast: Divergence in the cloud computing laws of the United States, European Union, and China. *Georgia Journal of International and Comparative Law, 41,* 481.

Cisco Systems Inc., (n.d.) *What is a firewall*. Retrieved on January 26, 2020, from https://www.cisco.com/c/en/us/products/security/firewalls/what-is-a-firewall.html.

Commons JR (1924) *The legal foundations of capitalism*. University of Wisconsin, Madison.

Congressional Research Service. (2020). The Defense Production Act of 1950: History, Authorities, and Considerations for Congress. Retrieved on May 7, 2020, from https://crsreports.congress.gov/product/pdf/R/R43767

Crisanto, J., Donaldson, C., Ocampo, D., & Prenio, J. (2018, December). Regulating and supervising the clouds: Emerging prudential approaches for insurance companies. *FSI Insights on Policy Implementation No 13*. Financial Stability Institute. Bank

for International Settlements. Retrieved on May 8, 2020, from https://www.centralbanking.com/media/download/31186.

Crosman, P. (2018, September 06). *Are fintechs a systemic risk?* Retrieved September 10, 2018, from https://www.americanbanker.com/news/are-fintechs-a-systemic-risk

Crouzet, N., & Eberly, J. (2018, August). *Understanding weak capital investment: the role of market concentration and intangibles*. Prepared for the Jackson Hole Economic Policy Symposium.

Daines, R. (2001). Does Delaware law improve firm value? *Journal of Financial Economics*, *62*(3), 525-558.

Daniel, C. (2018, December 1). Exploring Azure Service Fabric Mesh: A platform for building mission critical microservices. *InfoQ.* Retrieved on October 14, 2019, from https://www.infoq.com/articles/azure-service-fabric-mesh/

Dash, Eric. (2009, June 22). If it's too big to fail, is it too big to exist? *The New York Times*. Retrieved on February 15, 2021, from https://www.nytimes.com/2009/06/21/weekinreview/21dash.html?partner=rss&emc=rss.

Defense Production Act of 1950. (2009). US Code Chapter 55: Defense Production. Retrieved on May 7, 2020, from https://www.fema.gov/media-library/assets/documents/15666.

Delaware Division of Revenue. (n.d.). Doing Business in Delaware. Retrieved August 24, 2019, from https://revenue.delaware.gov/business-tax-forms/doing-business-in-delaware/

Denny, W. (2010). Survey of recent developments in the law of cloud computing and software as a service agreement. *The Business Lawyer, 66*(1), 237-242. Retrieved from http://www.jstor.org/stable/25758539.

Department of Homeland Security. (n.d.) *Critical Infrastructure Information Act*. Retrieved on August 10, 2019, from https://www.dhs.gov/publication/critical-infrastructure-information-act.

Dimson, E., & Mussavian, M. (1998). A brief history of market efficiency. *European Financial Management*, *4*(1), 91-103.

Dixon, A. (2019, May 30). America's 15 largest banks. *Bankrate*. Retrieved on October 14, 2019, from https://www.bankrate.com/banking/biggest-banks-in-america/

Drucker, P. F. (1962, March). Big business and the national purpose. *Harvard Business Review*, 40(2), 49-59. Retrieved on February 22, 2021, from https://hbr.org/1962/03/big-business-and-the-national-purpose.

EthicsPoint, Inc. (2020). *Ethicspoint – Amazon: Our commitment*. Retrieved on May 2, 2020, from https://secure.ethicspoint.com/domain/media/en/gui/44171/index.html.

Evans, J. L., & Archer, S. H. (1968). Diversification and the reduction of dispersion: An empirical analysis. *The Journal of Finance*, *23*(5), 761-767. Retrieved on April 18, 2020, from https://www.jstor.org/stable/pdf/2325905.pdf

Evans, K. (2018, November 28). Cloud native computing foundation announces Envoy graduation. *Cloud Native Computing Foundation.* Retrieved on October 14, 2019, from https://www.cncf.io/announcement/2018/11/28/cncf-announces-envoy-graduation/

Federal Communications Commission. (n.d.) *About the FCC: The FCC's mission*. Retrieved on August 10, 2019, from https://www.fcc.gov/about/overview.

