Vendor-Originated Vulnerabilities and Data Breaches: A Large-Scale Empirical Study of Threshold Effects and Mitigation

Santhosh Srinivas
Virginia Tech

Leting Zhang
University of Delaware

Huigang Liang
University of Memphis[1]

ABSRACT

Drawing on Routine Activity Theory, this study examines how vendor-originated vulnerabilities influence data breach likelihood. Using data from 13,005 U.S. firms between 2010 and 2018, we develop a comprehensive measure of vendor-originated vulnerabilities by aggregating CVSS scores for all CVEs that may affect a firm's systems and security posture. Importantly, we find novel interactions among RAT's dimensions: timely external guardianship, such as vendor patch availability and detailed vulnerability knowledge dissemination, mitigates vulnerability risk, whereas external visibility, driven by media attention, intensifies breach likelihood. Furthermore, we unpack the CVSS score, highlighting that confidentiality and availability sub-scores within the CVSS framework are particularly predictive of breaches. Additionally, using non-parametric threshold detection methods, we identify a threshold effect around a firm level aggregate CVSS score of 4.6, further validated through real-world breach cases, and provide consistent causal evidence supporting our main findings through Difference-in-Differences and a set of robustness checks. Our study makes a distinct theoretical contribution and addresses a longstanding gap in cybersecurity research on third-party software risks.

KEYWORDS: Vendor-Originated Vulnerability, CVSS, CVE, Media Exposure, Patch, Data Breach, Routine Activity Theory, Cyber Supply Chain Risk Management.

[1] Santhosh Srinivas, ssrinivas@vt.edu; Leting Zhang, letingz@udel.edu; Huigang Liang, hilang1@memphis.edu

INTRODUCTION

As organizations broaden their digital presence, external software vendors and commercial off-the-shelf (COTS) solutions have become integral to everyday operations (Loukas, 2015). This growing reliance on external parties in building technology ecosystem can create a “black box” of third-party code that harbors hidden vulnerabilities, as illustrated by high-profile data breaches at Target, Marriott, and Equifax (Berghel, 2020; Malatesta III, 2016). Cybercriminals have capitalized on these vendor dependencies, driving a 180% increase in vulnerability exploitation for initial entry between 2022 and 2023, while “island hopping” attacks that breach smaller vendors before moving into larger targets have surged (Poireault, 2024; Esther Kezia, 2019; Yeboah-Ofori & Opoku-Akyea, 2019). Recent evidence suggests that 95% of production assets contain at least one vulnerability ranking in the top 5% for exploit likelihood, highlighting widespread and critical exposure across enterprises (Cyentia Institute & Kenna Security, 2022). The financial stakes of such cyber exploits are formidable: each data breach can cost a firm up to 4.88 million dollars, and total cybercrime expenses are predicted to reach 10.5 trillion dollars by 2025 (Thomson Reuters, 2024). In this evolving threat landscape, understanding how vendor-originated vulnerabilities translate into realized breaches is increasingly urgent.

Despite growing awareness of vendor-centric threats, a critical theoretical and empirical gap persists in understanding how vendor-originated vulnerabilities affect firm cybersecurity. Routine Activity Theory (RAT) posits that crime occurs when three elements converge: a suitable target, motivated offenders, and the absence of capable guardianship (Cohen & Felson, 1979). Traditionally, RAT in cybersecurity predominantly focuses on internally generated vulnerabilities or isolated external threats (Leukfeldt & Yar, 2016), rather than the cumulative risk introduced by a firm's entire third-party software ecosystem (Boyson, 2014; Jacobs et al.,

2020). Consequently, the manner in which external vendor dependencies shape firms' "target suitability" by aggregating vulnerabilities remains under-theorized.

This oversight contributes to a critical empirical vacuum. Organizations widely rely on vulnerability severity metrics, such as CVSS scores, to prioritize vulnerability management and patching (Goodman, 2024). Theoretically, these metrics facilitate effective prioritization, enabling firms to proactively address vulnerabilities and mitigate susceptibility to attacks. However, the remediation rates remain low, with firms addressing only 16% of vulnerabilities on a monthly basis (Cyentia Institute & Kenna Security, 2022). Persistent high-profile breaches, such as the 2017 Equifax incident where vendor-originated vulnerabilities remained unpatched, highlight significant theoretical shortcomings in understanding vulnerabilities originating beyond organizational boundaries (Yeboah-Ofori & Opoku-Akyea, 2019; August et al., 2019). Such incidents illustrate the substantial challenges firms face, underscoring potential inefficiencies in current vulnerability management processes. Thus, vulnerability management must be reconceptualized theoretically as an inter-organizational issue (Boyson, 2014; Jacobs et al., 2020). Existing theoretical frameworks must better explain how external dependencies influence vulnerability prioritization and management outcomes, particularly given practical constraints on comprehensive patching. A theoretically grounded understanding is crucial to clarifying why certain vendor-originated vulnerabilities remain inadequately addressed despite their known severity, thus guiding organizations toward more effective cybersecurity strategies.

While vendor-originated vulnerabilities define the initial conditions for breach risks, the effectiveness of a firm’s vulnerability management processes ultimately determines whether these risks materialize into actual breaches. However, firms face significant challenges in vulnerability management due to their inherent dependencies on external players, including

software vendors for timely patch releases and security infomediaries for critical vulnerability disclosures and guidance (Kannan & Telang, 2005; Chen et al., 2007). Further complicating these challenges is media coverage, which publicly disseminates vulnerability information, unintentionally increasing its visibility and attractiveness to cyber attackers, complicating firms' vulnerability prioritization and remediation processes (Holt et al., 2020; Zorz, 2019).

When such external parties, including software vendors and security infomediaries, involved in vulnerability management processes, fail, or operate ineffectively, the risks associated with vendor-originated vulnerabilities escalate. Such failures extend the window for attackers to exploit vulnerabilities, significantly elevating breach risks (August et al., 2019; Jacobs et al., 2020). Although existing literature has explored certain aspects of vulnerability management, such as the economic incentives influencing patch deployment (Arora et al., 2010; August et al., 2019), comprehensive empirical investigations into these multifaceted dimensions remain absent. To address this gap and comprehensively understand how aggregated vendor-originated vulnerability severity impacts breach likelihood, we pose our research question:

Does a firm's vendor-originated vulnerability impact the likelihood of data breaches? If so, how do key external factors, including patch availability, patch release time, vulnerability knowledge, and media coverage, moderate this impact?

To theorize this question, we extend RAT by highlighting external factors beyond malicious actors, such as vendor-originated vulnerabilities, that significantly expand firms' suitability as targets. Further, we emphasize the critical role external actors, including vendors and infomediaries, play in shaping two key RAT dimensions: target visibility and capable guardianship. Our augmented RAT framework thus underscores external guardianship provided by vendors through timely patching and by infomediaries through vulnerability disclosures. Guided by this theoretical framework, we conduct comprehensive analyses. Using a novel

dataset comprising 13,005 U.S. firms between 2010 and 2018, we merge site-level vendor data with the National Vulnerability Database (NVD) to capture each firm's aggregated severity of vendor-specific vulnerabilities. Our empirical analysis reveals that higher aggregated vendor-originated vulnerability severity significantly elevates breach likelihood. Furthermore, we find that patch availability, and comprehensive vulnerability knowledge mitigates breach risks, whereas delayed patches and heightened media attention intensify these risks. In addition, we unpack the CVSS score and identify that Confidentiality and Availability sub-scores within the CVSS framework are particularly predictive of breaches. We also observe a distinct threshold around a CVSS score of 4.6. Based on the threshold, Difference-in-Differences and a set of robustness checks yield causal evidence supporting our main findings.

Our study offers several principal contributions. First, we extend RAT by empirically demonstrating how external vendors and infomediaries critically influence firm-level breach risk through their roles in patch availability, vulnerability disclosure, and inter-organizational coordination. Second, we uniquely examine the interactions among RAT dimensions, which have traditionally been explored independently rather than jointly, offering a richer theoretical explanation of cybercrime dynamics. Third, we introduce a comprehensive measure of firm-level, vendor-originated vulnerability severity, facilitating related empirical investigations by linking individual CVEs and CVSS to firms' vendor ecosystems. Fourth, we enhance the existing understanding of the CVSS framework by empirically identifying a threshold effect and pinpointing critical CVSS metrics, particularly Confidentiality and Availability, that significantly predict breaches, providing benchmarks for targeted patch prioritization. Lastly, we offer practical implications for vendor oversight, supply chain risk management, vulnerability prioritization practices, and strategic cybersecurity planning.

