Security Breaches in the U.S. Federal Government

Min-Seok Pang [a] and **Huseyin Tanriverdi** [b]

[a] Fox School of Business, Temple University, minspang@temple.edu

[b] McCombs School of Business, University of Texas-Austin, huseyin.tanriverdi@mccombs.utexas.edu

Abstract

Cybersecurity incidents in the U.S. federal government have increased by 1,121 percent between 2006 and 2014, leading to growing concerns on the security of the federal IT infrastructures. We examine potential drivers and mitigation mechanisms of security breaches in the U.S. federal government. Technologically, many argue that the large stock of legacy IT systems in federal agencies, which are not designed for security, cause security vulnerabilities. Some IT professionals, however, counter with a "security-by-antiquity" argument that legacy systems are more secure. We consider both arguments and empirically test how legacy systems are associated with security breach incidents in the federal government. Organizationally, federal agencies exhibit significant heterogeneity; some are highly centralized whereas others are highly decentralized geographically or functionally. We examine how their organizational forms affect security vulnerability. We find that agencies that invest more in new IT development and modernization experience fewer security breaches than ones that invest more in maintenance of legacy systems. Outsourcing legacy systems to the cloud also reduces the frequency of security breaches. Our results also find that effective IT governance, risk, and control mechanisms also mitigate security risks of the legacy systems. Finally, federal agencies that are geographically or functionally dispersed experience security breaches less frequently than centralized agencies.

1. Introduction

In June 2015, the U.S. Office of Personnel Management (OPM) announced that it experienced a breach to the federal human resources database by infiltrators abroad. It revealed that more than 22 million people's sensitive personal information including Social Security numbers was leaked in this incident (*The Washington Post* 2015). A subsequent investigation by the U.S. Congress revealed that a 30-year-old mainframe system written in COBOL for the personnel database was too technically obsolete to encrypt the personal information (*Nextgov* 2015). In January 2013, a group of hackers, believed to be from China, infiltrated a system of the U.S. Army Corps of Engineers and breached the database of 79,000 dams across the U.S., which includes confidential information on potential vulnerabilities of major dams (*Free Beacon* 2013). Indeed, the IT infrastructures in the U.S. federal government are becoming more and more vulnerable. Between 2006 and 2014, cybersecurity incidents in the U.S. federal government have increased by 1,121 percent.[1] In this study, we examine potential drivers of security breaches and mitigation mechanisms for vulnerabilities in the U.S federal government. We consider both technological and organizational factors.

Technologically, there is a growing concern among federal IT officials that many decades-old legacy IT systems in the federal government have serious security vulnerabilities, as illustrated in the OPM incident above. This contrasts with a counter-argument that legacy systems might be more "secure by antiquity." Some industry practitioners and government officials argue that there is too little documentation and knowledge about the antiquated systems for attackers to discover and exploit their potential vulnerabilities. We consider both sides of the argument and conduct an empirical test to find out whether legacy IT systems are more secure than modern systems. Organizationally, U.S. federal agencies exhibit significant heterogeneity; some are highly centralized whereas others are highly decentralized

[1] http://securityaffairs.co/wordpress/38699/security/incidents-federal-government-2014.html, accessed on Mar. 7, 2017

geographically and or functionally. We examine how such heterogeneity in organization forms affects the frequency of data breaches.

This study investigates how IT investment patterns of the federal agencies (i.e., investment into the maintenance of legacy systems versus the modernization and development of new IT systems) and their organizational structures influence security breach incidents. Our unit of analysis is a federal agency such as the Department of Transportation (DOT) and the Environmental Protection Agency. We use actual security incidents data from the Federal Information Security Management Act (FIMSA) Annual Reports to Congress in FY 2012-2015. Federal IT investments data is collected from the Federal IT Dashboard. For a supplementary analysis, we also use personal information breach records from an alternative, non-governmental source - Privacy Rights Clearinghouse.

Our empirical investigations produced several intriguing findings. We find that agencies that invest more in new IT development and modernization experience security breaches less frequently than ones that invest more in maintenance of legacy systems. There is a significantly negative relationship between the number of security incidents and the stock of new IT systems, which is measured by the percentage of IT spending in new IT development over total IT investments for five prior years. It is predicted that a 1%-point increase in the share of new IT development spending is associated with a 5% decrease in security breaches. In other words, federal agencies that spend more in maintenance of legacy systems experience more frequent security incident, a result that contradicts a widespread notion that legacy systems are more secure. This effect is consistent across many different types of security breaches such as unauthorized access, social engineering, and malicious codes. A supplementary analysis with security breach data from Privacy Rights Clearinghouse shows that the amount of new IT spending is associated with fewer unintentional breaches of personal information in federal agencies. Intriguingly, federal agencies that migrate their legacy systems to the cloud suffer from fewer security breaches.

We also find that the institution of effective IT governance, risk and control (IT-GRC) mechanisms, as evaluated by agency inspectors general audits, mitigates security risks of the legacy systems. This finding indicates that security vulnerabilities caused by unsecure legacy systems could be

mitigated by strong IT-GRC mechanisms such as close monitoring of network activities, strict access controls, continuous training of employees, and effective risk management.

In addition, counterintuitively, we found that federal agencies that are more geographically dispersed experience fewer security incidents than agencies that are concentrated in one geographic location such as Washington, D.C. It appears that it is more cost-effective for infiltrators to target agencies whose information is concentrated in one area. Additionally, federal agencies that perform more homogenous functions suffer from more frequent security incidents.

This study contributes to the information systems (IS) literature by studying how IT investment patterns affect security risks in large organizations such as U.S. federal agencies. To the best of our knowledge, there are only a few studies that empirically investigate what factors affect security vulnerabilities and compromises with actual incident records (Kwon and Johnson 2014, Mitra and Ransbotham 2015, Wang et al. 2015). The literature has paid scant attention on examining whether legacy systems create security risks, even though it has been a growing concern among security professionals both in the public and the private sectors. Our contribution therefore is to defy a conventional wisdom of "security-by-antiquity" and show that the presence of obsolete legacy systems in fact aggravates security risks in large organizations such as the federal government. Toward that end, we extend economic theories of criminal behaviors and software development to explain the risks of legacy infrastructures. This paper offers a new important finding to the literature that strong IT-GRC mechanisms are crucial in mitigating the vulnerabilities from legacy systems. In addition, ours is one of the first to theorize and test the relationship between organizational forms and security incidents. We offer counterintuitive findings that organizations that are geographically dispersed and ones that perform more diverse functions are less susceptible to security threats.

This paper offers a wide range of important managerial implications for IT managers and security professionals in both governments and business organizations. Specifically, this study proposes three mechanisms for security vulnerabilities: (i) modernization of the legacy systems, (ii) outsourcing of the legacy systems to the cloud, and (iii) institution of effective IT-GRC mechanisms to reduce or avoid the

security vulnerabilities. Our study dispels a taken-for-granted assumption of security-by-antiquity and demonstrates that organizations with outdated IT infrastructures are more susceptible to security threats. Hence, it is urgent for such organizations including the federal government to invest in modernizing the legacy systems and to augment IT-GRC measures to address the risks from the legacy systems. Our finding also shows that cloud computing is an effective means to modernize the legacy systems. This study provides a crucial implication that organizations whose information assets are more concentrated are more likely to be targeted by criminals and informs that security managers at such organizations exercise extra caution in guarding their data assets.

2. Related Work and Hypotheses

2.1. Related Work

Prior research identified three major conditions that increase IT security vulnerabilities: (i) system susceptibilities such as design or implementation flaws; (ii) threat accessibility such as system access points or services; and (iii) threat capabilities such as cyber criminal's knowledge and resources to discover, access, and exploit a flaw (Brumley et al. 2008, Hughes and Cybenko 2014). Threat capabilities can be reinforced by the attractiveness of the target based on its data and functionalities.

IT vulnerability points are notable at junctions where data cross over boundaries into, out of, and between systems: e.g., data collection points, data conversion points, data communication points, data storage points, and data disposition points (Fisher 1984). Such junctions create additional work to translate the outputs of one system to match the input requirements of another system (Leifer 1989). Baskerville et al. (forthcoming) find that the integration of IT systems reduces vulnerability points but increases the potential impact of a breach. The more integrated IT systems are, the more an organization spends on cybersecurity countermeasures. The positive relationship between IS integration and cybersecurity spending was stronger in external IS integration than in internal IS integration, since external IS integration involves more vulnerability points with higher risk exposure.

Cybersecurity defenses such as IT governance, risk, and control (IT-GRC) mechanisms could potentially reduce or avoid the risks at the vulnerability points. The IS literature focuses on IT-GRC mechanisms for reducing IT investment risks, IT application development risks, IT implementation risks, IT operational failure risks, IT security risks, and IT outsourcing risks (e.g., Barki et al. 2001, Keil et al. 1998, Lyytinen et al. 1998, Weill and Broadbent 1998). To reduce the IT risks, IS researchers suggest various IT-GRC mechanisms such as aligning business strategies, IT strategies and investment objectives (Henderson and Venkatraman 1993); setting up IT governance structures (Sambamurthy and Zmud 1999, Weill and Ross 2005); choosing proper control modes to regulate individual behaviors and obtain desired behaviors in IT development and implementation projects (e.g., Kirsch 1996, 1997); establishing IT security policies, training users for IT security awareness, and enforcing security countermeasures (e.g., D'Arcy et al. 2009, Ransbotham and Mitra 2009, Straub 1990).

Most studies on security defenses and outcomes to date have used qualitative case studies. However, to the best of our knowledge, there has been a dearth of large sample empirical studies (Kotulic and Clark 2004). Ransbotham and Mitra (2009) find that more frequent information scans against information systems are associated with more incidents of targeted probes, which in turn lead to more frequent targeted attacks. Kwon and Johnson (2014) examine the impact of security investments on information breaches in the healthcare sector and find that proactive security investments have stronger impacts than reactive security investments that are made after security incidents. Wang et al. (2015) study the relationship between the characteristics of information system applications and their vulnerabilities in the financial industry and find that applications with higher value, lower inertia, higher visibility, and higher accessibility experience more frequent unauthorized access attempts. Mitra and Ransbotham (2015) compare the security impacts of full and limited disclosures of potential vulnerabilities and find that the full disclosure of a vulnerability increase the speed in diffusion of attacks that exploit the vulnerability but reduce the lifespan of the vulnerability.

2.2. Economics of Criminal Behaviors

The criminology literature posits that criminals act rationally in an economics sense; in other words, they commit a crime only if potential gains from it outweighs expected or perceived costs of committing the crime and risks of punishment (Goel and Rich 1989, Ehrlich 1996, Machin and Meghir 2004). To put it differently, a crime occurs if and only if

$$E(G) \geq E(C) + P(A)\,E(P) \quad (1)$$

where G, C, and P indicate the expected gains, the expected costs, and the punishment, respectively, and $P(A)$ refers to the probability of getting caught and convicted for the crime

This is the case in information security as well. In planning an attack, cyber-criminals take into accounts; (i) the costs (C) associated with planning attacks, discovering security weaknesses, and executing intrusions; (ii) the risks (P) of getting punished for the crime; and (iii) the returns (G) from successfully infiltrating a system and obtaining information that otherwise could not have been acquired. The gains (G) do not need to be monetary. Criminals may pursue fame or reputation among their peers from pulling off high-profile security incidents, which could lead to future employment opportunities for them. Effective IT-GRC measures reduces security vulnerabilities by increasing either the costs of penetrating a system (C) or the chance of identifying and arresting the perpetrators (A). By doing so, an organization can dis-incentivize potential penetrators from executing attacks on its systems. The assumption that a hacker maximizes her payoff is widely used in the IS literature (e.g. Cavusoglu et al. 2008, 2009).

Drawing upon this economic theory, Ransbotham and Mitra (2009) put forth a conceptual model of security compromises. This model proposes that target attractiveness, active and passive Internet presence, and effectiveness of countermeasures contribute to security compromises. Information systems that are attractive to attackers in terms of tangible, iconic, and reprisal values (i.e. higher G in Eq. 1) are more likely to experience security probes and attacks more frequently. Likewise, it is less costly to find vulnerabilities in an organization with larger Internet presence (lower C), leading to more frequent security incidents. Ransbotham and Mitra (2009) further theorize that more effective countermeasures

that consist of the controls in access, vulnerability, feature, traffic, and audit can alleviate the security risks from high attractiveness and larger Internet presence, as such countermeasures increases the costs or the risks that would be borne by attackers.