Federal Risk and Authorization Management Program. (n.d.) About us. *FedRAMP.gov* Retrieved on December 28, 2019, from https://www.fedramp.gov/about/.

Federal Risk and Authorization Management Program. (2019, December 17) *Marketplace: Products*. Retrieved on December 28, 2019, from https://marketplace.fedramp.gov.

Federal Trade Commission Act (1914), US Code Chapter 2 Subchapter 1: Federal Trade Commission. Retrieved June 29, 2019, from http://uscode.house.gov/view.xhtml?req=granuleid%3AUSC-prelim-title15-chapter2-subchapter1&edition=prelim.

Federal Trade Commission. (2013). *Rules and guides*. Retrieved May 4, 2020, from https://www.ftc.gov/enforcement/rules/rules-and-guides.

Feiner, L. (2021, July 6). *Pentagon cancels $10 billion JEDI cloud contract that Amazon and Microsoft were fighting over.* CNBC Tech. Retrieved November 1, 2021, from https://www.cnbc.com/2021/07/06/pentagon-cancels-10-billion-jedi-cloud-contract.html.

Fortune Media IP Limited. (2018) *Biggest employers.* Retrieved February 5, 2019, from http://fortune.com/fortune500/list/filtered?sortBy=employees&first500.

FTC v. Sperry & Hutchinson Co., 405 U.S. 233, 92 S. Ct. 898, 31 L. Ed. 2d 170 (1972).

FTC v. Morton Salt Co., 334 U.S. 37, 68 S. Ct. 822, 92 L. Ed. 1196 (1948).

Fullwiler, S. T. (2009). The Social Fabric Matrix approach to central bank operations: An application to the Federal Reserve and the recent financial crisis. In *Institutional Analysis and Praxis* (pp. 123-169). Springer.

Galbraith, J. K. (2015). *The new industrial state*. Princeton University Press.

Garfinkel, S. (1999). *Architects of the information society: 35 years of the Laboratory for Computer Science at MIT*. MIT press.

Gartner. (2018, April 12). *Gartner forecasts worldwide public cloud revenue to grow 21.4 percent in 2018.* Retrieved August 4, 2018, from https://www.gartner.com/newsroom/id/3871416.

Gartner. (2018, September 12). *Gartner forecasts worldwide public cloud revenue to grow 17.3 percent in 2019.* Retrieved February 25, 2018, from https://www.gartner.com/newsroom/id/3871416.

General Services Administration. (2021, January 24). *FAR: Part 1 – Federal Acquisition Regulations System*. Retrieved on January 24, 2021, from https://www.acquisition.gov/far/part-1.

GitHub. (2017, May 22). Chaosmonkey: Readme.md. Retrieved on September 14, 2019, from https://github.com/netflix/chaosmonkey.

Greene, J. (2014, October 1). Amid rapid expansion, Amazon to shutter Kansas warehouse. *Seattle Times*. Retrieved August 25, 2019, from https://www.seattletimes.com/business/amid-rapid-expansion-amazon-to-shutter-kansas-warehouse/.

Greene, J. & Harwell, D. (2019, August 1). The Capital One hack couldn't have come at a worse time for Amazon's most profitable business. *Washington Post*. Retrieved on January 26, 2020, from https://www.washingtonpost.com/technology/2019/08/01/capital-one-hack-couldnt-have-come-worse-time-amazons-most-profitable-business/.

Grzadkowska, A. (2018, November 21). Measuring the scale of Amazon's threat to the insurance industry. *Insurance Business Magazine*. Retrieved August 25, 2019, from https://www.insurancebusinessmag.com/us/news/technology/measuring-the-scale-of-amazons-threat-to-the-insurance-industry-116810.aspx.

Hagel, J., Brown, J. S., & Davison, L. (2009, August 11). Why we need big organizations. *Harvard Business Review*. Retrieved on February, 15, 2021, from https://hbr.org/2009/08/why-we-need-big-organizations.

Haimes, Y. Y., Horowitz, B. M., Guo, Z., Andrijcic, E., & Bogdanor, J. (2015). Assessing systemic risk to cloud-computing technology as complex interconnected systems of systems. *Systems engineering*, *18*(3), 284-299.