BACKGROUND, THEORY, AND HYPOTHESIS

Cyber Supply Chains

Organizations do not possess all the resources necessary for their operation; therefore, they must acquire critical inputs from external sources (Pfeffer & Salancik, 1978). Therefore, understanding the various contingencies in the external environment where organizations interact is essential, with dependence and uncertainty being two key conditions that shape a firm's reliance on external partnerships (Hillman et al., 2009; Pfeffer & Salancik, 1978). In the context of digitization, organizations increasingly rely on Commercial Off the Shelf (COTS) solutions provided by external software vendors, which can accelerate deployment yet introduce added risk (Zhu & Zhou, 2012). On the one hand, engaging with external providers creates substantial value (Cho et al., 2013; Nagle, 2019; Taştan & Gönel, 2020). On the other hand, COTS software creates shared burdens and uncertainty for both the vendor and the client firm. Design shortcuts or failures to adhere to development standards can lead to accumulating technical debt, undermining reliability (Ramasubbu & Kemerer, 2016, Özkan, 2019). This uncertainty is magnified in vendor supply chains, where a single vulnerable product can be deployed at multiple client sites, compounding the risk of widespread exploits (Jacobs et al., 2020). As a result, Cyber Supply Chain Risk Management (CSCRM) has emerged as a cross-functional approach to address these externally introduced vulnerabilities, bringing together cybersecurity, enterprise risk management, and supply chain management perspectives (Boyson, 2014). While emerging technology vulnerabilities continue to expose organizations to significant cyber risks, few studies have systematically linked vendor-originated vulnerabilities to organization-level breach incidents and explored the factors that affect related cybercriminals.

This gap partly reflects the difficulty of attributing vulnerabilities to specific firms, especially when the same COTS product is deployed at thousands of client sites. To address

these challenges, we develop an innovative measure that aggregates all vulnerabilities within a firm's ecosystem and provide insights into the impact of vendor-originated vulnerabilities on organizations' data breaches, thereby contributing to cyber supply chain literature through a vendor-centric lens.

Cybercrime and Routine Activity Theory

Over the past decades, the widespread adoption of the internet and digital technologies has enabled increasingly sophisticated cyberattacks. Despite this technological evolution, cybercriminals' underlying motivations have largely remained stable (Rightley et al., 2023). Typically viewed as rational actors, cybercriminals weigh the potential benefits against the costs of their illicit activities (Cohen & Felson, 1979; Gibbs, 1968). Benefits include financial rewards from selling vulnerabilities or compromised data (Ablon & Libicki, 2015; Wegberg et al., 2020), moral satisfaction derived from perceived justice (Benjamin et al., 2016; D'Arcy et al., 2020), or recognition gained by demonstrating hacking prowess (Zhang et al., 2015). The associated costs involve the risk of detection, apprehension, and the time and resources required for intrusion (D'Arcy et al., 2009; Hui et al., 2017).

Routine Activity Theory (RAT) (Cohen & Felson, 1979) offers a robust framework for understanding these cybercrime dynamics. It posits that crime occurs when three elements converge: (1) a suitable target (valuable, visible, and accessible), (2) a motivated offender, and (3) the absence of a capable guardian. While existing studies frequently examine these factors independently (e.g., Hui et al., 2017; Wang et al., 2015; D'Arcy et al., 2009), rarely exploring their interactive effects. To address the gap, we extend RAT by shifting the theoretical focus from internal organizational vulnerabilities and defenses to external vulnerabilities introduced by third-party vendors, alongside the externally provided guardianship mechanisms addressing these risks.

Security Vulnerability Management

Software applications inevitably contain security vulnerabilities that can be exploited by malicious actors. These flaws arise due to programming errors (Anu et al., 2020), the inherent complexity of software structure (Chowdhury & Zulkernine, 2010), and insufficient testing time (Parwal, 2024). Firms adopting COTS solutions may discover too late that vendor-originated vulnerabilities pose critical risks (Pratt, 2022). Research on vulnerability management has focused on mitigating risks at different stages, including incentive structures for vulnerability discovery (Kannan & Telang, 2005; Ozment, 2004; Ransbotham, 2010; Zhao et al., 2018), vulnerability disclosure (Arora et al., 2008; Cavusoglu et al., 2007; Ruohonen & Allodi, 2018; Sen et al., 2020), and optimal patching policy (August & Tunca, 2008; Cavusoglu et al., 2008; August et al., 2019).

Beyond IT vendors, other external entities also shape how quickly and widely vulnerability information disseminates, acting as infomediaries (Kannan & Telang, 2005). Although the roles of vendors and infomediaries in cybersecurity are widely recognized (Arora et al., 2010; Cavusoglu et al., 2007; Ruohonen & Allodi, 2018), there is a shortage of empirical research on how the actions of external software suppliers. Our study fills in the gap by incorporating multiple vulnerability dimensions based on the widely used CVSS framework and examining contingency factors that include vendor patch availability, patch timing, and vulnerability knowledge. We provide valuable insights into how firms can effectively defend against exploitation attempts. Our results offer both theoretical and practical contributions to the security vulnerability management literature, underscoring the significant role that vendors and external infomediaries play in mitigating or exacerbating organizational exposure.

HYPOTHESIS DEVELOPMENT

Building on Routine Activity Theory in the context of security vulnerability management, we develop a research model to examine how vendor-originated vulnerabilities influence a firm's likelihood of experiencing data breaches. Our augmented RAT framework explicitly identifies externally introduced vulnerabilities as crucial determinants of firms' suitability as cyberattack targets. Specifically, we investigate whether higher aggregated vendor-originated vulnerabilities, externally originating from third-party software providers, raise breach risks. We further propose that this effect is moderated by several key management factors such as patch availability, patch release time, media coverage, and vulnerability knowledge. The research model is illustrated in Figure 1.

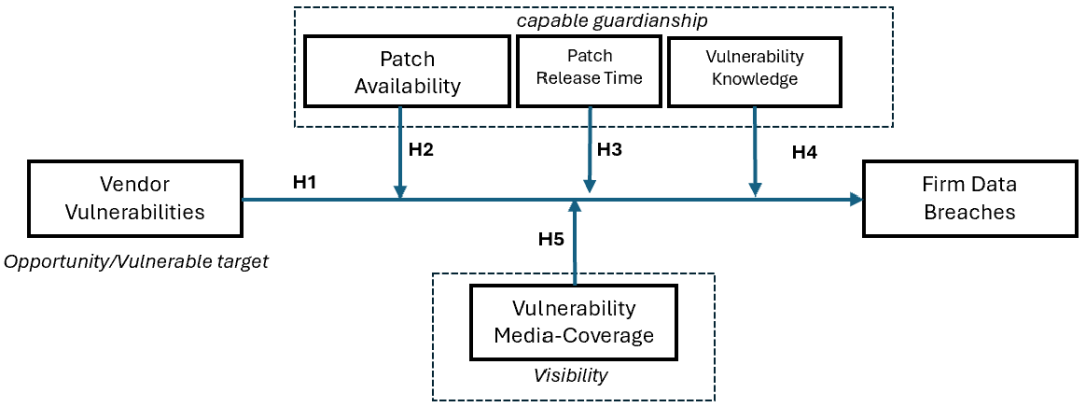


Figure 1. Research Model

As organizations increasingly pursue technological progress, they often rely on multiple external vendors to fulfill a wide range of technology requirements, from basic email services to complex ERP implementations. These partnerships are vital for operational success but also introduce additional security risks, broadening the organization's attack surface (Charney et al., 2011; Keskin et al., 2021). A recent survey of approximately 230,000 firms found that nearly 98% had at least one third-party partner who had experienced a breach, highlighting the widespread prevalence of third-party induced data compromise threats (Cyentia, 2023). One

primary form of vendor engagement is via Commercial Off-the-Shelf (COTS) software, which, despite its convenience and cost benefits, often functions as a "black box" that conceals hidden vulnerabilities (Chen et al., 2007; Özkan & Bulkan, 2019). In some cases, vulnerabilities in smaller embedded components within larger COTS products remain unpatched or overlooked, thereby compounding an organization's risk profile (Boyens et al., 2022). Routine Activity Theory (Cohen & Felson, 1979) helps explain why these vendor-originated vulnerabilities are so consequential: unaddressed or undisclosed vulnerabilities create more "suitable targets" for cybercriminals (August et al., 2019; Jacobs et al., 2020). Although traditional application of RAT in cyber domain primarily focuses on internally-generated vulnerabilities, externally-driven vulnerability conditions often lie beyond a firm's direct control (Boyson, 2014; Esther Kezia, 2019; Yeboah-Ofori & Opoku-Akyea, 2019). Therefore, we propose our baseline hypothesis on which more sophisticated hypotheses are developed:

H1. *Firms with a higher aggregated vendor-originated vulnerability experience a higher data breach likelihood.*