2.3. IT Investments into Legacy versus Modern IT Systems and Security Vulnerabilities

Whether or not legacy IT systems are more vulnerable to security threats than modern systems is a matter of continuing debates. While some argue that old legacy systems might have more vulnerabilities such as design or implementation flaws, others coin the term "security by antiquity" to argue that threat capabilities could be lower, due to cybercriminals' lack of knowledge and resources to discover, access, and exploit the potential flaws in antiquated legacy systems. Notably, Rep. Darrell Issa in the U.S. House Oversight and Government Reform Committee reportedly said that systems written in COBOL are "pretty much hack-proof" and "most hackers are not even old enough to understand the language" (*FCW* 2013). IT practitioners have expressed a similar opinion; one commented that "*"Newer" does not equal "better". There's a lot of "obsolete" COBOL that is handling your money, and a good deal of that code resides in comparatively unhackable environments. Code doesn't wear out, so it keeps on functioning. This is not a bad thing.*"[2] Another stated "*New systems are nearly useless. The "legacy" systems have been secure for years. The new systems get broken into every other day*."

Legacy systems could be more secure than newly-developed systems for several reasons. First, many decades-old legacy systems are relatively isolated from external networks, thereby reducing threat accessibility (Brumley et al. 2008, Hughes and Cybenko 2014). Due to limited system access points and services, it might be more difficult for cybercriminals to access and exploit the systems. Second, most of the legacy systems were developed with old programming languages or development tools such as COBOL and run over antiquated hardware systems such as mainframes. Many hackers in the current generation are unfamiliar with these technologies. Third, legacy systems are often undocumented or

[2] http://comments.us1.gigya.com/comments/rss/6407831/cw/3103585, accessed on May 16, 2017

poorly documented. How they were originally designed, developed, revised and maintained over the years remains to be obscure because many of those who developed the systems are retiring (*CIO* 2016). Legacy systems that are in working condition are often treated as black boxes. Hence, even if cybercriminals are willing to invest in learning the legacy systems, there is little they can discover and the costs entailed in discovering the flaws and vulnerabilities in the legacy systems could be very high (high C in Eq. 1).

On the other hand, there are many other reasons to expect legacy systems to be more susceptible to security threats. First, legacy systems have possibly accumulated a large amount of sensitive information over the years or decades. Thus, they are attractive targets, as they carry highly tangible value for an infiltrator (high G in Eq. 1, Ransbotham and Mitra 2009). For instance, the personal records of 22.1 million that were leaked from the federal personnel database in 2015 reportedly includes ones from 1985 (*The Washington Post* 2015, *Reuters* 2015). The U.S. Internal Revenue Service (IRS) still maintains the Individual Master File, which was developed 56 years ago with Assembly language code, but it still processes income tax filings and refunds of all American taxpayers (*Nextgov* 2016, *FCW* 2015). This system is a frequent target of security attacks (*Associated Press* 2015).

Second, the legacy systems that were designed and developed decades ago are very unlikely to have strong security features from the beginning, since awareness and knowledge of security defenses were limited at that time. Even if they had some security defenses, such features are unlikely to match the increasing sophistication of more recent and newly emerging security threats. For instance, the mainframe systems might not have a well-designed authentication system that closely monitors and deters malicious access attempts. They may not have strong identity governance and access management capabilities to manage access credentials of tens of thousands of employees and segregate potentially conflicting access privileges. In addition, because such systems are unlikely to have proper documentation and there might be few employees who know the systems well, they might not have been properly maintained or "patched" with new security features. Hence, it is difficult to apply effective countermeasures to legacy systems (Ransbotham and Mitra 2009). Such deficiencies in security protection reduce the costs of executing security attacks (C in Eq. 1). One way to go around this vulnerability is to treat a legacy system

as a black box and attach security protections around it. However, such an approach might exacerbate the complexity of the overall security infrastructures, exposing more security weaknesses.

Third, from an enterprise architecture perspective, an organization with a large stock of legacy systems is likely to have complex enterprise architectures. Under such a complex architecture, the systems are not tightly integrated but loosely connected with each other, making the overall systems more prone to security threats (Barkerville et al. forthcoming). Old legacy systems are likely to communicate with each other via connectors or converters, and information flows through such connectors can be easily intercepted by infiltrators. Such complexity thereby reduces the costs (C in Eq. 1) in executing a penetration. All these components in the complex enterprise architectures are less likely to be thoroughly documented, making it difficult to maintain security protection across the architecture up to recent emerging security threats. Such security holes reduce the execution costs (C in eq. 1) for hackers.

Fourth, the IS literature on software development and maintenance demonstrates that software quality is related to software complexity and development process maturity. Banker and Slaughter (2000) show that a software application that has higher data complexity or is updated more frequently is likely to have more defects. Harter et al. (2000) also found that software design complexity, measured by domain, data, and design complexity, is positively associated with defect density. Subramanyam et al. (2012) show that the number of data elements and data layer interfaces in a software component is positively associated with the number of component defects. These studies characterize that the more modified a system has been, the more entropy or complexity is accumulated in it. Whenever the system is modified with a new business requirement, new data elements or new interfaces layers are added to it. In each change with additional requirements, decision paths in an application component are altered, additional calculations and algorithms are added, and business logics become more complicated (Harter et al. 2012). Each of such changes exacerbates the complexity in software applications, increasing the likelihood of software failures. It is because these updates are unlikely to be done with careful consideration for coupling with other parts of the systems and ripple effects of the changes on them. It is highly unlikely

that maintenance activities on decades-old legacy systems are thoroughly documented, increasing the odd of software malfunctions.

A security breach is more likely to occur when software is more defective and error-prone. Defects in identification and access control components create security holes to an entire system. Attackers also can take advantage of malfunctions in any part of the system as a security hole (i.e. lower costs in attack execution). Hence, complexity and entropy accumulated throughout modifications in the legacy systems make them more defective and thereby more susceptible to security exploitation (Mitra and Ransbotham 2015). It is unlikely that software vendors can provide patches for vulnerabilities in old legacy systems, if the vendors even still exist (Ransbortham et al. 2012). The software development literature also argues that software structure and development process maturity are significantly related to higher software quality and fewer defects (Krishnan and Kellner 1999, Banker and Slaughter 2000, Harter et al. 2000, 2012, Krishnan et al. 2000). It is unlikely that the legacy systems were developed decades ago with a mature, structured development process.

Hence, while there are a few reasons to support the security-by-antiquity argument, the foregoing explanations with the perspectives of enterprise architecture, software defects, and target attractiveness suggest that legacy systems are likely to be less secure.

Hypothesis 1. *A federal agency with more legacy systems is likely to experience more frequent security breaches.*

Security risks posed by legacy systems could potentially be mitigated by strong IT governance, risk, and control (IT-GRC) mechanisms. Management control theory posits that formal and informal governance mechanisms could reduce the probability and magnitude of loss associated with risks (Simons 1991). Informal mechanisms include leadership, culture, values, and norms (Macintosh 1994), while formal mechanisms include agreements and assumptions about an organization's objectives and risks that could potentially inhibit their achievement (Goold and Quinn 1990). Managers institute formal GRC mechanisms to provide reasonable assurance for the achievement of the objectives and the minimization

of the risks. The scope of GRC mechanisms usually covers: (i) effectiveness and efficiency of operations, (ii) reliability of financial reporting, and (iii) compliance with applicable laws and regulations (COSO 1992). After designing and implementing the GRC mechanisms, managers also continuously monitor their operating effectiveness (Eisenhardt 1985, Ouchi and Maguire 1975). If there are deviations from the desired objectives, managers intervene by imposing sanctions, redesigning the GRC measures, or changing the objectives (Goold and Quinn 1990).

Consistent with the management control theory, IS research and practice recommend IT-GRC mechanisms for mitigating IT-related risks in the general computing infrastructure of an organization and the IT applications that automate and support the organization's business processes (IIA 2008, 2009, 2012). The organization institutes IT-GRC mechanisms to defend against security threats and other IT-related risks such as digital frauds, operational IT glitches, and incompliance with relevant laws and regulations.

The security vulnerabilities caused by complex enterprise architectures around the legacy systems can also potentially be alleviated by effective IT-GRC mechanisms such as security countermeasures and policies, close monitoring of network activities, and training of employees and contractors. Ransbotham and Mitra (2009) propose that effective vulnerability controls and feature controls can weaken chain reactions from information scans to targeted probes to targeted attacks. Likewise, effective traffic and access controls alleviate the risk of security compromises from attack scans and targeted attacks. Well-executed vulnerability controls on complex enterprise architectures enables an organization to properly patch security holes in the legacy systems on a constant basis. Effective access controls and policies enable the organization to closely monitor who has access privileges to the legacy systems and quickly detect those who violate access controls, increasing the risks for potential perpetuators (A in Eq. 1). Well-established feature controls document complete, accurate, and up-to-date configurations of the complex enterprise architectures, enabling security holes due to configuration deviations to be fixed quickly. In sum, effective IT-GRC measurse increase the costs in executing security attacks as well as the risks of apprehension, leading us to propose the following moderation hypothesis.

Hypothesis 2. *The impact of federal legacy systems on security breaches is weakened by strong IT governance, risk, and control mechanisms.*

We argue that an organization is likely to experience fewer security incidents when it outsources legacy IT systems to the cloud. Some argue that outsourcing to the cloud could create a single point of failure and pose more security risks than keeping on-premise legacy IT systems. If the cloud vendor is compromised, all information of the organization could be exposed. While such an incident, if occurs, would indeed be a high impact event, its probability is low (Baskerville et al. forthcoming).

We argue that the cloud has several security advantages vis-à-vis on-premise legacy systems as follows. A federal agency is unlikely to have more advanced IT-GRC capabilities than cloud computing vendors in the private sector. The cloud vendors can attract and retain more talented security professionals than the federal agencies, which have lower and more rigid salary scales for their employees. The cloud vendors can achieve economies of scale in security management for a range of systems that serve thousands of clients. For instance, it would be costly for a federal agency to constantly improve its on-premise security infrastructures in response to emerging threats and newly discovered vulnerabilities, while it is more seamless for a cloud vendor to address new threats across the systems it hosts on a continuous basis. Since the cloud vendor serves a large number of clients, it integrates, standardizes, and centralizes its IT systems and processes. As Baskerville et al. (forthcoming) find, organizations that integrate their IT systems to a greater extent can address more security vulnerability points.

The cloud vendor can enjoy economies of learning scope advantages as well, in a way that when it discovers a new security vulnerability from one system (Mitra and Ransbotham 2015), it can apply a patch for it across all the systems it hosts. To be hosted at a cloud, a client's system needs to adhere to several technical standards that the cloud vendor imposes, and such standards make it more straightforward to protect standardized systems from external security threats. This also implies that because of the technical standards at the cloud, migrating legacy systems to the cloud is likely to entail modernizing the system via reengineering, modularization, and standardization (Tanriverdi et al. 2007). If

a system is hosted at a cloud outside of a federal agency's premise, it is easier to control access and authentication, reducing the possibility of security compromise by insiders. As such, we put forth a negative relationship between cloud computing spending and security vulnerabilities.

Hypothesis 3. *A federal agency that spends more in cloud computing is likely to experience less frequent security breaches.*

2.4. Organizational Form and Security Vulnerabilities

Next, we hypothesize how organizational forms affect security risks. Federal agencies exhibit significant variance in terms of how they are organized. Some federal agencies such as the Department of Homeland Security have employees that are dispersed around the country while most personnel at some other agencies such as the Department of Education and the National Science Foundation are centralized around the Washington, D.C. area. We argue that such federal agencies that are geographically concentrated are more susceptible to security threats.