Hayden, F. G. (2009). Normative analysis of instituted processes. In *Institutional Analysis and Praxis* (pp. 103-120). Springer, New York, NY.

Hersher, R. (2017, March 3). Amazon and the $150 million typo. *NPR: The Two-Way*. Retrieved on October 8, 2020, from https://www.npr.org/sections/thetwo-way/2017/03/03/518322734/amazon-and-the-150-million-typo.

Hoffman, L., Mattioli, D., & Tracy, R. (2019, August 1). Fed examined Amazon's cloud in new scrutiny for tech: visit was made in April to amazon facility in Virginia. *The Wall Street Journal: Business.* Retrieved on December 29, 2020, from https://www.wsj.com/articles/fed-examined-amazons-cloud-in-new-scrutiny-for-tech-11564693812.

Homeland Security Act. (2002) Accessed on August 10, 2019, from https://www.dhs.gov/sites/default/files/publications/hr_5005_enr.pdf.

Hovenkamp, H. (2010). The Federal Trade Commission and the Sherman Act. *Fla. L. Rev.*, *62*, 871.

Institute on Taxation and Economic Policy. (August 2012). *Corporate income tax apportionment and the "single sales factor".* Retrieved August 24, 2019, from https://itep.org/wp-content/uploads/pb11ssf.pdf.

Internal Revenue Service. (2018, March). *Excise taxes (including fuel tax credits and refunds).* Retrieved August 24, 2019, from https://www.irs.gov/pub/irs-pdf/p510.pdf.

IT Sector Coordinating Council. n.d. *IT SCC Current Members*. Retrieved October 16, 2018, from https://www.it-scc.org/current-members.html.

Janakiram, M. (2016, July 18). The master plan behind Amazon's acquisition of Cloud9 IDE. *Forbes.* Retrieved March 14, 2021, from https://www.forbes.com/sites/janakirammsv/2016/07/18/the-master-plan-behind-amazons-acquisition-of-cloud9-ide/?sh=28bff9f372c1.

Kaufman, A., & Englander, E. J. (1993). Kohlberg Kravis Roberts & Co. and the restructuring of American capitalism. *Business History Review*, *67*(1), 52-97.

Kim, E. (2019, January 23). Amazon's executive org chart, revealed. *CNBC Tech.* Retrieved May 2, 2020, from https://www.cnbc.com/2019/01/23/who-are-amazons-top-executives-2019.html.

Klazema, M. (2018, April 18). What is Amazon's background check policy? *Backgroundchecks.com Blog.* Retrieved on January 26, 2020, from https://www.backgroundchecks.com/community/Post/5299/What-Is-Amazon-s-Background-Check-Policy.

Klein, M. (2017, March 13). *Lyft's Envoy: Experiences operating a large service mesh*. [Presentation] SREcon, San Francisco. Retrieved October 5th, 2019, from https://www.usenix.org/conference/srecon17americas/program/presentation/klein

Kopytoff, V. (2014, January 16). How Amazon crushed the union movement. *Time Magazine*. Retrieved August 25, 2019 from https://time.com/956/how-amazon-crushed-the-union-movement/.

Krazit, T. (2018, April 17). *Longtime Amazon Web Services customer Netflix said to be cozying up to Google.* GeekWire. Retrieved September 14, 2019, from https://www.geekwire.com/2018/longtime-amazon-web-services-customer-netflix-said-cozying-google/

Lambert, C. (2016, August 15*). Lyft saves infrastructure costs, enables massive growth of ridesharing platform using AWS*. Retrieved October 5, 2019, from https://aws.amazon.com/solutions/case-studies/lyft/

Lecher, C. (2019, Feb 22). Microsoft workers' letter demands company drop Army HoloLens contract. *The Verge*. Retrieved August 25, 2019, from https://www.theverge.com/2019/2/22/18236116/microsoft-hololens-army-contract-workers-letter.

Levy, A. (2017, Nov 8). *Kroger is using Google and Microsoft clouds to avoid paying Amazon.* CNBC. Retrieved April 15, 2019, from

https://www.cnbc.com/2017/11/08/kroger-using-google-and-microsoft-clouds-to-avoid-paying-amazon.html.