While vendor-originated vulnerabilities are a core risk factor, the extent to which they lead to actual breaches hinges on whether they can be fully addressed by software vendor. Since software vulnerabilities typically stem from flaws in internal code and logical errors, firms are generally incapable of developing patches on their own (August & Tunca, 2008; Cavusoglu et al., 2008). Researchers have consistently underscored the stakes involved in patch availability. For instance, Chen, Boehm, and Sheppard (2007) show that official patches can remove entire sets of potential attack paths in Commercial Off-the-Shelf (COTS) systems, thereby limiting opportunities for successful intrusions. By contrast, if vendors fail to release patches or do not support older software versions, client firms must rely on temporary workarounds or remain fully

exposed (Özkan, 2019; Schrader, 2024). This lack of reliable patch availability creates "silent risks" that attackers can exploit, amplifying the harm of vendor-originated vulnerabilities (Schryen, 2009). From our extended RAT perspective, prompt external remediation reduces a firm's attractiveness as a target by eliminating the window of exploitation available to attackers (Wang et al., 2015). Therefore, effective external guardianship, manifested as reliable and timely patch availability, mitigates vendor-originated vulnerabilities before they materialize into breaches. Therefore, explicitly grounded in our augmented RAT framework, we propose:

H2. *Increased patch availability across all vendors of a firm moderates the relationship between vendor-originated vulnerabilities and data breach likelihood, such that higher patch availability weakens the relationship.*

Given the software dependency between software vendor and firms, the speed of its release is critical for minimizing the risks posed by vendor-originated vulnerabilities. Although the importance of timely patching is well acknowledged, some vendors move swiftly to release patches, yet others do not (Schryen, 2009). Furthermore, the average patch release time was about 97 days in 2019, with enterprise software experiencing especially long delays (Roumani, 2021). The variations imply that vendors face tradeoffs: rushing a patch can introduce new flaws, while deferring a fix prolongs vulnerability (Cavusoglu et al., 2004). The overall cost of patch development varies with vulnerability severity, vendor resources, and market expectations, complicating the decision process (Arora et al., 2006). Extending traditional RAT, our augmented framework explicitly positions patch release speed as an external guardianship mechanism provided by software vendors. Slow external guardianship prolongs vulnerability exposure, thereby increasing a firm's attractiveness as a cybercrime target. Consequently, when vendors delay patches, client firms must rely on cumbersome interim solutions or remain

continuously exposed (Özkan, 2019; Schrader, 2024). Prolonged unpatched periods increase the likelihood of successful intrusions, as each additional day without a patch increases exploitation opportunities (Ransbotham, 2010; Ransbotham & Mitra, 2009, Amelia, 2025). Thus, aligned explicitly with our extended RAT perspective, we propose:

H3. *Patch release time moderates the relationship between vendor-originated vulnerabilities and data breach likelihood, such that longer patch release time amplifies the relationship.*

In additional to vendors, infomediaries play a critical role in disseminating information about vendor-originated vulnerabilities, thereby influencing their impact on security outcomes. One main type of information is vulnerability knowledge, reflected in the number of references for each CVE entry, offers defenders rich technical guidance on vulnerability causes, implications, and remediation strategies (Poireault, 2024). Public repositories such as the U.S. National Vulnerability Database, vendor advisories, and security bulletins aggregate these references, helping firms assess risk, prioritize patches, and shore up defenses (Trabelsi, 2015; Dissanayake, 2022). Although attackers could theoretically mine this information for exploit development (Emmitt, 2020; Beardsley, 2022), they more often turn to underground forums for ready-made exploits. In practice, then, public vulnerability knowledge functions as an external extension of the "capable guardian" in Routine Activity Theory, with third-party actors supplying detailed, actionable remediation roadmaps that firms cannot generate internally. By shortening the time between disclosure and fix deployment, this external guardianship narrows the window of attacker opportunity and should weaken the link between vendor-originated vulnerabilities and actual breaches. Accordingly, we propose:

H4. *Vulnerability knowledge moderates the relationship between vendor-originated vulnerabilities and data breach likelihood, such that higher levels of vulnerability knowledge weaken that relationship.*

While vulnerability knowledge is essential for organizations to design and implement effective vulnerability management strategies (Nir 2023), other types of vulnerability information can be leveraged by malicious hackers who are seeking exploitation opportunities (Kannan & Telang, 2005; Crews, 2018). A central proposition in the Routine Activity Theory perspective is that a target's "visibility" substantially influences its susceptibility to attack (Leukfeld & Yar, 2016). In the cybersecurity context, extensive media attention heightens the profile of a vulnerability, making it more appealing to potential attackers (Holt, 2020). Although technical disclosures may focus on patching procedures or risk assessments, mainstream news outlets and social media discussions often spotlight the severity and potential impact of a vulnerability (Zorz, 2019). Threat modeling and attack path analyses likewise find that when certain vulnerabilities gain "high popularity," they become focal points for intensified scrutiny by cybercriminals (Chen, Boehm, & Sheppard, 2007). Moreover, Grover and Kohli (2013) caution that organizations should carefully manage how much information they expose to the public, since greater "system visibility" can inadvertently disclose sensitive details to adversaries. In the cybersecurity domain, attackers monitor diverse news sources, technology blogs, and social media for leads, making increased publicity even more hazardous (Liang et al. 2025). Therefore, drawing on these insights, we propose:

H5. *Media coverage moderates the relationship between vendor-originated vulnerabilities and data breach likelihood, such that increased media coverage amplifies the relationship.*

In summary, we propose hypotheses on how vendor-originated vulnerabilities impact firms' data breach outcomes. Furthermore, we hypothesize how vendor-related contingent factors (i.e., patch availability and patch release time) and infomediary-related contingent factors (i.e., vulnerability knowledge and media coverage) moderate this impact.

METHODOLOGY

Data and Variables

To test our hypotheses, we develop a unique dataset by merging multiple sources of archival data. Our primary data source for all independent variables came from two main databases: the National Vulnerability Database (NVD) which hosts all publicly known security vulnerabilities and Aberdeen's Ci Technology Database (CiTDB), a comprehensive database on IT expenditures, technology adoption, vendor details, employee and IT staff information. CiTDB is widely used in Information Systems research (e.g., Xue et al., 2021, Kim et al., 2017).

Dependent Variable – Data Breach Likelihood. A data breach is the intentional or unintentional release of secure or private or confidential information to an untrusted environment without appropriate authorization. It occurs when sensitive, protected, or private information is copied, communicated, viewed, stolen, or utilized by unauthorized agents. We obtained data breach information for all firms from the Privacy Rights Clearinghouse (PRC), an organization dedicated to engaging, educating, and empowering individuals to protect their privacy. Many studies have quantified data breaches using this source (e.g., Wang, 2024). The dependent variable is a binary variable indicating breach occurrence each year.

Independent Variable - Vendor-Originated Vulnerability. Firm's vendor-originated vulnerability severity is the key independent variable in this study, representing the overall security weaknesses within a firm's technology ecosystem driven by vendors. To accurately measure this vulnerability severity, we integrate data from CiTDB and NVD. The CiTDB

provides detailed information on each firm's technical systems, including system names, types/makes, and associated vendor names across various sites. Meanwhile, the NVD serves as a comprehensive repository of publicly known security vulnerabilities, offering detailed records for each vulnerability, such as descriptions, affected products, associated patches, references, and their respective Common Vulnerabilities and Exposures (CVE) identifiers. Each CVE in the NVD is assigned a Common Vulnerability Scoring System (CVSS) score, which quantifies the severity of the vulnerability on a scale from 0 to 10. To calculate the score for vendor-originated vulnerability (VoV) for each firm i annually, we first identify all systems and their corresponding vendors at each site using CiTDB data. We then match these vendors to the NVD to identify all associated CVEs and retrieve the corresponding CVSS scores for each CVE. The score for VoV is obtained by averaging CVSS scores of all CVE's associated with all vendors at each site and then further aggregating this average across all sites within the firm for each year. It can be represented as below

$$VoV_i = \frac{1}{N_i} \sum_{s=1}^{S} \sum_{v=1}^{V_s} \sum_{j=1}^{J_{s,v}} CVSS_{s,v,j} \qquad (1)$$

Where VoV_i is average score for Vendor-Originated Vulnerability for firm i, S is Total number of sites for the firm i, V_s is Total number of vendors used at site s, $J_{s,v}$ Total number of CVEs associated with vendor v at site s, $CVSS_{s,v,j}$ is the CVSS score for the j-th CVE of vendor v at site s, N_i is Total number of CVEs for firm i.