On the one hand, the fact that subunits and employees of a federal agency are dispersed across many different geographic locations may increase security risks. There are more vulnerability points where data cross over boundaries into, out of, and between systems of organizational subunits (Fisher 1984). The agency has more work to translate the outputs of one system to match the input requirements of another system (Leifer, 1989). Maintaining secure environments in many dispersed locations would be more expensive. It would also be more challenging to authenticate employee access and prevent malicious attempts from or to remote offices. In order to do so, the security infrastructures would be more complex and require more layers and controls, adding more vulnerabilities to the overall enterprise architectures. It would also be less costly to guard the systems in one or few locations than in many dispersed locations. In terms of Ransbotham and Mitra (2009), an agency with more geographic footprints is likely to have more Internet presence, either active or passive, posing more security risks.

On the other hand, federal agencies that are geographically more centralized would likely be less secure. It is because it would be economically more viable for a criminal to target more geographically

concentrated agencies, where valuable information is likely to center around a few locations as well. Hence, the potential gains (G in Eq. 1) from executing security breaches against a geographically concentrated agency are higher than against one that is dispersed across the country. In the model of Ransbotham and Mitra (2009), such an agency carries higher attractiveness to potential infiltrators. It would also be costlier for cybercriminals to try to penetrate the systems on many different locations (high C in Eq. 1), and each attempt to intrude the systems at multiple locations increases the risks of getting apprehended (high A in Eq. 1). Hence, hackers are more likely to seek to breach federal agencies that are more geographically concentrated. Based on this argument, we propose the following.

Hypothesis 4. *A federal agency whose employees are more geographically scattered is likely to experience less frequent security breaches.*

In a similar vein, we posit that a federal agency that performs more diverse functions is less vulnerable to security threats. Many federal agencies are commissioned with a more variety of functions than their name might suggest. For example, the Department of Agriculture carries out such functions as agricultural research and farmland conservation, but it also operates a large scale of welfare programs such as food banks and income supports for farmers. The Department of Energy performs national defense activities related to nuclear energy programs. In addition to its law enforcement duties, the Department of Homeland Security (DHS) devoted more than 32% of its budgets to disaster relief and restoration programs in 2014, such as temporary housing, emergency foods and shelter, or flood insurances.

There is a reason to believe that a federal agency that performs diverse functions could be more vulnerable to security threats. Such an agency is likely to operate a variety of systems with diverse functionalities, which may expose more vulnerability points. We propose, however, that a federal agency with more diverse functions is less vulnerable to security attacks for the same reason with H4. If an agency performs homogenous functions, its information is more likely to be concentrated to a few systems, which can be an attractive target for cybercriminals (with high G in Eq. 1). On the other hand,

for a federal agency with diverse functions, its information is more likely to be scattered over multiple locations and systems, making it costly for an attacker to attempt a breach (high C in Eq. 1). For this reason, we propose a negative relationship between the agency functional diversity and security susceptibility as follows.

Hypothesis 5. *A federal agency that performs more diverse functions is likely to experience less frequent security breaches.*

Table 1. Chief Financial Officers (CFO) Act Agencies

Department of Agriculture
Department of Commerce
Department of Defense
Department of Education
Department of Energy
Department of Health and Human Services
Department of Homeland Security
Department of Housing and Urban Development
Department of Justice
Department of Labor
Department of State
Department of the Interior
Department of the Treasury
Department of Transportation
Department of Veterans Affairs
Environmental Protection Agency
General Services Administration
National Aeronautics and Space Administration
National Science Foundation
Nuclear Regulatory Commission
Office of Personnel Management
Small Business Administration
Social Security Administration
U.S. Agency for International Development

Table 2. Definition of Security Incidents

Incident Type	Definition
From Federal Information Security Management Act (FISMA) Reports to Congress (as defined by the U.S. Computer Emergency Readiness Team)	
Denial of Service	Successful DoS incidents, such as a flood of traffic, which render a web server unavailable to legitimate users
Improper Usage	Incidents in which a user violates acceptable computing policies or rules of behavior (including *Unauthorized Access* and *Policy Violation*)
Unauthorized Access	Incidents in which individual gains logical or physical access without permission to a Federal agency network, system, application, data or other resource (including *Social Engineering* and *Unauthorized Equipment*)
Social Engineering	Incidents involved with fraudulent web sites and other attempts to entice users to provide sensitive information or download malicious code (including *Phishing*)
Unauthorized Equipment	Incidents involved with lost, stolen or confiscated equipment, including mobile devices, laptops, backup disks or removable media
Policy Violation	Incidents of mishandling data in storage or transit, such as personally identifiable information (PII) found unsecured or emailed without proper encryption
Malicious Code	Successful executions or installations of malicious software, which are not immediately quarantined and cleaned by preventative measures such as antivirus tools
Non-Cyber Incidents	Incidents of PII spillages or possible mishandling of PII, which involve hard copies or printed material
From Privacy Rights Clearinghouse	
Total Breach	Incidents of breaches of personal information including Social Security numbers, account numbers, and driver's license numbers
Intentional Breach	Breaches via (i) electronic entries by an outside party, malware and spyware or (ii) by someone with legitimate access to intentionally breach information
Unintentional Breach	Breaches by unintended disclosure or via loss of electronic devices (portable, stationary) or physical documents

3. Data Sources and Empirical Methods

We use three data sources for our empirical analyses – Annual FISMA Reports to Congress in FY 2012-2015,[3] Privacy Rights Clearinghouse, and the Federal IT Dashboard. The FISMA Reports provide

[3] The FISMA reports before FY 2012 do not publish agency-level incident counts, and in FY 2016, the OMB revised the reporting guidelines, in which agencies do not need report security incidents that did not have an impact on agency operations. The FISMA 2016 report states that the 2016 data is not comparable to previous years. Indeed, the FISMA 2016 report tallies 30,899 incidents across the whole federal government, while the 2015 report counts 77,183.

the number of security incidents occurred at 24 federal agencies that are under purview of the Office of Management Budget (OMB) by the Chief Financial Officers (CFO) Act (Table 1). These "CFO Act" agencies include all cabinet departments such as the Department of Defense and the Department of Justice and several independent agencies including the National Science Foundation. Table 2 provides the types of security incidents tallied in the FISMA Reports. We use the log-transformed number of security incidents in the reports as the dependent variables.

In addition to the FISMA reports, which are available only in 2012-2015, we collected security breach data from Privacy Rights Clearinghouse in 2005-2016. It provides comprehensive records of breaches of personally identifiable information (PII) in business and government organizations that it has independently collected from media, law enforcement, and other online sources. We collected PII breach incidents occurred at federal agencies since 2005. As with the FISMA, we use the log-transformed number of PII breach incidents that occurred in each CFO agency as a dependent variable. In addition, we categorize incidents into two types – intentional breaches (incidents via hacking, malware, or by insiders) and unintentional incidents (including unintentional transmission of PII and loss of computers, physical devices or physical documents) – and investigate how our antecedents affect the two types differently.

The Federal IT Dashboard provides IT investment figures of the CFO Act agencies since FY 2003. It reports the amount of IT spending in development, modernization, and enhancement (DME)[4], and the rest of IT spending is devoted to operation and maintenance of existing IT systems. Our primary measure, DME_t,is the percentage of DME spending over total IT spending for five prior years (FY t-1 ~ t-5) as a proxy of the stock of new IT systems (Table 3). In the section to follow, we will present estimations with a three-year and a seven-year window, instead of five, as robustness checks. The Federal IT Dashboard for FY 2012-2015 also offers spending figures in cloud computing in the CFO Act

[4] DME investments include "costs for projects leading to new IT assets/systems and projects that change or modify existing IT assets to: substantively improve capability or performance; implement legislative or regulatory requirements; or to meet an agency leadership request."
http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/fy13_guidance_for_exhibit_300_a-b_20110715.pdf

agencies. $Cloud_t$ is the percentage of cloud computing spending over total IT expenditures in FY *t*. H1 and H2 predict that the coefficients of DME and Cloud are negative.

Table 3. Variable Definitions and Data Sources

Variable	Definition	Data Sources
Independent Variables		
DME_t (Development, Modernization, and Enhancement)	% of DME spending in FY *t*-1 ~ *t*-5 to total IT spending in FY *t*-1 ~ *t*-5	Federal IT Dashboard
$Cloud_t$	% of cloud computing spending in FY *t* to total IT spending in FY *t*	
Disperse (geographic personnel dispersion)	$1 - \sum$ (Agency FTE in state i / Total agency FTE)^2	Federal HR Database
Diversity (functional diversity)	$1 - \sum$ (Agency budget in function i / Total agency budget)^2	Budget of U.S. Government
IG Score	IT GRC scores evaluated by agency inspectors general (max 100)	FISMA Reports to Congress
Control Variables		
IT Invest	% of IT spending to total agency budget	Federal IT Dashboard
Security Invest	Log (Spending in cybersecurity in million $)	FISMA Reports to Congress
Budget	Log (Total agency budget in million $)	Budget of U.S. Government
Bureau	$1 - \sum$ (Budget in subordinate bureau i / Total agency budget)^2	
FTE	Log (Total full-time equivalent employee)	Federal HR Database
Age	Agency age in years	Federal Register
Defense	% of national defense function budget to total agency budget	Budget of U.S. Government
Welfare	% of public welfare function budget to total agency budget	
Law Enforce	% of law enforcement function budget to total agency budget	
Management	% of government management function budget to total agency budget	
Regulation	% of regulatory function budget to total agency budget	

Table 4. Descriptive Statistics

Variable	*N*	Mean	Std. Dev.	Minimum	Maximum
Incidents from FISMA Reports					
Total Incidents	96	1964.219	2566.310	3	11263
Denial of Services	96	4.260	9.561	0	59
Improper Access	96	1027.104	1326.784	1	4988
Unauthorized Access	96	551.146	756.528	0	2766
Social Engineering	96	159.146	342.637	0	1755
Unauthorized Equipment	96	367.490	575.655	0	2306
Policy Violation	96	440.958	696.811	0	2563
Malicious Code	96	338.677	496.683	0	1900
Non-Cyber Incidents	96	594.177	1283.127	0	4877
Incidents from Privacy Rights Clearinghouse					
Total Breach	106	0.349	0.769	0	5
Intentional Breach	106	0.198	0.506	0	3
Unintentional Breach	106	0.104	0.336	0	2
Independent Variables					
DME	106	22.817	11.945	5.319	55.074
Cloud	106	5.874	12.945	0	99.826
Disperse	106	0.798	0.233	0.016	0.981
Diversity	106	0.428	0.318	0	0.999
IG Score	91	73.978	19.789	19	99
Control Variables					
IT Invest	106	0.009	0.019	0	0.157
Security Invest	106	3.915	2.050	0	9.397
Budget	106	17.186	2.037	12.802	21.066
Bureau	106	0.371	0.323	0	0.888
FTE	106	10.002	1.454	7.262	12.809
Age	106	90.047	63.255	10	240
Defense	106	3.511	15.916	0	100
Welfare	106	29.092	40.840	0	100
Law Enforce	106	4.832	17.670	0	85.218
Management	106	9.282	26.123	0	100
Regulation	106	21.025	34.802	0	100

The FISMA Reports also provide the assessment in the effectiveness of IT-GRC mechanisms by agency inspectors general (IG). They assess the effectiveness of their agencies' IT-GRC mechanisms in ten management areas – continuous monitoring management, configuration management, identity and access, incident response and reporting, risk management, plan of action, security training, remote access, contingency planning, and contractor systems. The IGs are required to use 91 reporting metrics created by

the DHS Office of Cybersecurity and Communications. For example, for configuration management, an IG assesses whether the agency documents policies and procedures for configuration management and whether it documents proposed and actual changes to hardware and software configurations. For identity and access management, the IG evaluates, among others, whether the agency ensures that accounts are terminated or deactivated once access is no longer required. The FISMA reports in FY 2012-2015 offer the agency IG scores in the range of 0 to 100, which we use as a measure for IT GRC effectiveness.