Lift, Inc. (2019, March 1). Form S-1. Retrieved on October 5, 2019, from SEC EDGAR website https://www.sec.gov/Archives/edgar/data/1759509/000119312519059849/d633517ds1.htm

Lindsay, J. R. (2013). Stuxnet and the limits of cyber warfare. *Security Studies*, *22*(3), 365-404.

LinkedIn. (n.d.). *Amazon: Come build the future with us: People.* Retrieved September 27, 2020, from https://www.linkedin.com/company/amazon/people/.

LinkedIn. (n.d.). *Official global LinkedIn page for Amazon Web Services: People.* Retrieved September 27, 2020, from https://www.linkedin.com/company/amazon-web-services/people/

Mankiw, G. (2015). *Principle of Economics 7th Edition*. Cengage Learning.

Maughan, A. (2015, November 26). *FCA Cloud Guidance – How financial services should treat cloud services.* Retrieved December 20, 2019, from https://www.cio.co.uk/opinion/legal-briefing/fca-cloud-guidelines-how-financial-services-cloud-3630594/.

McKinnon, J. D. (2020, April 16). Watchdog finds few problems in Pentagon award of JEDI contract to Microsoft over Amazon. *Wall Street Journal*. Retrieved October 7, 2020, from https://www.wsj.com/articles/defense-department-watchdog-finds-few-problems-in-awarding-jedi-contract-to-microsoft-over-amazon-11586960694.

McLaughlin, K. (2018, April 17). Netflix, long an AWS customer, tests waters on Google Cloud. *The Information.* Retrieved on September 14, 2019, from https://www.theinformation.com/articles/netflix-long-an-aws-customer-tests-waters-on-google-cloud

Menell, P. S. (2019). Economic analysis of network effects and intellectual property. In *Research Handbook on the Economics of Intellectual Property Law*. Edward Elgar Publishing.

Microsoft. (2020, July 22). *Earnings release FY20 Q4.* Redmond, Washington. Retrieved September 27, 2020. https://www.microsoft.com/en-us/investor/earnings/FY-2020-Q4/press-release-webcast.

MITRE Corporation. (2019, August 19). *Frequently asked questions.* Retrieved January 27, 2019, from https://cve.mitre.org/about/faqs.html.

Nair, M. (2017, October 17). *How Netflix works: the (hugely simplified) complex stuff that happens every time you hit play*. Retrieved on September 14, 2019, from https://medium.com/refraction-tech-everything/how-netflix-works-the-hugely-simplified-complex-stuff-that-happens-every-time-you-hit-play-3a40c9be254b

Narayanan, M. (1985). Managerial incentives for short-term results. *The Journal of Finance*, *40*(5), 1469-1484.

Netflix. (2019, July 17). *Q2 2019 Shareholder letter.* Retrieved September 14, 2019, from https://s22.q4cdn.com/959853165/files/doc_financials/quarterly_reports/2019/q2/Q2-19-Shareholder-Letter-FINAL.pdf

Newcomb, A., (2019, September 16). Google workers to walk out, along with Amazon and Microsoft employees, for Sept. 20's climate strike. *Fortune.* Retrieved on May 8, 2020, from https://fortune.com/2019/09/16/global-climate-strike-protest-google-amazon-microsoft-walkout/.

Kissel, R., Regenscheid, A., Scholl, M., & Stine, K. (2014). *Guidelines for media sanitization*. US Department of Commerce, National Institute of Standards and Technology.

Nix, N. (2018, June 20). CIA Tech Official calls Amazon cloud project 'transformational'. *Bloomberg Technology*. Retrieved December 1, 2019, from https://www.bloomberg.com/news/articles/2018-06-20/cia-tech-official-calls-amazon-cloud-project-transformational

Olano, G. (2018, September 18). Amazon set to file for insurance license. *Insurance Business Magazine.* Retrieved August 25, 2019, from https://www.insurancebusinessmag.com/us/news/breaking-news/amazon-set-to-file-for-insurance-licence-111581.aspx.

Paul, S. (2020). Antitrust as allocator of coordination rights. *UCLA Law Review*, 378–431.

Panetta, K. (2019, October 10). Gartner offers recommendations for developing a cloud computing strategy and predictions for the future of cloud security. *Smarter With Gartner.* Gartner, Inc. Retrieved January 25, 2020, from https://www.gartner.com/smarterwithgartner/is-the-cloud-secure/.