Controls: Control variables were derived from the CiTDB, including the number of sites, IT budget, Revenue, and employee size. Collectively, firms in our sample experienced 1158 breach incidents. Overall, our dataset included information for 13,005 American firms from 2010 to 2018. Of these firms, 683 firms were public, and the remaining were private. After aligning

data between CiTDS and breach sources, we constructed a unique dataset consisting of 115,425 firm-year observations. Appendix A, Table A1 shows the pairwise correlations between the variables. All variables' sources and measures are summarized in Appendix A, Table A2. Table A3 presents descriptive statistics of all variables.

Moderator 1 – Patch Availability[2]. To measure patch availability, we calculated an aggregate patch availability score for each firm annually. For every CVE identified within a firm's technology ecosystem, we determined whether a patch was available by referencing the NVD's "References" section, which is updated upon the release of a patch (Please see Appendix D for more comprehensive details on the conceptualization of Patch Availability). If a patch was available for a particular CVE, we assigned a value of 1; otherwise, we assigned a value of 0. These binary indicators were then aggregated across all CVEs, vendors, and sites within each firm for each year, mirroring the aggregation process used for the Vendor-Vulnerability Score. This firm-level patch availability score captures the overall availability of patches relative to the firm's total vendor-vulnerabilities, providing a comprehensive measure of how effectively a firm could address its security weaknesses through patching across its entire technology ecosystem. Although not perfectly efficient, this information in NVD is still utilized on daily basis to find information on patches from across the world (Wunder et al., 2024). The patch availability equation is:

$$Patch\ Availability_i = \frac{\sum_{j=1}^{N_i} 1(Patch_j)}{N_i} \qquad (2)$$

[2] We use an example to demonstrate how to measure Patch Availability. For instance, if a firm faces 100 vulnerabilities in a given year and 85 of them have an available patch, the Patch Availability measure for that firm is 85/100 ≈ 0.85.

Where N_i is the total number of vulnerabilities (CVEs) that firm i faces in a given year, and $1(Patch_j)$ is an indicator function that equals 1 if a patch is available for vulnerability j and 0 otherwise. This measure ranges from 0 to 1, representing the proportion of CVEs that have patches.

Moderator 2 – Patch Release Time. We measured the time lag between the publication of each vulnerability and the availability of its corresponding patch, as indicated by the "patch" tag in the NVD "References" section (Please see Appendix D for more comprehensive details on the conceptualization of Patch availability speed). For every CVE identified within a firm's technology ecosystem, we calculated the number of days between the CVE's publish date and the date a patch became available. These time lags were then aggregated across all CVEs, vendors, and sites within each firm annually, mirroring the aggregation process used for the Vendor-Originated Vulnerability Score and Patch Availability Score. This firm-level patching speed score captures the critical window during which a firm's ecosystem remains vulnerable to potential exploitation. A shorter patching speed score indicates a more rapid response to vulnerabilities, thereby minimizing the time frame for potential attacks, whereas a longer score suggests slower patching practices and increased exposure risk. The patch release time is represented as below.

$$Patch\ Release\ Time_i = \frac{\sum_{j=1}^{N_i} T_j}{N_i} \qquad (3)$$

Where N_i is the total number of vulnerabilities (CVEs) that firm i faces in a given year, and T_j represent the time (in days) between vulnerability j's publish date and the date its patch becomes available. Then the firm-level Patch Release Time measure is simply the average of

these time lags[3]: A higher $Patch\ Release\ Time_i$ indicates that patches become available more slowly (delayed) on average, increasing a firm's window of exposure.

Moderator 3 – Vulnerability Knowledge. To measure vulnerability knowledge, we counted the number of information/resources available for each vulnerability (Please see Appendix D for more comprehensive details on the conceptualization of Vulnerability Knowledge). Specifically, we utilized the number of hyperlinks in the "References to Advisories, Solutions, and Tools" section of each vulnerability entry in the NVD database. These hyperlinks direct users to relevant advisories, patches, or related solutions, indicating the breadth of external resources available to remediate vulnerabilities. For every CVE identified within a firm's technology ecosystem, we counted the number of references provided in this section. These counts were then aggregated across all CVEs, vendors, and sites within each firm annually, mirroring the aggregation processes used for the Firm Vendor-Originated Vulnerability Score and Patch Availability Score. This firm-level vulnerability knowledge score captures the overall knowledge of external guidance and tools relative to the firm's total vulnerabilities. These references are among the most widely used sections of a CVE entry in NVD, with studies showing that 85 percent of practitioners rely on advisory and patch information for their security workflows (Miranda, 2023). The average vulnerability knowledge[4] is represented as below.

$$Vulnerability\ Knowledge_i = \frac{\sum_{j=1}^{N_i} V_j}{N_i} \qquad (4)$$

[3] For instance, if a firm has 5 CVEs with time lags of 3, 5, 6, 10, and 11 days, then $Patch\ speed_i$= (3+5+6+10+11)/5=7 days.

[4] For instance, if a firm has 5 vulnerabilities with reference hyperlink counts of 2, 3, 3, 5, and 1, then $Vulnerability\ Knowledge_i$=(2+3+3+5+1)/5=2.8. A larger value indicates that a firm's vulnerabilities tend to have more comprehensive external documentation, potentially aiding remediation/aiding attackers.

Where N_i is the total number of vulnerabilities (CVEs) that firm i faces in a given year, and $V_j$represent the number of references (hyperlinks) listed for vulnerability j. A higher $Vulnerability\ Knowledge_i$ indicates that the firm's vulnerabilities are accompanied by more comprehensive external documentation and advisories.

Moderator 4 – Media Coverage. To measure media coverage that a firm's vulnerabilities receive, we computed a media coverage score by counting the number of web articles mentioning each CVE using the Google Search API. The search window extended from one month before the vulnerability's publication date to six months after, capturing instances where vulnerability details circulated prior to their formal listing in the NVD. These articles include sources such as hacking forums, GitHub pages, and technical news outlets, where discussions may involve exploitation details, root causes, fixes, and attack methodologies. Additionally, typical media articles can popularize CVEs globally, increasing their visibility. We then aggregated the total number of articles across all CVEs, vendors, and sites for each firm annually, resulting in a firm-level media coverage metric. The media coverage metric can be represented as below[5]:

$$Media\ Coverage_i = \frac{\sum_{j=1}^{N_i} M_j}{N_i} \qquad (4)$$

Where N_i is the total number of vulnerabilities (CVEs) that firm i faces in a given year, and M_j represent the number of articles referencing vulnerability j. A higher $Media\ Coverage_i$ implies that, on average, a firm's vulnerabilities received more public and industry attention.

[5] For instance, if a firm has 5 vulnerabilities with article counts of 0, 2, 2, 10, and 6, then Media Coverage is calculated as (0+2+2+10+6)/5=4. Because the Google Search API covers both mainstream and niche outlets, some widely publicized CVEs yielded extremely large numbers of results, raising the overall average of this measure.

Control Variables. CiTDB provides essential firm metrics such as Number of Employees, Revenue, and Number of Sites, which act as proxies for firm size and profitability. We include Revenue as a control variable because hackers often target profitable and reputable firms (Dahmash et al., 2009). The Number of Employees accounts for variations in firm size, while the Number of Sites reflects the geographic and operational spread of a firm Additionally, we control for the number of CVEs and the number of systems a firm possesses. Controlling for number of CVEs distinguishes the impact of a firm's vendor-originated vulnerabilities from the total number of vulnerabilities, which captures volume. Similarly, controlling for the number of systems accounts for technological scale differences across firms, ensuring that our findings reflect true relationships rather than being influenced by larger infrastructures. Lastly, a firm's IT budget reflects its commitment to technological assets and significantly influences its security posture (Sen & Borle, 2015). Including the IT budget as a control variable allows us to account for the resources allocated to security measures.

Model Specifications

We employ the following Linear Probability Model (LPM) to test the hypothesis that a firm's vendor-originated vulnerabilities lead to a higher likelihood of data breaches, moderated by patch availability, patch release time, vulnerability knowledge, and media coverage.

$$y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 M_{i,t} + \beta_3 (X_{i,t} \cdot M_{i,t}) + \sum \beta_k C_{i,t} + \alpha_i + \gamma_t + \varepsilon_{i,t} \qquad (3)$$

where the dependent variable $y_{i,t}$ is the binary variable indicating if firm i had a breach event year t. $X_{i,t}$ is the independent variable - the firm's average vendor-originated vulnerabilities in year t. $M_{i,t}$ is the moderator – the moderators, including patch availability, patching speed, media coverage, and vulnerability knowledge of firm i in year t. $\beta_k C_{i,t}$represents a collection of time-varying control variables. α_i represents firm fixed effects, while γ_t denotes

year-fixed effects. $\varepsilon_{i,t}$ is the idiosyncratic error term. We control for both firm and year fixed effects to address concerns about omitted variables related to firm-specific attributes and temporal shocks.