Geographical dispersion and functional diversity of federal agencies are measured as follows. We use the federal human resource database from the OPM, which provides data on how many federal employees work in each of the 50 states and Washington, D.C. Using this information, we measure the geographic dispersion of federal agencies (Disperse) with an inverse of Herfindahl index ($1 - \sum$ (Agency employees in state i / Total agency employees)^2). The smaller this value is, the more geographically concentrated a federal agency is. We also measure the functional diversity of federal agencies in a similar way. In the federal government budgets published by the OMB, each of the budget items is designated to one of the functional categories such as national defense, energy, transportation, and general government administration. Using these budget functional categories, we calculate an inverse of Herfindahl index of agency budget ($1 - \sum$ (Agency budget in function i / Total agency budget)^2). Again, a smaller value of this measure (Diversity) indicates that an agency performs more homogeneous functions. H4 and H5 predict that the coefficients of Disperse and Diverse are negative, respectively.

Table 3 provides the descriptions of the explanatory and control variables. We control for the overall size of IT investments and security-related spending. We also control for the scale of agencies with the size of agency budgets and the number of full-time equivalent employees. Many cabinet agencies have sub-agency bureaus (e.g. the Drug Enforcement Administration under the Department of Justice, the Food and Drug Administration under the Department of Health and Human Services). We control for the bureau-level organizational complexity with a Herfindhal index of bureau budgets. Since older agencies are likely to have more legacy systems, we control for the ages (years) of federal agencies. Lastly, security vulnerabilities could be associated with agency functions. For instance, agencies that perform

defense or law enforcement responsibilities are more likely to be targeted by cybercriminals. To account for this possibility, our estimations include five indicators for agency functions (Table 3). Table 4 offers the descriptive statistics.

As a primary estimation approach, we use Driscoll and Kraay (2006) fixed-effects estimation for spatial autocorrelation, which accounts for agency-specific unobserved heterogeneity. Since the federal agencies are under the umbrella of the U.S. government, unobserved heterogeneity of a federal agency could be contemporarily correlated with other peer agencies. The estimations also include year fixed-effects and temporal autocorrelation in residuals for two lagged years. Since the dependent variables are count variables (the number of security breaches), we will present the results of Poisson fixed-effects regressions as a robustness check below.

In our estimation, endogeneity is unlikely to be of concern. Security incidents in year t are unlikely to significantly affect investment decisions in new IT development in the previous five years. In addition, migration of systems in cloud computing is more likely to be motivated by cost savings. The geographic dispersion of agency workforce or the diversity of agency functions are likely to be exogenous as well, as it is challenging for a federal agency to move around its personnel across the country or change its responsibilities mandated by Congress in a short term. We do not have a reason to believe that IT investments, federal budgets, or agency headcounts are measured with significant errors that are correlated with residuals. In addition, our fixed-effects estimation accounts for agency-specific unobserved heterogeneity; hence, omitted variable bias is not a serious concern.

4. Results

Table 5 presents our estimation results. It shows that the coefficient of DME (development, modernization and enhancement) is negative and significant in all but two columns. This indicates that fewer security breaches occurred at federal agencies that spent more in new IT development. It is consistent with Hypothesis 1 that the legacy systems are more vulnerable to security failures than more

recently-developed systems. From the coefficient of DME in Column 1, we calculated that a 1%-point increase in DME spending over total IT spending is associated with a 5% decrease[5] in total security breaches. The impact of DME is strongest for social engineering incidents (Column 5), which include phishing; a 1%-point increase in DME is related to a 9% reduction in social engineering incidents. Likewise, DME spending is negatively related to security breaches involved with malicious codes (Column 8). While incidents of social engineering and malicious codes might primarily target on personal computing devices rather than enterprise systems, given that the FISMA reports tally all *successful* security breaches, these results demonstrate that legacy systems are vulnerable to breaches conducted via personal devices. It is also interesting to see that investments in DME are negatively associated with policy violation incidents (Column 7), which are involved with mishandling personal information without proper encryption. It appears to be challenging to completely protect personal information that flows throughout enterprise architectures tangled with a myriad of legacy systems. Table 5, Column 9 demonstrates that the legacy systems are vulnerable even to non-cyber incidents, in which personal information is mishandled or breached via printed materials.

Likewise, Table 5 shows that the coefficient of cloud computing is negative and significant for all columns but Columns (1), (2) and (9), providing support to Hypothesis 3. The impact of cloud computing is strongest for policy violation; when a federal agency spends a 1%-point more in cloud computing over total IT investments, it experiences 1.5% fewer incidents in violation of personal information policies by employees and contractors. This finding is in accordance with Hypothesis 2 that it is more straightforward and flexible to impose security protocols to the cloud computing than to on-premise ones. While the coefficient of DME is not significant for equipment incidents (Column 6), that of Cloud is significant. It seems that cloud computing vendors may have stronger physical access controls that restrict unauthorized access to IT equipment as well as effective process controls for the custody of digital equipment.

[5] $1 - e^{-0.052} = 0.0507$

Table 5. Estimation with FISMA Incident Data

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**								
Data Source	FISMA Reports to Congress in FY 2012-2015								
Incident Types (DV in Log)	Total Incidents	Denial of Service	Improper Use				Policy Violation	Malicious Code	Non-Cyber Incidents
			Improper Use Total	Unauthorized Access					
				Unauthorized Access Total	Social Engineering	Unauthorized Equipment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DME	-0.052***	0.005	-0.061***	-0.032**	-0.094***	0.017	-0.072***	-0.064***	-0.072***
	(0.007)	(0.008)	(0.007)	(0.011)	(0.003)	(0.013)	(0.002)	(0.011)	(0.005)
Cloud	0.002	-0.001	-0.012***	-0.008**	-0.011***	-0.011**	-0.015***	-0.012**	0.006*
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.001)	(0.004)	(0.002)
Disperse	-11.760***	8.664***	-15.866***	-15.315***	-32.402***	-13.644***	-22.865***	-6.774*	-21.685***
	(1.629)	(1.099)	(0.784)	(1.146)	(1.709)	(0.795)	(1.967)	(2.661)	(1.611)
Diversity	-0.816***	1.360***	-0.913***	-0.688***	0.223	-1.098***	-1.029***	-0.415	0.125
	(0.087)	(0.246)	(0.076)	(0.142)	(0.212)	(0.164)	(0.139)	(0.385)	(0.321)
IT Invest	-26.808***	-6.076	-24.580***	-36.167**	-35.310+	-30.511*	-24.379**	-82.363***	-2.174
	(0.595)	(9.004)	(3.432)	(10.719)	(18.626)	(12.208)	(7.259)	(12.091)	(7.643)
Security Invest	0.201***	-0.246***	0.183**	0.410***	0.347***	0.265***	-0.041	0.248	-0.008
	(0.030)	(0.030)	(0.051)	(0.034)	(0.051)	(0.060)	(0.095)	(0.170)	(0.080)
Budget	-0.681***	0.850***	-0.510***	-0.497+	-1.499***	0.111	-0.413***	-1.769***	-0.407
	(0.167)	(0.129)	(0.104)	(0.274)	(0.179)	(0.369)	(0.034)	(0.281)	(0.260)
Bureau	-0.123	-0.823	0.417	-1.127	-1.670	-0.074	1.238*	-0.945	0.414**
	(0.934)	(0.793)	(0.661)	(0.833)	(0.980)	(0.831)	(0.587)	(1.231)	(0.120)
FTE	3.420***	-1.834*	4.882***	0.630	4.657***	1.356+	5.741***	0.945	5.091**
	(0.269)	(0.871)	(0.962)	(0.761)	(0.830)	(0.668)	(0.678)	(3.003)	(1.375)
Age	-0.040	-0.027	-0.229*	0.270**	0.167	0.025	-0.297**	0.466	-0.247
	(0.050)	(0.118)	(0.100)	(0.076)	(0.136)	(0.091)	(0.097)	(0.344)	(0.211)
Defense	-0.015	0.051***	0.015	-0.006	-0.002	-0.005	0.032*	-0.054***	-0.024+
	(0.015)	(0.012)	(0.022)	(0.021)	(0.017)	(0.027)	(0.012)	(0.001)	(0.013)
Welfare	-0.014	0.050***	0.015	-0.007	-0.017	-0.002	0.034*	-0.051***	-0.022
	(0.015)	(0.011)	(0.022)	(0.022)	(0.017)	(0.028)	(0.012)	(0.009)	(0.014)
Law Enforce	-0.005	0.076*	0.014	-0.001	0.040	-0.058*	0.027	-0.030	0.030*
	(0.008)	(0.035)	(0.013)	(0.019)	(0.035)	(0.022)	(0.016)	(0.027)	(0.012)
Management	-0.387***	-0.085	-0.417***	-0.554**	-0.528+	-0.569**	-0.370***	-0.719***	0.019
	(0.007)	(0.120)	(0.038)	(0.147)	(0.256)	(0.147)	(0.081)	(0.171)	(0.075)
Regulation	0.019**	-0.016	0.018**	0.020*	-0.021+	0.034**	0.015***	-0.002	-0.021***
	(0.005)	(0.018)	(0.006)	(0.009)	(0.012)	(0.012)	(0.001)	(0.006)	(0.004)
F	320.92***	113.61***	754.99***	149.23***	2502.74***	151.89***	79.76***	306.07***	130.40***
Within R^2	0.323	0.530	0.412	0.299	0.456	0.301	0.446	0.340	0.276

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; $N = 96$; # of Agency = 24; Agency and year fixed-effects included; Standard errors are in parentheses.

Table 6. Moderating Effect of Inspector General Compliance Score with FISMA Incident Data

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**								
Data Source	FISMA Reports to Congress in FY 2012-2015								
Incident Types (DV in Log)	Total Incidents	Denial of Service	Improper Use				Policy Violation	Malicious Code	Non-Cyber Incidents
			Improper Use Total	Unauthorized Access					
				Unauthorized Access Total	Social Engineering	Unauthorized Equipment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
DME	-0.060***	-0.034	-0.060***	-0.090***	-0.101*	-0.100**	-0.076***	-0.085***	-0.078***
	(0.013)	(0.020)	(0.005)	(0.015)	(0.042)	(0.029)	(0.018)	(0.012)	(0.014)
DME × IG Score	0.000***	0.001**	0.000***	0.001***	0.000**	0.001***	0.000***	0.000***	0.000**
	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*	(0.000)*
Cloud	0.001	-0.005	-0.013***	-0.010**	-0.014***	-0.014**	-0.016***	-0.012**	0.007**
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
Disperse	-12.498***	10.063***	-16.798***	-17.977***	-37.013***	-17.816***	-21.232***	-10.450***	-20.516***
	(1.490)	(1.610)	(0.857)	(1.954)	(1.911)	(1.551)	(2.432)	(1.747)	(2.460)
Diversity	-0.868***	1.138***	-0.981***	-0.706**	0.150	-1.060***	-1.149***	-0.360	0.200
	(0.119)	(0.269)	(0.106)	(0.225)	(0.232)	(0.211)	(0.124)	(0.503)	(0.297)
IG Score	-0.008+	-0.005+	-0.007	-0.032***	-0.034**	-0.055***	0.006*	-0.028*	0.009
	(0.004)	(0.003)	(0.005)	(0.002)	(0.008)	(0.004)	(0.003)	(0.013)	(0.007)
IT Invest	-25.461***	-1.926	-23.121***	-30.828***	-27.748+	-22.311**	-24.101**	-77.494***	-4.884
	(1.143)	(9.020)	(4.484)	(7.278)	(14.002)	(6.574)	(6.250)	(6.188)	(10.040)
Security Invest	0.207***	-0.274***	0.206**	0.440***	0.425***	0.291***	-0.033	0.261	-0.064
	(0.031)	(0.034)	(0.057)	(0.030)	(0.047)	(0.031)	(0.094)	(0.165)	(0.082)
Budget	-0.668***	1.042***	-0.510***	-0.585*	-1.634***	-0.030	-0.330***	-1.893***	-0.381
	(0.151)	(0.095)	(0.068)	(0.276)	(0.175)	(0.352)	(0.070)	(0.328)	(0.250)
Bureau	0.183	0.477	0.656	-0.984	-1.848+	0.163	1.710**	-1.288	0.374+
	(0.972)	(0.938)	(0.610)	(1.031)	(0.987)	(1.125)	(0.474)	(1.575)	(0.182)
FTE	3.335***	-2.823**	4.752***	0.936	5.156***	2.004+	5.024***	1.851	5.295**
	(0.152)	(0.715)	(0.772)	(0.939)	(0.833)	(0.988)	(0.426)	(2.849)	(1.678)
Age	-0.024	0.023	-0.202*	0.289***	0.183	0.046	-0.251**	0.430	-0.281
	(0.043)	(0.093)	(0.093)	(0.062)	(0.118)	(0.069)	(0.085)	(0.338)	(0.225)
F	1393.98***	384.26***	242.86***	131.05***	2998.95***	224.81***	80.38***	131.63***	82.48***
Within R^2	0.330	0.561	0.424	0.351	0.487	0.416	0.455	0.373	0.299

+$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; $N = 91$; # of Agency = 23; Agency and year fixed-effects included; Standard errors are in parentheses.
Other control variables are omitted for brevity.