Perrow, C. (2007). *The next catastrophe: Reducing our vulnerabilities to natural, industrial, and terrorist disasters*. Princeton University Press.

Perrow, C. B. (2008). Complexity, catastrophe, and modularity. *Sociological Inquiry*, *78*(2), 162-173.

Perrow, C. (2011). *Normal accidents: Living with high risk technologies-updated edition*. Princeton University Press.

Petrou, K. (2018, September 6). *The crisis next time: The risk of new-age fintech and last-crisis financial regulation.* Retrieved September 10, 2018, from http://www.fedfin.com/images/stories/client_reports/FedFin Policy Paper on The Risk of New-Age Fintech and Last-Crisis Financial Regulation.pdf.

Polanyi K (1957) The economy as instituted process. In: Polanyi K et al (eds) *Trade and market in early empire*. (pp. 243-270).

Poloz, S. (2018, August 25). *The Fourth Industrial Revolution and Central Banking.* [Remarks] At Jackson Hole Economic Policy Symposium. Retrieved November 1, 2021 from https://www.bis.org/review/r180827b.htm.

Posner, R. A. (1978). Natural monopoly and its regulation. *Journal of Reprints and Antitrust Law & Economics*, *9*, 767.

Rallo, A. (2018, September 24). New research from TSO logic shows AWS costs get lower every year. *TSO Logic*. Retrieved April 13, 2019, from https://aws.amazon.com/blogs/apn/new-research-from-tso-logic-shows-aws-costs-get-lower-every-year/

Reason, J. (1990) The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society* (London), series B. 327: 475-484.

Reason, J., Hollnagel, E., & Paries, J. (2006). Revisiting the Swiss cheese model of accidents. *Journal of Clinical Engineering*, *27*(4), 110-115.

Reinert, H., & Reinert, E. S. (2006). Creative destruction in economics: Nietzsche, Sombart, Schumpeter. In *Friedrich Nietzsche (1844–1900)* (pp. 55-85). Springer.

Reuters. (2021). *Companies: Amazon.com, Inc. Key metrics*. Retrieved February 15, 2021, from https://www.reuters.com/companies/AMZN.O/key-metrics.

Rockoff, H. (2004). *Until it's over, over there: the US economy in World War I* (No. w10580). National Bureau of Economic Research. Retrieved November 1, 2021, from https://www.nber.org/system/files/working_papers/w10580/w10580.pdf

Schoar, A. (2018, August 24). *Changing market structure and monetary policy: comments.* [Remarks] At Jackson Hole Economic Policy Symposium. Retrieved from https://www.kansascityfed.org/documents/6981/SchoarRemarks_Jh2018.pdf

Schoolov, K. (2019, August 22). *How Amazon is fighting back against workers' increasing efforts to unionize.* CNBC.com. Retrieved on August 25, 2019, from https://www.cnbc.com/2019/08/22/how-amazon-is-fighting-back-against-workers-efforts-to-unionize.html

Schumpeter, J. A. (2010). *Capitalism, socialism and democracy*. Routledge.

Securities and Exchange Commission. (2013). *The laws that govern the securities industry.* Retrieved on August 10, 2019, from https://www.sec.gov/answers/about-lawsshtml.html.

Shane, S., & Wakabayashi, D. (2018, April 4). Google employees protest work for the pentagon. *New York Times*. Retrieved August 25, 2019, from https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

Shead, S. (2021, January 18). *Parler's website shows signs of life but mobile apps remain offline.* CNBC.com: Tech. Retrieved March 6, 2021, from https://www.cnbc.com/2021/01/18/parlers-website-shows-signs-of-life-after-aws-fallout-.html

Sherman Anti-Trust Act (1890). *Enrolled Acts and Resolutions of Congress*, 1789-1992; General Records of the United States Government; Record Group 11; National Archives.

Shevlin, R. (2019, July 22). Banks' inevitable race to the cloud. *Forbes*. Retrieved December 20, 2019, from https://www.forbes.com/sites/ronshevlin/2019/07/22/banks-inevitable-race-to-the-cloud/#5bf79ddb1135.