RESULTS

We begin by inspecting the model-free relationship between firm's vendor-originated vulnerabilities and data breach likelihood across all firms over time (A detailed model-free descriptive and trend analysis of our data is available in Appendix D). As shown in Figure 2, there appears to be a general trend where an increase in firms' vendor-originated vulnerabilities is associated with a corresponding rise in the occurrences of data breaches. This visual correlation suggests that higher vulnerability levels in firms' technological ecosystems may contribute to an elevated risk of data breaches.

Main effect: We then evaluate the main and interaction models to test the hypothesis that higher vendor-originated vulnerabilities lead to higher data breach likelihood. Table 1 presents the results from our Linear Probability Model. To account for potential heteroscedasticity, we estimate clustered robust standard errors for the regression coefficients at the firm level, ensuring more reliable inference (Wooldridge, 2010). The model also controls both firm and year fixed effects. Importantly, we control for cumulative breach count, capturing a firm's historical breach incidents, as a control variable across all models. This accounts for potential path dependency, where past breaches may influence both vulnerability management and subsequent breach likelihood. We also control for the firm's historic vendor-originated vulnerability levels, accounting for prior patterns in our key independent variable. This helps address endogeneity by acknowledging that previously high or low vulnerabilities may systematically influence current security outcomes. As seen in Table 1, a higher Firm vendor-originated vulnerability score is associated with higher data breach likelihood ($\beta=0.079$, $p<0.05$), corroborating our Hypothesis 1.

Next, we introduce all our moderators and other firm attributes as controls and see similar results (β=0.093, p<.05), further strengthening our H1.

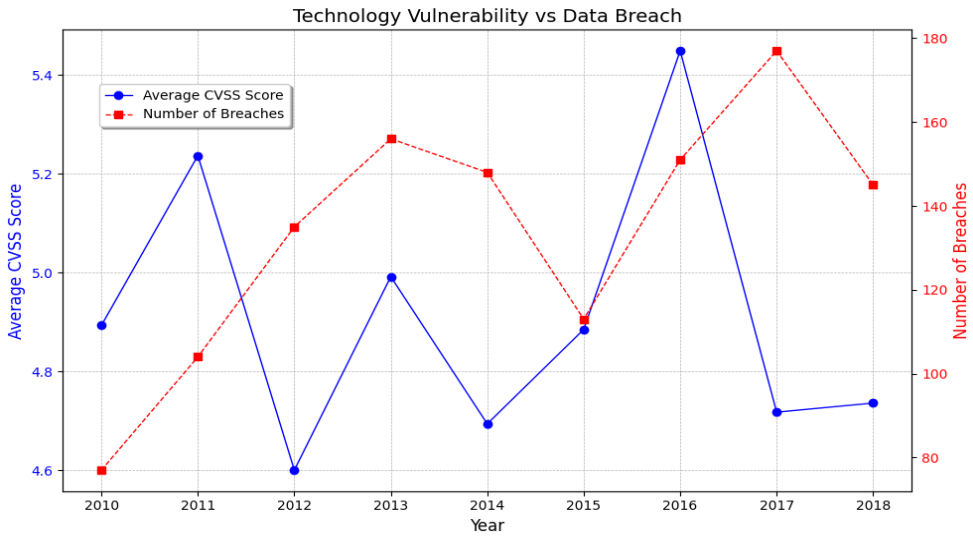


Figure 2: Firm Vendor-Originated Vulnerability vs YoY breach Trend

Interaction effects: We then introduce our moderators and test the moderating effect of patch availability, patch release time, vulnerability knowledge, and media coverage. First, as seen in Table 1, patch availability exhibits a significant negative interaction effect on data breaches ($\beta = -1.142$, $p < 0.01$), supporting H2. This finding demonstrates that firms whose vendors provide patches for a larger proportion of vulnerabilities face fewer breaches, suggesting that more comprehensive patch coverage alleviates the damaging effects of vendor-originated vulnerabilities.

Second, patch release time shows a significant positive interaction coefficient ($\beta = 0.056$, $p < 0.01$), consistent with H3. Longer patch release intervals extend the window during which attackers can exploit unpatched vulnerabilities, thereby increasing firms' breach risk.

Turning to infomediary‐related factors, vulnerability knowledge has a negative effect on breach likelihood ($\beta = -0.066$, $p < 0.05$). This supports H4 and a higher level of knowledge on vulnerabilities assist defenders by enabling more efficient and timely remediation.

Finally, we find that media coverage is positively associated with data breaches ($\beta = 0.049$, $p < 0.01$), supporting H5. This underscores that widespread public discussion and media attention for vendor-originated vulnerabilities can attract malicious actors' attention, effectively complicating firms' mitigation strategies and heightening the probability of a breach.

Table 1. Hypothesis testing results

DV: Data Breaches	Base Model		Interaction			
	(1)	(1.1)	(2)	(3)	(4)	(5)
	Base Model	Base Model with Controls	Patch Availability	Patch Release Time	Vulnerability Knowledge	Media Coverage
Vendor-Originated Vulnerability (VoV)	0.079* (0.034)	0.093* (0.037)	0.171** (0.044)	-0.260* (0.109)	0.530** (0.159)	-0.414+ (0.249)
Patch Availability		-0.846+ (0.504)	5.183* (2.115)			
VoV × Patch Availability			-1.142** (0.405)			
Patch Release Time		0.016 (0.034)		-0.312** (0.100)		
VoV × Patch Release Time				0.056** (0.017)		
Media Coverage		0.064 (0.042)				-0.192 (0.134)
VoV × Media Coverage						0.049* (0.024)
# Vulnerability Knowledge (VK)		-0.109 (0.095)			0.261 (0.172)	
VoV × Vulnerability Knowledge					-0.066** (0.024)	
Cumulative Breach	-0.245** (0.011)	-0.245** (0.011)	-0.245** (0.011)	-0.245** (0.011)	-0.245** (0.011)	-0.246** (0.011)
Historic VoV	-0.002 (0.032)	-0.003 (0.032)	0.001 (0.032)	-0.003 (0.032)	-0.000 (0.032)	-0.003 (0.032)
Log IT Budget	0.340** (0.050)	0.336** (0.050)	0.336** (0.050)	0.335** (0.050)	0.337** (0.050)	0.335** (0.050)

Log Employees	-0.335** (0.103)	-0.348** (0.103)	-0.338** (0.103)	-0.333** (0.103)	-0.342** (0.103)	-0.348** (0.103)
Log Revenue	-0.091+ (0.047)	-0.093* (0.047)	-0.089+ (0.047)	-0.090+ (0.047)	-0.089+ (0.047)	-0.094* (0.047)
Log Num of Sites	0.214+ (0.121)	0.212+ (0.122)	0.208+ (0.121)	0.212+ (0.121)	0.211+ (0.121)	0.214+ (0.122)
Log CVE Count	0.209** (0.046)	0.325** (0.085)	0.214** (0.049)	0.196** (0.049)	0.280** (0.083)	0.302** (0.084)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared (Adjusted)	0.165	0.164	0.165	0.165	0.165	0.164
Log Likelihood	-361827	-360185	-361724	-361367	-361823	-360183
Observations	101,721	101,241	101,690	101,577	101,721	101,241

Note: + p < 0.1, * p < .05, ** p < .01, *** p < .001. Values in parentheses are robust standard errors clustered at the firm level. Coefficients reflect percentage point changes as the dependent variable was multiplied by 100 for readability

VENDOR-ORIGINATED VULNERABILITIES – UNDERLYING DYNAMICS

To understand precisely how vendor-originated vulnerabilities contribute to data breach risks, we examined the two critical *subcomponents* of the overall CVSS score separately: (a) Exploitability Score, reflecting the ease of exploiting a vulnerability, and (b) Impact Score, reflecting the potential damage to confidentiality, integrity, and availability (CIA).

While our main analysis confirms aggregated CVSS scores predict breaches, unpacking the CVSS score into these distinct components allows us to clarify the practical significance of each dimension. This understanding enables firms to more effectively prioritize vulnerability remediation efforts- a strategy regarded as more effective than merely increasing remediation capacity (Cyentia Institute & Kenna Security, 2022). Our results show that the aggregate Exploitability Score shows a marginally significant positive effect on data breaches[6]. While this suggests vendor-originated vulnerabilities easier to exploit do somewhat increase breach risk, the relatively weak statistical significance indicates exploitability alone may not be the primary driver. One explanation might be that firms can effectively mitigate easily exploitable

[6] Detailed results are available upon request.

vulnerabilities once identified. Conversely, the Impact Score strongly and positively affects data breach likelihood, emphasizing that attackers prioritize vendor-originated vulnerabilities capable of inflicting significant damage. Thus, vulnerabilities with higher potential impact pose greater threats to organizations and should be prioritized in cybersecurity strategies. Furthermore, we conduct analyses on the interacting impact of these scores and other key actions of vendors and infomediaries. The results reveal that the interaction effects for Exploitability Scores generally align with the overall CVSS results, whereas Impact Score interactions were mostly non-significant except for vulnerability knowledge, underscoring the complexity in managing high-impact vendor-originated vulnerabilities.