The coefficients of Disperse and Diversity are estimated to be negative and significant in Table 5, Column 1, supporting Hypotheses 4 and 5, respectively. Comparing Column 2 to other columns gives an interesting result. Column 2 supports an intuition that DoS attacks occur more frequently at federal agencies that are geographically distributed or perform diverse functions. However, the geographic footprints and the function diversity have an opposite effect on other types of security incidents. For example, security incidents via social engineering such as phishing occur more frequently at agencies whose employees are more geographically concentrated.

Table 6 presents the interaction effects of IT-GRC effectiveness measured by inspectors general and new IT development/modernization to test Hypothesis 2. The coefficient of DME × IG Score is positive and significant for denial of service (DoS, Column 2), unauthorized access (Column 4), and unauthorized equipment use (Column 6). With a higher IG Score, the coefficient of DME on security breaches becomes less negative. Since the rest of IT spending goes to maintenance of existing systems, this indicates that in an agency with stronger IT-GRC mechanisms, an increase in maintenance spending leads to a smaller increase in security breach incidents. This provides an encouraging finding that the undesired effects of legacy systems on security incidents is mitigated when a federal agency has stronger IT-GRC mechanisms. Figure 1 illustrates this effect on security failures involved with unauthorized equipment use (from Table 6, Column 6). When a federal agency has a higher IG Score, not only does it experience less frequent incidents, but also the effect of its legacy systems (lower DME) is attenuated. This finding offers a counterintuitive implication that modernizing legacy systems and improving IT-GRC are in fact substitutes with respect to cybersecurity risks.

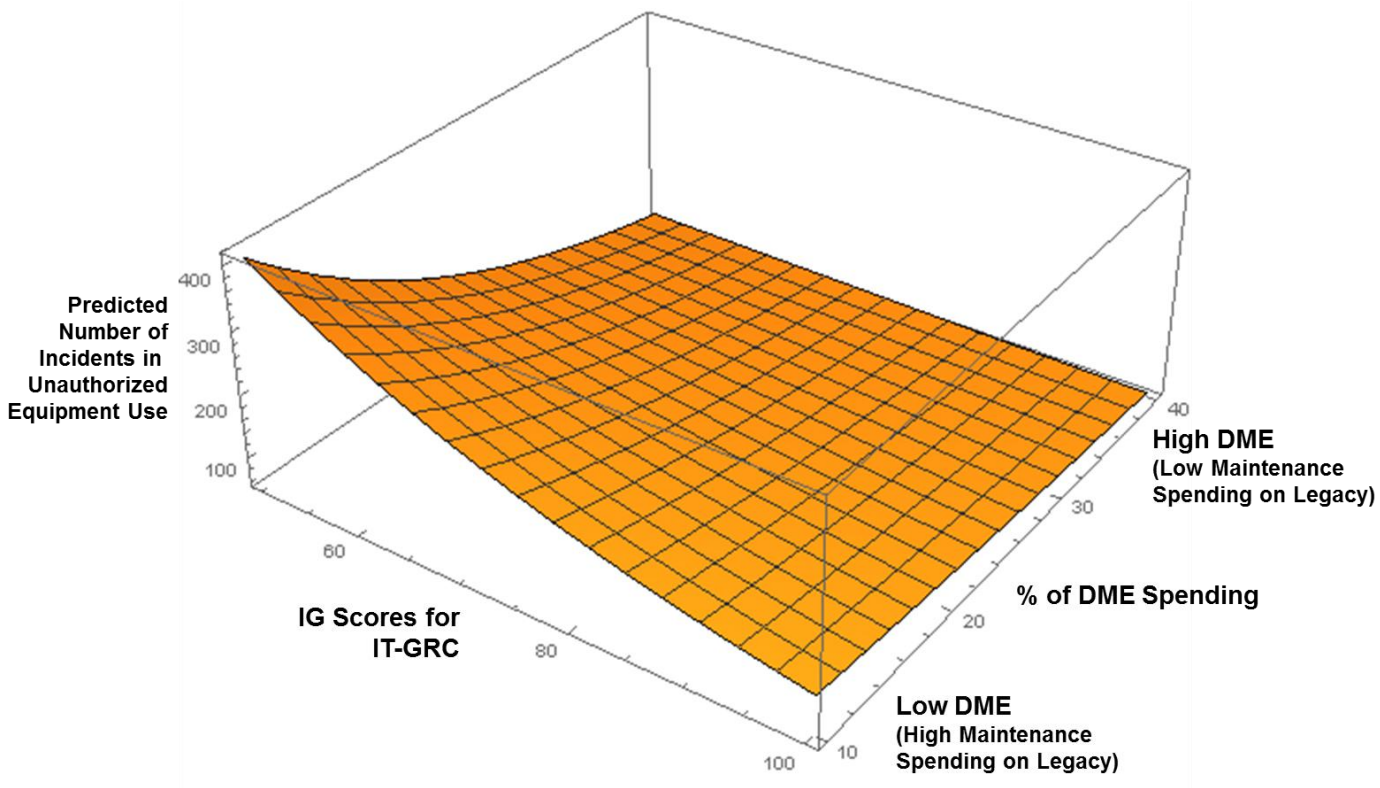

Figure 1. Interaction Effects of IG Score and DME Spending

It is intriguing to see that while the coefficient of DME on DoS and unauthorized equipment incidents (Table 5, Columns 2 and 6, respectively) is insignificant, the interaction effect of DME and IG Score in Table 6, Columns 2 and 6 is positive and significant. This finding implies that effective IT governance and control mechanisms could prevent DoS attacks from turning into security incidents. Agencies whose IT systems are well governed and well controlled may have instituted mechanisms to neutralize the effects of the DoS attacks. They could institute effective network layer controls and analytical intelligence to detect DoS attack traffics at the network layer and block them there before the attacks have a chance to reach the IT applications. Incidents of unauthorized equipment use are likely to occur under weak security controls, such as under weak physical access controls that may allow anyone to enter a secure facility or under weak equipment safeguards that could lead to loss of devices and equipment in transit (by travelling employees or third-party couriers) between sites. The result in Table 6 illustrates that effective IT-GRC prevents IT devices from being lost or stolen.

Table 7. Estimation with Privacy Rights Clearinghouse Breach Data in FY 2005-2016

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**					
Data Source	Privacy Rights Clearinghouse (FY 2005-2016)					
Breach Types (DV in Log)	Total Breach		Intentional Breach		Unintentional Breach	
	(1)	(2)	(3)	(4)	(5)	(6)
DME	-0.003^{+} (0.002)		0.000 (0.002)		-0.003^{+} (0.002)	
DME × I(2005-2007)		0.000 (0.002)		0.000 (0.002)		0.000 (0.002)
DME × I(2008-2010)		0.000 (0.003)		0.001 (0.003)		-0.001 (0.003)
DME × I(2011-2013)		-0.011** (0.003)		0.001 (0.003)		-0.010** (0.003)
DME × I(2014-2016)		-0.013***(0.002)		-0.003 (0.003)		-0.010** (0.003)
Disperse	0.457^{+} (0.256)	0.176 (0.318)	-0.169 (0.267)	-0.197 (0.294)	0.727* (0.302)	0.483^{+} (0.270)
Diversity	-0.014 (0.197)	0.008 (0.194)	0.031 (0.054)	0.065 (0.059)	-0.021 (0.212)	-0.020 (0.201)
IG Score	-2.333* (0.915)	-2.111* (0.917)	-1.310^{+} (0.755)	-1.382^{+} (0.744)	-1.259 (0.799)	-1.009 (0.736)
IT Invest	-0.005** (0.002)	-0.003^{+} (0.002)	0.000 (0.001)	0.000 (0.001)	-0.006* (0.002)	-0.004* (0.002)
Security Invest	0.038 (0.064)	0.034 (0.071)	0.044^{+} (0.022)	0.052* (0.025)	-0.025 (0.063)	-0.034 (0.067)
Budget	0.121 (0.253)	0.006 (0.223)	0.052 (0.256)	0.006 (0.270)	0.099 (0.254)	0.013 (0.316)
Bureau	-0.416***(0.090)	-0.320** (0.105)	-0.126 (0.081)	-0.121 (0.075)	-0.412***(0.091)	-0.326** (0.085)
FTE	0.044* (0.018)	0.037* (0.017)	0.007 (0.008)	0.006 (0.007)	0.056***(0.014)	0.051** (0.014)
Age	-0.022* (0.010)	-0.022^{+} (0.011)	-0.005 (0.003)	-0.005 (0.004)	-0.022* (0.010)	-0.022^{+} (0.011)
Defense	-0.009 (0.010)	-0.008 (0.010)	0.001 (0.003)	0.001 (0.003)	-0.014 (0.009)	-0.013 (0.010)
Welfare	-0.025^{+} (0.014)	-0.026* (0.012)	-0.009 (0.005)	-0.009^{+} (0.005)	-0.019 (0.012)	-0.019^{+} (0.010)
Law Enforce	-0.013 (0.008)	-0.012 (0.010)	-0.011** (0.003)	-0.010** (0.003)	-0.004 (0.008)	-0.004 (0.010)
Management	0.009* (0.004)	0.014** (0.005)	0.010***(0.003)	0.011** (0.003)	0.000 (0.004)	0.004 (0.005)
Regulation	-0.003^{+} (0.002)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)	-0.003^{+} (0.002)	0.000 (0.002)
N	286	286	286	286	286	286
# of Groups	24	24	24	24	24	24
F	2633.52***	6575.95***	3145.86***	91.71***	139222.31***	247.85***
Within R^2	0.256	0.281	0.168	0.172	0.267	0.291

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Agency and year fixed-effects included; Standard errors are in parentheses.;

Table 8. Estimation with Privacy Rights Clearinghouse Breach Data in FY 2012-2016

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**					
Data Source	Privacy Rights Clearinghouse (FY 2012-2016)					
Breach Types (DV in Log)	Total Breach		Intentional Breach		Unintentional Breach	
	(1)	(2)	(3)	(4)	(5)	(6)
DME	-0.001 (0.005)	-0.046* (0.017)	0.003 (0.009)	-0.041*** (0.006)	-0.002 (0.005)	-0.015 (0.010)
DME × IG Score		0.0004* (0.000)		0.001** (0.000)		0.000 (0.000)
Cloud	0.000 (0.001)	0.001 (0.001)	-0.001 (0.001)	0.000 (0.000)	0.002* (0.001)	0.003** (0.001)
Disperse	0.632 (0.646)	0.352* (0.155)	0.344 (0.768)	0.158 (0.106)	-0.579+ (0.321)	0.267*** (0.058)
Diversity	0.268 (0.185)	-0.619 (1.556)	0.073 (0.146)	-1.048 (0.749)	0.233*** (0.053)	-1.387 (0.976)
IG Score	-0.195 (0.351)	-0.015** (0.004)	-0.817 (2.192)	-0.016*** (0.002)	0.818 (1.744)	0.000 (0.003)
IT Invest	0.044 (0.032)	-2.055 (3.577)	0.014 (0.016)	-1.551 (3.737)	0.043+ (0.022)	-0.945 (1.009)
Security Invest	0.197*** (0.030)	0.125*** (0.012)	0.219* (0.090)	0.040* (0.017)	-0.044 (0.056)	0.099*** (0.014)
Budget	-0.214 (0.271)	-0.028 (0.034)	0.261* (0.125)	0.055 (0.069)	-0.492+ (0.273)	-0.198** (0.054)
Bureau	-1.125*** (0.187)	-1.531** (0.433)	-0.598 (0.496)	-0.562*** (0.055)	-0.379 (0.553)	-1.339** (0.378)
FTE	0.078* (0.030)	-0.917** (0.251)	0.019 (0.052)	0.104 (0.458)	0.052 (0.057)	-0.762 (0.696)
Age	0.009 (0.007)	0.131*** (0.008)	0.001 (0.004)	0.001 (0.049)	0.004 (0.004)	0.145+ (0.075)
Defense	0.022** (0.007)	0.009 (0.006)	0.012** (0.004)	-0.001 (0.004)	0.009** (0.003)	0.004* (0.001)
Welfare	-0.032** (0.009)	0.019** (0.005)	-0.041*** (0.010)	0.007+ (0.004)	0.004 (0.006)	0.000 (0.001)
Law Enforce	-0.006 (0.006)	-0.036** (0.012)	-0.033** (0.008)	-0.034* (0.013)	0.037** (0.012)	-0.002 (0.005)
Management	0.009** (0.003)	-0.022 (0.066)	0.009* (0.003)	-0.021 (0.045)	0.002 (0.002)	0.012 (0.029)
Regulation	-0.001 (0.005)	0.010+ (0.005)	0.003 (0.009)	0.007 (0.005)	-0.002 (0.005)	0.002 (0.001)
N	120	91	120	91	120	91
# of Groups	24	23	24	23	24	23
F	36.64***	10.27***	17.99***	7.62***	4.46***	3.01**
Within R^2	0.405	0.211	0.335	0.156	0.309	0.299