Smith, A. (2013). *On the wealth of nations*. Simon and Schuster.

Social Security Administration. (2017, March). What is FICA. Retrieved August 24, 2019, from https://www.ssa.gov/thirdparty/materials/pdfs/educators/What-is-FICA-Infographic-EN-05-10297.pdf

Society for Human Resource Management. (2018, March 6). *Disciplinary issues: What constitutes insubordination?* Retrieved August 25, 2019, from https://www.shrm.org/resourcesandtools/tools-and-samples/hr-qa/pages/cms_020144.aspx

Sterman, John D. (2000). *Business dynamics: Systems thinking and modeling for a complex world.* Irwin/McGraw-Hill.

Supreme Court Of The United States. (1975) *U.S. Reports: Buckley v. Valeo, 424 U.S. 1*. [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/usrep424001/.

Sverdlik, Y. (2017, November 21). *CIA's on-prem Amazon cloud now available to other agencies.* DataCenter Knowledge. Retrieved on December 3, 2017 from https://www.datacenterknowledge.com/amazon/cia-s-prem-amazon-cloud-now-available-other-agencies

Synergy Research Group. (2018, July 27) *Cloud revenues continue to grow by 50% as top four providers tighten grip on market.* Retrieved from https://www.srgresearch.com/articles/cloud-revenues-continue-grow-50-top-four-providers-tighten-grip-market

Taber, N. (2017, June 30). Blue/green deployments with Amazon Elastic Container Service. *AWS Compute Blog.* Retrieved on February 3, 2020, from https://aws.amazon.com/blogs/compute/bluegreen-deployments-with-amazon-ecs/

Tak, B. C., Urgaonkar, B., & Sivasubramaniam, A. (2011, June). To move or not to move: The economics of cloud computing. In *HotCloud*.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder*(Vol. 3). Random House Incorporated.

Tchernykh, A., Schwiegelsohn, U., Talbi, E. G., & Babenko, M. (2019). Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. *Journal of Computational Science, 36*, 100581.

Thongtanunam, P., Tantithamthavorn, C., Kula, R. G., Yoshida, N., Iida, H., & Matsumoto, K. I. (2015, March). Who should review my code? a file location-

based code-reviewer recommendation approach for modern code review. In *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)* (pp. 141-150). IEEE.

Thorne, J. (2019, November 2). *Amazon and JP Morgan to roll out new health insurance plans as part of haven joint venture.* GeekWire. Retrieved on May 8, 2020, from https://www.geekwire.com/2019/amazon-jpmorgan-roll-new-health-insurance-plans-part-haven-joint-venture/

Tseitlin, A. (2013). The antifragile organization. *Communications of the ACM*, 56(8), 40-44.

Turecek, V. (2018, December 10). Using Envoy for data aware traffic routing in Azure Service Fabric. In *EnvoyCon 2018, Seattle, Washington*. Retrieved from https://www.youtube.com/watch?v=V33ZN7SH9m4

United States District Court for the Western District of Washington at Seattle. (2019, August 28). Indictment No. CR19-159: United States of America v. Paige A. Thompson. Retrieved on January 25, 2020, from https://www.justice.gov/usao-wdwa/press-release/file/1198481/download

US PATRIOT Act (2001), Accessed July 7, 2019. From https://www.congress.gov/107/plaws/publ56/PLAW-107publ56.pdf

U.S. Department of Defense. (2018, July 26). Contract milestone brings enterprise cloud solution one step closer to warfighter. Retrieved on October 7, 2020 from https://www.defense.gov/Newsroom/Releases/Release/Article/1584975/contract-milestone-brings-enterprise-cloud-solution-one-step-closer-to-warfight/.

U.S. Department of Homeland Security. (2016). *Information technology sector-specific plan: An annex to the NIPP 2013.* Retrieved from https://www.dhs.gov/sites/default/files/publications/nipp-ssp-information-technology-2016-508.pdf

Van Eeeten, M., & Bauer, J. M. (2012). Mega-crises and the internet: Risks, incentives, and externalities. *Mega-Crises: Understanding the Prospects, Nature, Characteristics, and the Effects of Cataclysmic Events*, 356.