VULNERABILITY THRESHOLD IDENTIFICATION

All organizations possess a base degree of vulnerability in their technology infrastructure because of unavoidable vendor reliance and outsourcing requirements. While every CVSS score represents a potential weakness, firms cannot realistically eliminate all vulnerabilities; our goal is therefore to identify an alarm point (i.e., threshold)- a level of vulnerability that remains manageable versus one that signals a "point of no return," beyond which defenses rapidly degrade and breach risk escalates sharply. Organizations face substantial practical constraints in addressing the vulnerabilities within their technology ecosystems due to limited cybersecurity resources and operational complexities. Consequently, identifying a meaningful vulnerability threshold becomes crucial. While the official CVSS document categorizes vulnerabilities with a score of 4.0 as the threshold between "Low" and "Medium/High" severity (NVD, 2025), we let the empirical structure of our data determine a robust and practically meaningful vulnerability threshold. Initially, we split firms at the median CVSS score (4.66) into low and high vulnerability groups, providing a basic distribution-driven reference point. However, to ensure

the validity and practical utility of this threshold, we further employed two independent, nonparametric machine learning approaches[7] explicitly designed for threshold identification.

First, we applied a single-split decision tree (Banerjee & McKeague, 2007; Breiman et al., 2017), a method well suited to finding a "break" in a continuous predictor. In our case, the tree tests every possible vulnerability score as a split, and for each candidate it fits two simple predictions: one average breach rate for firms at or below that score and another for firms above it. It then calculates the mean squared error (MSE) between those group-specific averages and each firm's actual breach rate, and picks the split that makes that MSE as small as possible. This procedure identified a threshold of 4.63 (Figure B1, Appendix B). To further validate our decision-tree result, we implemented an exhaustive sum-of-squared-errors (SSE) grid search (Hansen, 1999; Yeh et al., 2010). This approach also pinpointed a similar threshold at 4.64.

CAUSAL IDENTIFICATION

Although our main results indicate that higher vendor-originated vulnerabilities increases data breach likelihood, potential confounders and unobserved heterogeneity might bias these findings. To address these concerns and strengthen causal inference, we employ a Difference-in-Differences (DiD) approach, leveraging the identified vulnerability threshold as a plausibly exogenous "shock". Specifically, firms crossing this "threshold" serve as the treatment group, while firms that do not cross it form the control group. This design provides a quasi-experimental setting to evaluate whether crossing the vulnerability threshold causally increases breach occurrence.

To ensure robustness, we first used Coarsened Exact Matching (CEM) to pair firms by year and industry, and subsequently applied Propensity Score Matching (PSM) to additionally

[7] The decision-tree threshold approach was implemented using the DecisionTreeRegressor from scikit-learn (Python), and the SSE threshold identification through a custom calculation built using pandas and NumPy (Python).

align firms on characteristics such as employee count, revenue, and IT budget. We applied one-to-one (1:1) matching, yielding a final matched sample of 1,088 firms. Standardized mean differences after matching were below the recommended threshold of 0.1 (Zhang, Kim et al. 2019), confirming good covariate balance. The common support condition was clearly satisfied, as illustrated by overlapping propensity score distributions in density plots (Appendix B, Figures B3 & B4).

We then conducted DiD analysis using the LPM model as shown below.

$$y_{i,t} = \alpha + \beta\,(Treat_i * Post_{it}) + \delta_i + \gamma_t + \varepsilon_{i,t} \quad (4)$$

Where $y_{i,t}$ is the data breach occurrence for firm i in year t, α is the intercept, β is the coefficient of interest, measuring the differential effect of crossing the vulnerability threshold. δ_i captures firm-specific fixed effects, γ_t captures year-specific effects (year fixed effects) and $\varepsilon_{i,t}$ is the error term. $Treat_i$ is a binary indicator equal to 1 if firm i crosses the vendor-originated vulnerability threshold (treatment group) and 0 otherwise (control group). $Post_{it}$ is time-varying indicator (1 in years t after the crossing year for treated firms, 0 before; always 0 for controls). Thus, the interaction term ($Treat_i * Post_{it}$) captures the differential impact on data breach likelihood specifically attributable to crossing the threshold. As seen in Table 4, the average treatment effect is significant (β=1.031, p<.05) & (β=1.032, p<.05) with PSM 1:1 and MDM 1:1 Matching respectively, suggesting that firms crossing critical vendor-originated vulnerability threshold leads to a higher likelihood of data breaches.

While the DiD have been widely used to identify causal effects of staggered events, recent econometrics methodology research shows that it can yield biased results because the estimation of treatment effects includes a problematic comparison between the treatment group and the already treated group (Goodman-Bacon, 2021). To correct the bias, we follow an

approach proposed by Callaway & Sant'Anna (2021) to further validate the LPM results. The analysis shows that the average treatment effect on the treated (ATT) is still significant (ATT=1.886, $p<0.05$), confirming the effect of firms' vendor-originated vulnerabilities on data breaches. [8] Further, a relative time model based on Callaway and Sant'Anna (2021) shows that the treatment effects before the event (Firms crossing the CVSS threshold) do not differ from zero, and several treatment effects after the event differ from zero, thus supporting the parallel trend assumption (Appendix B, Figure B5). Overall, these results indicate that there is likely a causal relationship between a firm's vendor-originated vulnerabilities and the likelihood of data breaches.

Table 4: Difference-in-difference analysis on matched data

Base model			Difference-in-difference					
	Threshold Analysis		Base Model		Interaction			
	Low Vulnerability	High Vulnerability	PSM +CEM 1:1	MDM +CEM 1:1	Patch Availability	Patch Release Time	Vulnerability knowledge	Media Coverage
	(6.1)	(6.2)	(6.3)	(6.4)	(6.5)	(6.6)	(6.8)	(6.7)
Vendor Originated Vulnerability (VoV)	**-0.275 (0.181)**	**0.119* (0.051)**						
Threshold × Post (DiD Effect)			**1.031* (0.502)**	**1.032* (0.494)**	0.302* (0.700)	-4.352* (1.898)	15.546** (3.403)	-1.212* (2.430)
Moderator					9.002* (3.931)	-0.213 (0.304)	2.242** (0.623)	0.498+ (0.268)
DiD*Moderator					**-9.134* (4.150)**	**0.591* (0.284)**	**-2.408** (0.522)**	**0.035 (0.226)**
Log IT Budget	0.458** (0.088)	0.224** (0.058)	-0.105 (0.226)	-0.102 (0.225)	0.370 (0.241)	0.346 (0.240)	0.343 (0.241)	0.359 (0.242)
Log Employees	-0.472* (0.187)	-0.294* (0.129)	0.102 (0.594)	0.104 (0.568)	0.468 (0.568)	0.444 (0.567)	0.454 (0.572)	0.444 (0.575)
Log Revenue	-0.088 (0.079)	-0.051 (0.059)	-0.208 (0.197)	-0.202 (0.197)	-0.245 (0.193)	-0.229 (0.193)	-0.219 (0.192)	-0.229 (0.195)
Log Num of Sites	0.184 (0.185)	0.119 (0.189)	1.211* (0.577)	1.213* (0.574)	1.258* (0.575)	1.301* (0.574)	1.270* (0.575)	1.309* (0.574)

[8] We utilize 'csdid' module in stata.

Log CVE Count	0.313** (0.079)	0.128* (0.063)	0.154 (0.186)	0.145 (0.188)	0.264 (0.217)	0.048 (0.229)	0.371 (0.353)	0.314 (0.244)
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared (Adjusted)	0.125	0.177	0.121	0.122	0.238	0.238	0.240	0.238
Log Likelihood	-172647	-172734	-36109	-36071	-31242	-31231	-31241	-31132
Num of Firms	11737	11888	1,088	1,088	1,075	1,075	1,075	1,074

Note: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$, *** $p < .001$. Values in parentheses are robust standard errors clustered at the firm level. PSM: propensity score matching; MDM: Mahalanobis distance matching; CEM: coarsened exact matching. Coefficients reflect percentage point changes as the dependent variable was multiplied by 100 for readability

CONCLUSION

This study extends Routine Activity Theory by empirically examining how a firm's external vendor-originated vulnerabilities, captured through aggregated CVSS scores, influences data breach likelihood. By mapping vulnerabilities from the NVD to each firm's specific vendors and active systems, we construct a measure of firm-level technology risk that highlights the critical role of vulnerability severity in shaping breach outcomes. Our findings show that higher vendor-originated vulnerabilities significantly increase the likelihood of data breaches, with a distinct threshold effect emerging around a CVSS score of 4.6. Firms exceeding this threshold face sharply elevated breach risks, emphasizing the need for consistent vulnerability monitoring and rapid patch deployment. We also find that patch availability, faster patch deployment, and vulnera bility knowledge, can help mitigate these threats, while media attention to CVE's exacerbates the cyber risk. In addition, metrics related to confidentiality and availability within the CVSS framework pose particularly high risks, underlining the multifaceted nature of vulnerability management.