+$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; Agency and year fixed-effects included; Standard errors are in parentheses.;

Our estimations with the alternative dataset from Privacy Rights Clearinghouse also provide intriguing results (Table 7 and 8). Table 7 presents estimation results for the data in 2005-2016 with 286 observations. Columns 1 and 5 show that the coefficient of DME is negative and marginally significant for total breaches and unintended disclosures of personal information. Unintended disclosures include cases in which sensitive information was accidently posted on the Internet, inadvertently delivered to a wrong party, or accessible mistakenly by unauthorized personnel. For instance, it was reported in 2013 that users in System for Award Management (SAM) at the General Administration Services, including external contractors and vendors, were able to view financial information and trade secrets of other SAM users. Interestingly, when we added interaction terms of DME and time dummies (e.g. I(2011-2013) is one for FY 2011, 2012, and 2013), the share of DME spending is significantly associated with fewer breaches in 2011-2016 (Table 7, Columns 2 and 6). Column 2 illustrates that a 1%-point increase in DME leads to a 1% decrease in total breaches of personal information after 2011. This suggests that the complex legacy systems have become more vulnerable to security threats for personal information that are increasingly more sophisticated in the recent years.

Table 8 includes Cloud variable[6] and presents the interaction effect of IT-GRC. While the coefficient of DME is not significant for intentional breaches (Table 7, Columns 2-3 and Table 8, Column 3), it is significantly moderated by IG Score (Table 8, Column 4). This finding supplements the results from Table 6 and shows that in a federal agency with weaker IT-GRC, legacy systems are more susceptible to a breach of private information with a malicious intent.

[6] Since Federal IT Dashboard reports cloud computing spending only in 2012-2016, we present this result in a separate table.

Table 9. Fixed-Effects Poisson Regression

Method	Conditional Fixed-Effects Poisson Regression					
Data Source	FISMA Report					
Incident Types (DV in Log)	Total Incidents	Improper Use	Unauthorized Access	Social Engineering	Policy Violation	Non-Cyber Incidents
	(1)	(2)	(3)	(4)	(5)	(6)
DME	-0.047***	-0.036***	-0.037***	-0.146***	-0.051***	-0.049***
	(0.002)	(0.002)	(0.003)	(0.006)	(0.004)	(0.004)
Cloud	0.000	-0.007***	0.001	-0.003	-0.016***	0.000
	(0.000)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Disperse	-15.672***	-13.726***	-13.506***	-12.795***	-18.312***	-56.710***
	(0.544)	(0.707)	(1.098)	(2.052)	(1.089)	(1.759)
Diversity	0.138***	0.861***	0.872***	3.120***	1.010***	-0.111+
	(0.036)	(0.056)	(0.078)	(0.231)	(0.085)	(0.068)
IT Invest	-18.855***	-21.305***	-24.237***	-56.619***	-22.944***	-43.909***
	(0.947)	(1.206)	(1.748)	(6.252)	(2.732)	(9.303)
Security Invest	-0.089***	-0.125***	-0.033	-0.618***	-0.403***	-0.020
	(0.013)	(0.018)	(0.026)	(0.055)	(0.029)	(0.032)
Budget	-0.007	-0.146*	-0.417***	-1.122***	0.324***	-0.029
	(0.034)	(0.063)	(0.102)	(0.309)	(0.088)	(0.082)
Bureau	1.016***	0.609**	-0.190	3.630***	2.439***	6.260***
	(0.153)	(0.191)	(0.232)	(1.000)	(0.385)	(0.483)
FTE	0.953***	0.265+	0.407+	7.722***	0.419+	-0.034
	(0.100)	(0.155)	(0.229)	(0.738)	(0.227)	(0.263)
Age	-0.011**	0.025***	0.048***	0.225***	-0.014+	-0.002
	(0.003)	(0.005)	(0.008)	(0.023)	(0.008)	(0.010)
Defense	0.007**	0.030***	0.011**	0.040***	0.073***	0.000
	(0.002)	(0.003)	(0.004)	(0.011)	(0.006)	(0.012)
Welfare	0.008**	0.031***	0.010*	0.026*	0.072***	0.004
	(0.002)	(0.003)	(0.004)	(0.011)	(0.006)	(0.012)
Law Enforce	-0.007*	-0.004	-0.018*	-0.036*	0.015**	0.063***
	(0.003)	(0.004)	(0.007)	(0.015)	(0.005)	(0.009)
Management	-0.236***	-0.258***	-0.269***	-0.914***	-0.270***	0.297***
	(0.009)	(0.010)	(0.015)	(0.064)	(0.017)	(0.033)
Regulation	-0.011***	-0.018***	-0.016***	-0.103***	-0.030***	-0.025***
	(0.001)	(0.002)	(0.003)	(0.007)	(0.003)	(0.004)
Wald χ^2	3182.55	2190.57	1299.14	2221.94	1438.22	3509.89
Log Likelihood	-2531.453	-1823.072	-1483.550	-799.554	-1369.831	-1848.146

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; $N = 96$; # of Agency = 24; Agency and year fixed-effects included; Standard errors are in parentheses.; For other incident types, maximum likelihood estimation failed to converge.

We conducted a few robustness checks as follows. First, since the dependent variables are count variables (the number of security incidents), we estimated the model with conditional fixed-effects Poisson regressions. Table 9 shows consistent results with our baseline estimation in Table 5. The coefficient of DME is consistently negative in all columns, and so is that of Disperse (geographical dispersion). A random-effects Poisson regression and a negative binominal regression provide similar results. Second, as stated above, our sample agencies are the CFO Act agencies that include both cabinet departments and independent agencies. In Table A1 (Appendix), we re-ran the model without the non-

cabinet agencies (e.g. the OPM, the National Science Foundation), which are smaller than the cabinet agencies and perform more homogenous functions. The results in Table A1 are relatively less significant but consistent with Table 5. The coefficient of Diversity is still negative and significant in Columns 1, 3, 4, and 5, even though we excluded the non-cabinet agencies, which perform much less heterogeneous functions that the cabinet ones.

Third, we re-estimate the model with alternative DME variables. As mentioned in Table 3, DME was measured by the share of new IT development to total IT spending for a five-year window. Table A2 provides estimation results with different windows for DME (three-year and seven-year). We did not obtain substantially different results. In Table A3, we use two alternative explanatory variables. In Columns 1-4, instead of the proportion of DME over total IT expenditures, we used the total amount of DME spending for the previous five years (as transformed in log). These estimations generate similar results. In Columns 5-8, in calculating the percentage of DME to total IT spending, we excluded security-related IT investments, including expenditures related to identification and authentication, access control, and incident response. On average, these security-related IT investments only account for 2.33% of the total federal IT spending. Excluding such spending in DME does not change our main findings substantially. The signs and the magnitudes in the coefficient of DME in Table A3, Columns 5-8 are very similar to Table 6.

5. Discussions and Conclusion

In an increasingly digital economy, both the public- and the private-sector organizations face more escalating cybersecurity risks, which could threaten the very existence of organizations. In this study, we focus on potential drivers and mitigation mechanisms of IT security incidents in the U.S. federal government. To do so, we considered both technological and organizational factors.

Technologically, we found that the large stock of legacy IT systems in the U.S. federal government is a major driver of the IT security incidents. To the best of our knowledge, this is the first

empirical test of a widely-held belief among IT security professionals that legacy IT systems might have an advantage of "security by antiquity." Drawing upon economic theories of criminal behaviors and the literature on software economics, we hypothesized that federal agencies with more legacy systems are likely to experience more security breaches. Our empirical findings support this hypothesis. This finding is also corroborated by supplemental analyses with the Privacy Rights Clearinghouse data. Thus, it is important to recognize that legacy IT systems create grave vulnerabilities in federal IT infrastructures, which cybercriminals can easily exploit. This finding suggests that modernization of legacy systems is one of the key mitigation mechanisms for the IT vulnerabilities.

Another potential mitigation is the use of the cloud. We find that federal agencies that outsource more legacy systems to the cloud experience fewer security breaches than others. The third potential mitigation mechanism we consider is the institution of effective IT governance, risk, and control (IT-GRC) measures around the legacy systems. Our analyses show that federal agencies that institute more effective IT-GRC mechanisms experience fewer security incidents from the legacy systems.

Organizationally, the agencies that are geographically or functionally more dispersed experience security breaches less frequently than the ones that are more centralized. These findings are counterintuitive. Our theoretical discussions and empirical evidence indicate that modernizing IT systems could make it easier to consolidate, integrate, and standardize the IT systems, in turn limiting potential vulnerability points available to attackers. However, centralization of federal agencies can lead to the opposite result. A centralized agency may also have highly valuable "treasures" available for cybercriminals. The treasure value is higher because the sensitive data of the agency is likely to at a central location, rather than being scattered across many different locations. Instead of having to design many different attacks across many dispersed IT system, attackers can design a common attack for the central location and scale it up across many units and functions within the central location.

This research makes an important contribution to the IS literature as follows. To the best of our knowledge, our study is the first to examine how IT investment patterns (maintain legacy systems versus modernize and develop new IT systems) and organizational forms (geographic concentration and

functional diversity) affect security vulnerabilities. The U.S. federal government is an ideal setting that allows us to investigate the technical and organizational antecedents of security failures across several large-scale organizations. It would be challenging to collect security incident data from many large organizations in the private sector, where most security incidents are not reported publicly. This study makes an important theoretical contribution by extending the economic theory of criminal behaviors in explaining the role of IT investment patterns and organizational forms in information security. We also contribute to the literature by investigating the crucial roles of IT-GRC mechanisms such as security management capabilities in alleviating security vulnerabilities.

Our study provides ample managerial implications for IT managers in the public sector and security professionals in the private sector, to whom governments are among their largest customers. IT managers are advised to abandon their conventional wisdom of security-by-antiquity and take security threats posed by legacy enterprise architectures seriously. Our empirical analyses raise a wake-up call that old and technologically outdated systems cause a wide range of security incidents including intentional breaches with malicious intent (Table 5) or unintentional disclosures of sensitive information (Table 7, Column 5). Therefore, this study informs that it is imperative for an organization to modernize decades-old legacy systems and to consistently monitor and patch the vulnerabilities from the antiquated legacy infrastructures. Security managers at organizations with antiquated enterprise architectures also need to improve IT-GRC mechanisms, an effort that can alleviate security vulnerabilities from the legacy infrastructures (Table 6). Our counterintuitive findings in Tables 6 and 8 inform federal officials that under a circumstance where it is challenging to secure budgets for IT modernization, enhancing agencies' IT-GRC mechanisms can be an alternative resort to mitigate the cybersecurity risks.