Van Reenen, J. (2018, July). *Increasing differences between firms: Market power and the macro-economy*. [Remarks] At Jackson Hole Economic Policy Symposium. Retrieved November 1, 2021, from https://www.kansascityfed.org/documents/6974/VanReenenPaper_JH2018.pdf

Wakabayashi, D. & Shane, S. (2018, June 1). Google will not renew Pentagon contract that upset employees. *New York Times*. Retrieved May 8, 2020, from https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html

Washington State Department of Revenue. (2010, June 29). *Special Notice: Digital Products.* Retrieved August 24, 2019, from

https://dor.wa.gov/sites/default/files/legacy/Docs/Pubs/SpecialNotices/2010/sn_10_DigitalProducts.pdf

Weise, E. (2017, March 01). Massive Amazon cloud service outage disrupts sites. *USA Today.* Retrieved September 10, 2018, from https://www.usatoday.com/story/tech/news/2017/02/28/amazons-cloud-service-goes-down-sites-scramble/98530914/

Temin, P., & Galambos, L. (1987). *The fall of the Bell system: A study in prices and politics*. Cambridge University Press.

The White House, Office of Management and Budget. (n.d.) *The Office of Federal Procurement Policy*. Retrieved January 24, 2021, from https://www.whitehouse.gov/omb/management/office-federal-procurement-policy/.

The White House, Office of Management and Budget. (2004, February 20). *Federal Acquisition Council manager's guide to competitive sourcing,* Second Edition. Retrieved on January 24, 2021, from https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/assets/procurement_guides/fac_manager_guide.pdf.

The White House, Office of Management and Budget. (2011, July). *Improving the way our government buys: Getting the best value for our taxpayers*. Retrieved on January 24, 2021, from https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/procurement/reports/improving-the-way-government-buys.pdf.

The White House, Office of the Press Secretary. (2013, February 12). *Presidential policy directive—Critical infrastructure security and resilience* (PPD-21). Retrieved August 10, 2019 from https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil.

The White House, Office of the Press Secretary. (2019, July 5). *Remarks by President Trump at a salute to America.* Retrieved July 6, 2019, from https://www.whitehouse.gov/briefings-statements/remarks-president-trump-salute-america/

Wintermeyer, L. (2019, April 4). OakNorth is Europe's most valuable fintech and in profit: A rare breed of fintech unicorn. *Forbes.* Retrieved on December 20, 2019, from https://www.forbes.com/sites/lawrencewintermeyer/2019/04/04/oaknorth-is-europes-most-valuable-fintech-and-in-profit-a-rare-breed-of-fintech-unicorn/#26f6d1c854a2.

Yegulalp, S. (2019, January 23). *What is ISTIO? the Kubernetes service mesh explained*. InfoWorld. Retrieved on October 14, 2019, from https://www.infoworld.com/article/3328817/what-is-istio-the-kubernetes-service-mesh-explained.html.

Zhang, L. (2016). *Price trends for cloud computing services.* Honors Thesis Collection. 386. http://repository.wellesley.edu/thesiscollection/386

VITA

Andrew Douglas Cotton was born on August 17, 1988 in Springfield, Missouri. He was educated in public schools, graduating from Hillcrest High School in Springfield, MO. He enrolled at Truman State University, and in 2010 earned his Bachelor of Arts in Business Administration with a concentration in finance and a minor in economics.

Upon graduating, Mr. Cotton began his professional career working for a small broker-dealer as a high frequency stock trader, developing and refining algorithms for the purchase and sale of stocks. In this context, he developed skills in SQL and Python, working closely with database administrators and software developers to implement new trading strategies. His next venture was as an entrepreneur and consultant, starting a business and providing financial consultation for other ventures.

In 2014, Mr. Cotton joined the Federal Reserve Bank of Kansas City as a financial analyst. He served various roles in finance and information technology, including leading a team of software developers working on several payment processing applications. He has coordinated, contributed to, and delivered presentations at information technology conferences within the Federal Reserve System and for the broader industry. He served on a steering committee for Money Smart Kansas City and has volunteered and taught economic literacy classes to elementary aged students across the Kansas City Metro area. In his current role for the Federal Reserve Bank of Kansas City, he provides leadership and oversight for bank-wide strategic planning, enterprise risk management, and business continuity.