Our study makes several important contributions to research and practice in cybersecurity risk management. Theoretically, we contribute to the cybersecurity risk management literature by extending Routine Activity Theory to explicitly recognize the critical role external entities

play in shaping organizational vulnerability and guardianship. Additionally, we advance theoretical understanding by empirically examining the interactions among RAT's core dimensions (suitable targets, capable guardianship, and target visibility) in the specific context of vendor-originated vulnerabilities. Empirically, our approach systematically links vendor-originated vulnerabilities to organizational technology ecosystems, enabling deeper analysis of firm-level cybersecurity outcomes.

Beyond its contribution to research, our study also provides practical implications for vulnerability management. First, our findings support the effectiveness of CVSS in assessing cybersecurity risks and demonstrate the values of systemically mapping vendor-originated vulnerabilities to organizational IT systems. Second, our findings offer guidance for security resources allocation by identifying a threshold effect and key dimensions of CVSS. Third, our study highlights the importance of coordination efforts in vulnerability management by empirically examining the role of software vendors and infomediaries in impacting breach risks induced by vendor-originated vulnerabilities.

Our study has certain limitations that warrant acknowledgment. While our research relies on detailed data from the NVD, this source is not entirely free from practical constraints. First, although the NVD's publish date approximates when a vulnerability becomes publicly known, it may not always align perfectly with the exact initial disclosure date. A second limitation involves how we map firms' systems to known vendor-originated vulnerabilities. Future studies might expand this framework by mapping vulnerabilities to particular system versions or specific business units. Third, while our firm-level average CVSS score provides a broad gauge of the severity of vendor-originated vulnerabilities, it may not fully address real-time patch prioritization needs.

REFERENCES

Ablon, L., & Libicki, M. C. (2015). Hackers' bazaar: The markets for cybercrime tools and stolen data. *Defense Counsel Journal, 82*(2), 143–152. https://doi.org/10.7249/RR610

Alkhadra, R., Abuzaid, J., AlShammari, M., & Mohammad, N. (2021, July). SolarWinds hack: In-depth analysis and countermeasures. In *Proceedings of the 12th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICCCNT51525.2021.9579611

Allodi, L. (2017, October). Economic factors of vulnerability trade and exploitation. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17)* (pp. 1483–1499). Association for Computing Machinery. https://doi.org/10.1145/3133956.3133960

Amelia, N. T. (2025). Implications of the WannaCry ransomware attack on personal security: Analysis of human security concepts. *PROIROFONIC, 1*(1), 634–641.

Anu, V., Sultana, K. Z., & Samanthula, B. K. (2020, October). A human-error-based approach to understanding programmer-induced software vulnerabilities. In *Proceedings of the 2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)* (pp. 49–54). IEEE. https://doi.org/10.1109/ISSREW51248.2020.00036

Anwar, A., Khormali, A., Nyang, D., & Mohaisen, A. (2018). Understanding the hidden cost of software vulnerabilities: Measurements and predictions. In R. Beyah, T. Yu, & C. Wang (Eds.), *Security and privacy in communication networks: 14th international conference, SecureComm 2018, Singapore, August 8–10, 2018—Proceedings, Part I* (pp. 377–395). Springer. https://doi.org/10.1007/978-3-030-01701-9_21

Arora, A., Krishnan, R., Telang, R., & Yang, Y. (2010). An empirical analysis of software vendors' patch release behavior: Impact of vulnerability disclosure. *Information Systems Research, 21*(1), 115–132. https://doi.org/10.1287/isre.1080.0226

Arora, A., Telang, R., & Hao, X. (2008). Optimal policy for software vulnerability disclosure. *Management Science, 54*(4), 642–656. https://doi.org/10.1287/mnsc.1070.0771

August, T., Dao, D., & Kim, K. (2019). Market segmentation and software security: Pricing patching rights. *Management Science, 65*(10), 4575–4597. https://doi.org/10.1287/mnsc.2018.3153

August, T., & Tunca, T. I. (2008). Let the pirates patch? An economic analysis of software security patch restrictions. *Information Systems Research, 19*(1), 48–70. https://doi.org/10.1287/isre.1070.0142

Banerjee, M., & McKeague, I. W. (2007). Confidence sets for split points in decision trees. The Annals of Statistics, 35(2), 543–574. https://doi.org/10.1214/009053606000001415

Beardsley, T. (2022, February 10). The hidden harm of silent patches. Rapid7 Blog. https://www.rapid7.com/blog/post/2022/06/06/the-hidden-harm-of-silent-patches/

Benjamin, V., Zhang, B., Nunamaker, J. F., Jr., & Chen, H. (2016). Examining hacker participation length in cybercriminal internet-relay-chat communities. Journal of Management Information Systems, 33(2), 482–510. https://doi.org/10.1080/07421222.2016.1205918

Berghel, H. (2020). The Equifax hack revisited and repurposed. Computer, 53(5), 85–90. https://doi.org/10.1109/MC.2020.2979525

BlueVoyant. (2020). DACH report: Managing cyber risk across the vendor ecosystem. https://www.bluevoyant.com/resources/dach-report-managing-cyber-risk-across-the-extended-vendor-ecosystem

Bomgar. (2016). Vendor vulnerability index 2016. https://assets.beyondtrust.com/assets/documents/Bomgar-Vendor-Vulnerability-Index-2016.pdf

Boyens, J., Smith, A., Bartol, N., Winkler, K., Holbrook, A., & Fallon, M. (2022). Cybersecurity supply chain risk management practices for systems and organizations (NIST Special Publication 800-161 Revision 1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-161r1

Boyson, S. (2014). Cyber supply chain risk management: Revolutionizing the strategic control of critical IT systems. Technovation, 34(7), 342–353. https://doi.org/10.1016/j.technovation.2014.02.001

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. Routledge. https://doi.org/10.1201/9781315139470

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Cavusoglu, H., Cavusoglu, H., & Raghunathan, S. (2007). Efficiency of vulnerability disclosure mechanisms to disseminate vulnerability knowledge. IEEE Transactions on Software Engineering, 33(3), 171–185. https://doi.org/10.1109/TSE.2007.26

Cavusoglu, H., Cavusoglu, H., & Zhang, J. (2008). Security patch management: Share the burden or share the damage? Management Science, 54(4), 657–670. https://doi.org/10.1287/mnsc.1070.0794

Charney, S., & Werner, E. T. (2011). Cyber supply chain risk management: Toward a global vision of transparency and trust [White paper]. Microsoft Corporation. https://icscsi.org/library/Documents/Risk_Management/Microsoft%20-%20Cyber%20Supply%20Chain%20Risk%20Management%20(White%20Paper).pdf

Chen, Y., Boehm, B., & Sheppard, L. (2007). Measuring security investment benefit for off-the-shelf software systems: A stakeholder value–driven approach. In Proceedings of the Workshop on the Economics of Information Security (WEIS). Retrieved from https://econinfosec.org/archive/weis2007/papers/46.pdf

Chingombe, R. (2023). Impact of risk attributes on vendor risk assessment and classification. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4620534

Cho, W., Subramanyam, R., & Xia, M. (2013). Vendors' incentives to invest in software quality in enterprise systems. Decision Support Systems, 56, 27–36. https://doi.org/10.1016/j.dss.2013.04.005

Chowdhury, I., & Zulkernine, M. (2010). Can complexity, coupling, and cohesion metrics be used as early indicators of vulnerabilities? In Proceedings of the 2010 ACM Symposium on Applied Computing (pp. 1963–1969). Association for Computing Machinery. https://doi.org/10.1145/1774088.1774504

Clarke, D., & Tapia-Schythe, K. (2021). Implementing the panel event study. The Stata Journal, 21(4), 853–884. https://doi.org/10.1177/1536867X211063144

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. American Sociological Review, 44(4), 588–608. https://doi.org/10.2307/2094589