This paper defies another conventional wisdom that organizations that are geographically distributed are more susceptible to security threats. It turns out that federal agencies whose employees are more concentrated and ones that serve more homogenous functions suffer from more frequent breaches. This finding provides an important implication to security managers in both the public and the private

sector that cybercriminals are more likely to target organizations whose data assets are more concentrated. Such organizations are advised to take extra precaution in protecting their information resources.

This study carries a few limitations. First, the FISMA reports do not report the number of security incidents at the sub-agency bureau level, such as the Transportation Security Administration under the DHS or the Federal Aviation Administration under the DOT. Security breach information at a more granular level would have afforded us a more in-depth analysis on the relationship between organizational characteristics and security vulnerabilities. Second, the FISMA dataset covers only four years in 2012-2015. With a longer timeframe, we could have examined how changes in political environments affect security vulnerabilities in federal agencies. Third, we were not able to obtain more detailed information in security breaches, other than incident categories (Table 2) and the number of incidents in each category. We could not find out, for example, the magnitude of security breaches, information on potential attackers, or which system was targeted and compromised. While Privacy Rights Clearinghouse does provide the number of records that were breached, it does so for only 39% of the incidents that occurred in federal agencies, and many of such figures are estimates.

The public sector provides a range of interesting and promising opportunities for future IS and security research. For example, one can examine how political environments affect security vulnerabilities in governments. Given the political ramifications from a security failure, which we observed from the OPM incident, IS researchers can study how political environments in the legislative branch influence security investments or capabilities in the executive branch. It would be also interesting to study the relationship between IT outsourcing and security vulnerabilities, since outsiders such as vendors and contractual personnel who have access to internal systems can pose security risks. This risk is clearly exemplified by the case of Edward Snowden, who was an external contractor for the National Security Agency when he exposed its surveillance program. Researchers can also study how high-profile incidents of privacy breaches such as ones in Target and the OPM affect security investments in for-profit firms and governments.

Reference

Associated Press (2015) “IRS: Computer Breach Bigger than First Thought, with 700K Victims,” available at: http://www.pbs.org/newshour/rundown/irs-computer-breach-bigger-than-first-thought-with-700k-victims/, accessed on Jan 5, 2017.

Banker, B.D. and Slaughter, S.A. (2000) “The Moderating Effects of Structure on Volatility and Complexity in Software Enhancement,” *Information Systems Research* (11:3) pp. 219-240.

Barki, H., Rivard, S., and Talbot, J. (2001). "An Integrative Contingency Model of Software Project Risk Management," *Journal of Management Information Systems* (17:4), pp. 37-69.

Baskerville, R., Rowe, F., and Wolff, F-C. (forthcoming) “Integration of Information Systems and Cybersecurity Countermeasures: An Exposure to Risk Perspective.” *The Data Base for Advances in Information Systems*.

Brumley, D., Newsome, J., Song, D., Wang, H., and Jha, S. (2008) “Theory and Techniques for Automatic Generation of Vulnerability-Based Signatures,” *IEEE Transactions on Dependable and Secure Computing* (5:4) pp. 224-241.

Cavusoglu, H., Raghunathan, S., and Cavusoglu, H. (2009) “Configuration of and Interaction between Information Security Technologies: The Case of Firewalls and Intrusion Detection Systems,” *Information Systems Research* (20:2) pp. 198-217.

Cavusoglu, H., Raghunathan, S., and Yue, W.T. (2008) “Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment,” *Journal of Management Information Systems* (25:2) pp. 281-304.

CIO (2016) *Why It’s Time To Learn COBOL*, available at http://www.cio.com/article/3050836/developer/why-its-time-to-learn-cobol.html, accessed on Mar 7, 2017.

COSO (1992) *Internal Control - Integrated Framework*. New York, NY: AICPA.

D'Arcy, J., Hovav, A., and Galletta, D. (2009). "User Awareness of Security Countermeasures and Its Impact on Information Systems Misuse: A Deterrence Approach," *Information Systems Research* (20:1) pp. 79-98.

Driscoll, J.C. and Kraay, A.C. (2006) “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data,” *Review of Economics and Statistics* (80:4) pp. 549-560.

Ehrlich, I. (1996) “Crime, Punishment, and the Market for Offenses,” *Journal of Economic Perspectives* (10:1) pp. 43-67.

Eisenhardt, K.M. (1985). "Control: Organizational and Economic Approaches," *Management Science* (31:2), pp. 134-149.

FCW (2013) “The Bright Side of Obsolescence,” available at: https://fcw.com/Blogs/FCW-Insider/2013/01/cobol-benefit.aspx, accessed on Mar 7, 2017.

FCW (2015) “The Taxman’s Tech Troubles,” available at: https://fcw.com/Articles/2016/04/08/taxman-tech-troubles.aspx, accessed on Jan 5, 2017.

Fisher, R. (1984). *Information Systems Security*. Englewood Cliffs: Prentice-Hall.

Free Beacon (2013) “The Cyber-Dam Breaks,” available at: http://freebeacon.com/national-security/the-cyber-dam-breaks/, accessed on May 21, 2016.

Goel, R.K. and Rich, D.P. (1989) “On the Economic Incentive for Taking Bribes,” *Public Choice* (61:3) pp. 269-275.

Goold, M., and Quinn, J.J. (1990). "The Paradox of Strategic Controls," *Strategic Management Journal* (11:1), pp. 43-57.

Harter, D.E., Kemerer, C.F., and Slaughter, S.A. (2012) “Does Software Process Improvement Reduce the Severity of Defects? A Longitudinal Field Study,” *IEEE Transactions on Software Engineering* (38:4) pp. 810-827.

Harter, D.E., Krishnan, M.S., and Slaughter, S.A. (2000) “Effects of Process Maturity on Quality, Cycle Time, and Effort in Software Product Development,” *Management Science* (46:4) pp. 451-466.

Henderson, J.C., and Venkatraman, N. (1993). "Strategic Alignment: Leveraging Information Technology for Transforming Organizations," *IBM Systems Journal* (32:1) pp. 472-484.

Hughes. J. and Cybenko, G. (2014). Three Tenets for Secure Cyber-Physical System Design and Assessment. in I.V. Ternovskiy & P. Chin (Eds.), *Cyber Sensing 2014: SPIE Defense+ Security* (Vol. 9097, pp. 90970A-90915): International Society for Optics and Photonics.

IIA. (2008). *GAIT for IT General Control Deficiency Assessment: An Approach for Evaluating ITGC Deficiencies in Sarbanes-Oxley Section 404 Assessments of Internal Controls over Financial Reporting*. Altamonte Spring, FL.: The Institute of Internal Auditors.

IIA. (2009). *Global Technology Audit Guide (GTAG) 8: Auditing Application Controls*. Altamonte Spring, FL: The Institute of Internal Auditors.

IIA. (2012). *Global Technology Audit Guide (GTAG) 1: Information Technology Controls*, (2nd ed.). Altamonte Spring, FL: The Institute of Internal Auditors.

ITGI. (2007). *COBIT 4.1*. Rolling Meadows, IL: IT Governance Institute.

Keil, M., Cule, P.E., Lyytinen, K., and Schmidt, R.C. (1998). "A Framework for Identifying Software Project Risks," *Communications of the ACM* (41:11) pp. 76-83.

Kirsch, L.J. (1996). "The Management of Complex Tasks in Organizations: Controlling the Systems Development Process," *Organization Science* (7:1) pp. 1-21.

Kirsch, L.J. (1997). "Portfolios of Control Modes and IS Project Management," *Information Systems Research* (8:3) pp. 215-239.

Kotulic, A. and Clark, J.G. (2004). “Why There Aren’t More Information Security Research Studies.” *Information and Management* (41:5) pp. 597-607.

Krishnan, M.S. and Kellner, M.I. (1999) “Measuring Process Consistency: Implications for Reducing Software Defects,” *IEEE Transactions on Software Engineering* (25:6) pp. 800-815.

Krishnan, M.S., Kriebel, C.H., Kekre, S., and Mukhopadhyay, T. (2000) “An Empirical Analysis of Productivity and Quality in Software Products,” *Management Science* (46:6) pp. 745-759.

Kwon, J. and Johnson, M.E. (2014) “Proactive versus Reactive Security Investments in the Healthcare Sector,” *MIS Quarterly* (38:2) pp. 451-471.

Leifer, R. (1989) “Understanding Organizational Transformation using a Dissipative Structure Model,” *Human Relations* (42:10) pp. 899-916.

Lyytinen, K., Mathiassen, L., and Ropponen, J. (1998). "Attention Shaping and Software Risk - a Categorical Analysis of Four Classical Risk Management Approaches," *Information Systems Research* (9:3) pp. 233-255.

Machin, S. and Meghir, C. (2004) “Crime and Economic Incentives,” *Journal of Human Resources* (39:4) pp. 958-979.

Macintosh, N.B. (1994) *Management Accounting and Control Systems: An Organizational and Behavioral Approach*. New York, NY: John Wiley.

Mitra, S. and Ransbotham, S. (2015) “Information Disclosure and the Diffusion of Information Security Attacks,” *Information Systems Research* (26:3) pp. 565-584.

Nappa, A., Rafique, M.Z., and Caballero, J. (2015). “The Malicia Dataset: Identification and Analysis of Driven by Download Operations,” *International Journal of Information Security* (14:1) pp 15-33.

Nextgov (2015) “Heated House Hearing Offers New Clues into How Hackers Broke into OPM Networks,” available at: http://www.nextgov.com/cybersecurity/2015/06/heated-house-hearing-offers-new-clues-how-hackers-broke-opm-networks/115474/, accessed on May 21, 2016.

Nextgov (2016) “Here Are 10 of the Oldest IT Systems in the Federal Government,” available at: http://www.nextgov.com/cio-briefing/2016/05/10-oldest-it-systems-federal-government/128599/, accessed on Jan 5, 2017.

Office of Management and Budget (2016) *Federal Information Security Management Act Annual Report to Congress*, available at: https://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/final_fy14_fisma_report_02_27_2015.pdf, accessed on May 21, 2016.

Ouchi, W.G. and Maguire, M.A. (1975). "Organizational Control: Two Functions," *Administrative Science Quarterly* (20:4) pp. 559-569.

Ransbotham, S. and Mitra, S. (2009) “Choice and Chance: A Conceptual Model of Paths to Information Security Compromise,” *Information Systems Research* (20:1) pp. 121-139.

Ransbotham, S., Mitra, S., and Ramsey, J. (2012) “Are Markets for Vulnerabilities Effective?,” *MIS Quarterly* (36:1) pp. 43-64.

Reuters (2015) “Data Hacked from U.S. Government Dates Back to 1985: U.S. Official,” available at: http://www.reuters.com/article/us-cybersecurity-usa-idUSKBN0OL1V320150606, accessed on Jan 5, 2017.

Sambamurthy, V., and Zmud, R.W. (1999). "Arrangements for Information Technology Governance: A Theory of Multiple Contingencies," *MIS Quarterly* (23:2), pp. 261-290.

Simons, R. (1991). "Strategic Orientation and Top Management Attention to Control Systems," *Strategic Management Journal* (12:1), pp. 49-62.

Straub, D.W. (1990). “Effective IS Security: An Empirical Study.” *Information Systems Research* (1:3) pp. 255-276.

Subramanyam, R., Ramasubbu, N., and Krishnan, M.S. (2012) “In Search of Efficient Flexibility: Effects of Software Component Granularity on Development Effort, Defects, and Customization Effort,” *Information Systems Research* (23:3) pp. 787-803.

Suleiman, H. and Svetinovic, D. (2013) Evaluating the Effectiveness of the Security Quality Requirements Engineering (SQUARE) Method: A Case Study Using Smart Grid Advanced Metering Infrastructure,” *Requirements Engineering* (18:3) pp. 251-279.

Tanriverdi, H., Konana, P., and Ge, L. (2007) “The Choice of Sourcing Mechanisms for Business Processes,” *Information Systems Research* (18:3) pp. 280-299.

The Washington Post (2015) “Hacks of OPM Databases Compromised 22.1 Million People, Federal Authorities say,” available at: https://www.washingtonpost.com/news/federal-eye/wp/2015/07/09/hack-of-security-clearance-system-affected-21-5-million-people-federal-authorities-say/, accessed on May 21, 2016.

Wang, J., Gupta, M., and Rao, H.R. (2015) “Insider Threats in a Financial Institution: Analysis of Attack-Proneness of Information Systems Applications,” *MIS Quarterly* (39:1) pp. 91-112.

Weill, P. and Broadbent, M. (1998) *Leveraging the New Infrastructure: How Market Leaders Capitalize on Information Technology*, Boston MA: Harvard Business School Press.

Weill, P., and Ross, J. (2005) "A Matrixed Approach to Designing IT Governance," *MIT Sloan Management Review* (46:2) pp. 26-34.

Appendix A – Additional Estimation Tables

Table A1. Estimation without Non-Cabinet Agencies

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**					
Data Source	FISMA Report					PRC
Incident Types (DV in Log)	Total Incidents	Denial of Service	Improper Use	Unauthorized Access	Social Engineering	Unintentional Breach
	(1)	(2)	(3)	(4)	(5)	(6)
DME	-0.032**	-0.010	-0.016+	-0.032*	-0.131***	-0.017**
	(0.010)	(0.027)	(0.007)	(0.013)	(0.011)	(0.005)
Cloud	-0.005	-0.016*	-0.005	-0.011	-0.004	0.004
	(0.005)	(0.007)	(0.007)	(0.010)	(0.010)	(0.002)
Disperse	-10.222**	7.781	3.629+	-4.797	-17.712***	-0.479
	(3.015)	(9.313)	(2.001)	(7.445)	(3.997)	(0.720)
Diversity	-0.634***	1.306**	-0.862***	-0.911***	-0.380+	-0.041
	(0.111)	(0.336)	(0.105)	(0.119)	(0.179)	(0.114)
IT Invest	-69.472**	-308.515***	-59.156*	-70.930	-61.691	10.959
	(19.911)	(30.17)	(26.625)	(42.552)	(36.614)	(7.896)
Security Invest	-0.002	0.101	-0.284***	0.093	-0.084	0.015
	(0.043)	(0.090)	(0.052)	(0.057)	(0.113)	(0.011)
Budget	-0.602*	0.462	-0.560*	-0.699+	-1.594***	-0.255*
	(0.266)	(0.323)	(0.196)	(0.338)	(0.232)	(0.087)
Bureau	0.801	3.148	0.633	-3.982	-6.488***	-1.884*
	(2.246)	(2.515)	(1.392)	(2.478)	(1.278)	(0.823)
FTE	3.766***	-6.745***	6.504***	0.364	3.632+	-0.512
	(0.153)	(0.931)	(0.887)	(1.179)	(1.991)	(0.484)
Age	-0.129*	0.575***	-0.559***	0.225	0.187	0.119
	(0.059)	(0.121)	(0.086)	(0.138)	(0.239)	(0.070)
Defense	-0.003	0.056***	0.008	0.002	-0.003	-0.002
	(0.017)	(0.005)	(0.023)	(0.023)	(0.019)	(0.005)
Welfare	-0.001	0.056***	0.009	0.000	-0.018	0.003
	(0.017)	(0.006)	(0.024)	(0.024)	(0.019)	(0.005)
Law Enforce	0.010	0.168***	-0.024*	-0.018	0.001	-0.017**
	(0.013)	(0.029)	(0.011)	(0.037)	(0.039)	(0.006)
Management	-0.316***	0.087	-0.353***	-0.535**	-0.501+	-0.011
	(0.016)	(0.089)	(0.007)	(0.168)	(0.242)	(0.046)
Regulation	0.007	-0.069**	0.015	0.024	-0.018	0.011*
	(0.008)	(0.019)	(0.009)	(0.016)	(0.011)	(0.004)
F	5.96***	41.97***	70.91***	29.08***	7.04***	65.11***
Within R^2	0.403	0.662	0.453	0.224	0.377	0.382

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; $N = 68$; # of Agency = 17; Agency and year fixed-effects included; Standard errors are in parentheses.;

Table A2. Estimation with Different Windows for DME

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**							
Data Source	FISMA Report							
Window	Three-Year for DME				Seven-Year for DME			
Incident Types (DV in Log)	Total Incidents	Improper Use	Unauthorized Access	Malicious Code	Total Incidents	Improper Use	Unauthorized Access	Malicious Code
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DME	-0.008*	-0.005	-0.009*	-0.010*	-0.051*	-0.050*	-0.020	-0.053**
	(0.003)	(0.005)	(0.004)	(0.004)	(0.018)	(0.020)	(0.021)	(0.018)
Cloud	0.002	-0.012***	-0.008**	-0.012**	0.003**	-0.010***	-0.007**	-0.010*
	(0.001)	(0.002)	(0.002)	(0.004)	(0.001)	(0.002)	(0.002)	(0.004)
Disperse	-9.801***	-13.836***	-13.792***	-4.351+	-10.473***	-14.344***	-14.505***	-5.179*
	(1.377)	(0.292)	(0.649)	(2.395)	(1.708)	(0.530)	(0.931)	(2.290)
Diversity	-0.796***	-0.884***	-0.684***	-0.391	-0.646***	-0.738***	-0.614**	-0.231
	(0.102)	(0.083)	(0.142)	(0.395)	(0.118)	(0.093)	(0.154)	(0.362)
IT Invest	-26.545***	-24.480***	-35.749**	-82.023***	-29.272***	-27.071***	-37.182**	-84.986***
	(0.724)	(3.765)	(10.475)	(12.090)	(1.076)	(3.655)	(11.276)	(10.747)
Security Invest	0.176***	0.156*	0.393***	0.217	0.186***	0.164**	0.400***	0.229
	(0.015)	(0.062)	(0.027)	(0.159)	(0.023)	(0.057)	(0.031)	(0.164)
Budget	-0.419**	-0.205*	-0.335	-1.448***	-0.618**	-0.401*	-0.416	-1.657***
	(0.135)	(0.074)	(0.224)	(0.321)	(0.215)	(0.161)	(0.306)	(0.287)
Bureau	0.617	1.249*	-0.630	-0.034	0.009	0.664	-0.929	-0.689
	(0.801)	(0.532)	(0.622)	(1.349)	(1.010)	(0.744)	(0.898)	(1.199)
FTE	2.686***	4.155***	0.020	0.036	3.046***	4.432***	0.385	0.475
	(0.142)	(0.986)	(0.610)	(2.867)	(0.275)	(1.028)	(0.755)	(2.933)
Age	-0.040	-0.242*	0.287**	0.468	-0.021	-0.215*	0.272**	0.481
	(0.037)	(0.112)	(0.072)	(0.333)	(0.056)	(0.095)	(0.079)	(0.336)
Defense	-0.019	0.010	-0.007	-0.058***	-0.016	0.013	-0.008	-0.057***
	(0.013)	(0.018)	(0.020)	(0.008)	(0.015)	(0.021)	(0.020)	(0.009)
Welfare	-0.018	0.010	-0.008	-0.056***	-0.016	0.013	-0.009	-0.054***
	(0.013)	(0.019)	(0.021)	(0.007)	(0.015)	(0.022)	(0.021)	(0.008)
Law Enforce	0.002	0.023+	0.004	-0.020	-0.006	0.014	0.000	-0.029
	(0.010)	(0.012)	(0.016)	(0.020)	(0.006)	(0.009)	(0.015)	(0.022)
Management	-0.382***	-0.411***	-0.552**	-0.713***	-0.418***	-0.446***	-0.565**	-0.749***
	(0.018)	(0.046)	(0.145)	(0.173)	(0.014)	(0.043)	(0.155)	(0.151)
Regulation	0.023***	0.024**	0.023*	0.003	0.026**	0.026**	0.025*	0.006
	(0.005)	(0.006)	(0.009)	(0.004)	(0.007)	(0.008)	(0.011)	(0.005)
F	3847.43***	13645.50***	15.56***	118.00***	498.56***	547.45***	10.87***	3847.43***
Within R^2	0.265	0.354	0.287	0.313	0.285	0.369	0.287	0.319

+$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; $N = 96$; # of Agency = 24; Agency and year fixed-effects included; Standard errors are in parentheses.;

Table A3. Estimation with Alternative DME Variables

Method	**Driscoll and Kraay Fixed-Effects Estimation for Spatial Autocorrelation**							
Data Source	FISMA Report							
DME	Log (Total DME Spending in Five Previous Years)				% of DME Excluding Security-Related Spending			
Incident Types (DV in Log)	Total Incidents	Improper Use	Unauthorized Access	Policy Violation	Total Incidents	Improper Use	Unauthorized Access	Policy Violation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DME	-0.414^{+}	-0.624***	-0.579***	-0.745***	-0.049***	-0.055***	-0.027**	-0.068***
	(0.200)	(0.146)	(0.072)	(0.155)	(0.006)	(0.005)	(0.008)	(0.000)
Cloud	0.002	-0.012***	-0.008**	-0.015***	0.002	-0.011***	-0.008**	-0.015***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Disperse	-0.753***	-0.830***	-0.627***	-0.931***	-0.807***	-0.902***	-0.681***	-1.017***
	(0.095)	(0.070)	(0.137)	(0.140)	(0.088)	(0.076)	(0.142)	(0.140)
Diversity	-10.778***	-14.851***	-15.028***	-21.662***	-11.744***	-15.808***	-15.224***	-22.841***
	(1.391)	(0.562)	(0.834)	(1.945)	(1.584)	(0.741)	(1.073)	(1.977)
IT Invest	-26.405***	-23.907***	-35.450**	-23.575**	-26.918***	-24.712***	-36.240**	-24.533**
	(0.795)	(4.195)	(10.341)	(8.122)	(0.603)	(3.452)	(10.716)	(7.283)
Security Invest	0.197***	0.184**	0.421***	-0.039	0.202***	0.183**	0.410***	-0.039
	(0.026)	(0.049)	(0.030)	(0.096)	(0.029)	(0.054)	(0.033)	(0.098)
Budget	-0.500*	-0.323**	-0.447^{+}	-0.192***	-0.673***	-0.493***	-0.477^{+}	-0.401***
	(0.177)	(0.105)	(0.244)	(0.043)	(0.163)	(0.099)	(0.265)	(0.036)
Bureau	0.293	0.820	-1.067	1.715*	-0.124	0.435	-1.088	1.237*
	(0.936)	(0.625)	(0.764)	(0.726)	(0.920)	(0.639)	(0.812)	(0.568)
FTE	3.386***	4.982***	0.935	5.863***	3.381***	4.825***	0.582	5.686***
	(0.353)	(1.091)	(0.770)	(0.846)	(0.274)	(0.965)	(0.754)	(0.692)
Age	-0.073	-0.268*	0.248**	-0.344**	-0.038	-0.228*	0.270**	-0.294**
	(0.043)	(0.109)	(0.075)	(0.107)	(0.048)	(0.103)	(0.076)	(0.101)
Defense	-0.019	0.012	-0.006	0.028*	-0.016	0.014	-0.007	0.031*
	(0.014)	(0.021)	(0.021)	(0.012)	(0.015)	(0.021)	(0.021)	(0.012)
Welfare	-0.017	0.013	-0.007	0.031*	-0.015	0.014	-0.008	0.032*
	(0.015)	(0.021)	(0.022)	(0.012)	(0.015)	(0.021)	(0.021)	(0.012)
Law Enforce	0.000	0.019	0.000	0.033*	-0.005	0.015	0.000	0.028^{+}
	(0.009)	(0.012)	(0.017)	(0.015)	(0.008)	(0.013)	(0.018)	(0.016)
Management	-0.383***	-0.412***	-0.553**	-0.364***	-0.386***	-0.415***	-0.553**	-0.368***
	(0.015)	(0.045)	(0.142)	(0.089)	(0.007)	(0.039)	(0.146)	(0.083)
Regulation	0.020***	0.019**	0.019^{+}	0.017***	0.018**	0.018**	0.021*	0.015***
	(0.005)	(0.006)	(0.009)	(0.001)	(0.005)	(0.006)	(0.009)	(0.002)
F	1275.97***	776.59***	197.86***	129.28***	304.94***	754.20***	163.45***	87.94***
Within R^2	0.274	0.371	0.298	0.394	0.319	0.406	0.296	0.442

$^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; $N = 96$; # of Agency = 24; Agency and year fixed-effects included; Standard errors are in parentheses.;