Cowles, C. (2011, August). Target relaunches website, dumps Amazon.com. The Cut. https://www.thecut.com/2011/08/target_website-missoni.html
Crews, P., Kuszmaul, W., & Penkov, P. (2018). To disclose or not disclose: The ethics of vulnerability disclosure [Blog post]. Medium. https://medium.com/@ptcrews/to-disclose-or-not-disclose-the-ethics-of-vulnerability-disclosure-aaf09c1ab4b0
Cyentia Institute & Kenna Security (2022) Prioritization to Prediction Volume 8: Measuring and Minimizing Exploitability. https://library.cyentia.com/report/report_008756.html
Cyentia Institute & SecurityScorecard. (2023). Close encounters of the third (and fourth) party kind [Report]. Cyentia Institute & SecurityScorecard. https://library.cyentia.com/report/report_014201.html
D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. Information Systems Research, 20(1), 79–98. https://doi.org/10.1287/isre.1070.0160
D'Arcy, J., Adjerid, I., Angst, C. M., & Glavas, A. (2020). Too good to be true: Firm social performance and the risk of data breach. Information Systems Research, 31(4), 1200–1223. https://doi.org/10.1287/isre.2020.0939
Dahmash, F. N., Durand, R. B., & Watson, J. (2009). The value relevance and reliability of reported goodwill and identifiable intangible assets. The British Accounting Review, 41(2), 120–137. https://doi.org/10.1016/j.bar.2009.03.002
Doan, D. (2006). Commercial off the shelf (COTS) security issues and approaches (Master's thesis, Naval Postgraduate School). Retrieved from https://apps.dtic.mil/sti/tr/pdf/ADA456996.pdf
Dissanayake, N., Jayatilaka, A., Zahedi, M., & Babar, M. A. (2022). Software security patch management—a systematic literature review of challenges, approaches, tools and practices. Information and Software Technology, 144, 106771. https://doi.org/10.1016/j.infsof.2021.106771
Dobrovoljc, A., Trček, D., & Likar, B. (2017). Predicting exploitations of information systems vulnerabilities through attackers' characteristics. IEEE Access, 5, 26063–26075. https://doi.org/10.1109/ACCESS.2017.2769063
Emmitt, J. (2020). The National Vulnerability Database (NVD) explained [Blog post]. Kaseya. https://www.kaseya.com/blog/national-vulnerability-database-nvd/
Esther Kezia, T. (2019). 50% of cyber attacks now use island hopping. ITPro. https://www.itpro.com/security/33946/50-of-cyber-attacks-now-use-island-hopping
Thorpe, E. K. (2019, July 3). 50% of cyber attacks now use island hopping. IT Pro. https://www.itpro.com/security/33946/50-of-cyber-attacks-now-use-island-hopping
Gibbs, J. P. (1968). Crime, punishment, and deterrence. The Southwestern Social Science Quarterly, 515–530.
Goodman, C. (2024). What is the common vulnerability scoring system (CVSS)? Balbix. https://www.balbix.com/insights/understanding-cvss-scores/
Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014
Grover, V., & Kohli, R. (2013). Revealing your hand: Caveats in implementing digital business strategy. MIS Quarterly, 37(2), 655–662. https://doi.org/10.5555/2535658.2535679
Hillman, A. J., Withers, M. C., & Collins, B. J. (2009). Resource dependence theory: A review. Journal of Management, 35(6), 1404–1427. https://doi.org/10.1177/0149206309343469

Holt, T. J., Leukfeldt, E. R., & van de Weijer, S. (2020). An examination of motivation and routine activity theory to account for cyberattacks against Dutch web sites. Criminal Justice and Behavior, 47(4), 487–505. https://doi.org/10.1177/0093854819900322

Hui, K.-L., Kim, S. H., & Wang, Q.-H. (2017). Cybercrime deterrence and international legislation: Evidence from distributed denial of service attacks. MIS Quarterly, 41(2), 497–A11. https://doi.org/10.25300/MISQ/2017/41.2.08

Ilori, O., Nwosu, N. T., & Naiho, H. N. N. (2024). Third-party vendor risks in IT security: A comprehensive audit review and mitigation strategies. World Journal of Advanced Research and Reviews, 22(3), 213–224. https://doi.org/10.30574/wjarr.2024.22.3.1727

Iyengar, R. (2020, December 19). Massive SolarWinds hack has big businesses on high alert. CNN. https://www.cnn.com/2020/12/19/tech/solarwinds-hack-companies/index.html

Jacobs, J., Romanosky, S., Adjerid, I., & Baker, W. (2020). Improving vulnerability remediation through better exploit prediction. Journal of Cybersecurity, 6(1), 1–12. https://doi.org/10.1093/cybsec/tyaa015

Kahn, L. B., & Hershbein, B. (2018). Do recessions accelerate routine‐biased technological change? American Economic Review, 108(7), 1737‐1772. https://www.nber.org/papers/w22762

Kannan, K., & Telang, R. (2005). Market for software vulnerabilities? Think again. Management Science, 51(5), 726–740. https://doi.org/10.1287/mnsc.1040.0357

Keskin, O. F., Caramancion, K. M., Tatar, I., Raza, O., & Tatar, U. (2021). Cyber third‐party risk management: A comparison of non‐intrusive risk scoring reports. Electronics, 10(10), Article 1168. https://doi.org/10.3390/electronics10101168

Kim, K., Mithas, S., & Kimbrough, M. (2017). Information technology investments and firm risk across industries. MIS Quarterly, 41(4), 1347–1368. https://doi.org/10.25300/MISQ/2017/41.4.15

KPMG International. (2022). Model risk management | Third‐party risk management outlook 2022. https://home.kpmg/xx/en/home/insights/2022/01/third-party-risk-management-outlook.html

Krebs, B. (2015, October). At Experian, security attrition amid acquisitions. Krebs on Security. https://krebsonsecurity.com/2015/10/at-experian-security-attrition-amid-acquisitions/

Li, H., & Yoo, S. (2021). Information systems sourcing strategies and organizational cybersecurity breaches. IEEE Transactions on Engineering Management, 68(2), 481–490. https://doi.org/10.1109/TEM.2021.3127485

Liang, H., Srinivas, S., & Xue, Y. (2025). How mergers and acquisitions increase data breaches: A complexity perspective. Management Information Systems Quarterly, 49(1), 211–242. https://doi.org/10.25300/MISQ/2023/17703

Loukas, G. (2015). Cyber‐physical attacks: A growing invisible threat. Butterworth‐Heinemann.

Malatesta, J. T. A., III, & Glover, S. S. (2016). A clear and present danger: Mitigating the data security risk vendors pose to businesses. Sedona Conference Journal, 17, 761.

Martínez, J., & Durán, J. M. (2021). Software supply chain attacks, a threat to global cybersecurity: SolarWinds' case study. International Journal of Safety and Security Engineering, 11(5), 537–545. https://doi.org/10.18280/ijsse.110505

Mell, P., Scarfone, K., & Romanosky, S. (2006). Common Vulnerability Scoring System. IEEE Security & Privacy, 4(6), 85–89. https://doi.org/10.1109/MSP.2006.145

Mell, P., & Spring, J. (2025). Likely Exploited Vulnerabilities: A Proposed Metric for Vulnerability Exploitation Probability (NIST CSWP 41). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.41

Miller, C. (2006). Security considerations in managing COTS software. Cigital, Inc.

Miranda, L., Figueiredo, C., Menasché, D. S., & Kocheturov, A. (2023). Patch or exploit? NVD assisted classification of vulnerability-related GitHub pages. In S. Dolev, E. Gudes, & P. Paillier (Eds.), Cyber Security, Cryptology, and Machine Learning (pp. 511–522). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34671-2_36

Nagle, F. (2019). Open source software and firm productivity. Management Science, 65(3), 1191–1215. https://doi.org/10.1287/mnsc.2017.2977

National Vulnerability Database. (2025, May 1). National Vulnerability Database CVSS scoring. https://nvd.nist.gov/vuln-metrics/cvss

Nelson, N. (2021). How many vendors is too many? EnjoyTech Web. https://medium.com/enjoytechweb/how-many-vendors-is-too-many-6ee99d5c2038

Nir, O. (2023, December 19). The CVE patching dilemma: Why we need to reframe vulnerability management [Blog post]. ArmoSec Blog. https://www.armosec.io/blog/cve-patching/

Ozment, A. (2004). Bug auctions: Vulnerability markets reconsidered. In Proceedings of the Workshop on the Economics of Information Security (WEIS) (pp. 1–23).

Özkan, B. E., & Bulkan, S. (2019, May). Hidden risks to cyberspace security from obsolete COTS software. In Proceedings of the 11th International Conference on Cyber Conflict (CyCon) (Vol. 900, pp. 1–19). IEEE. https://doi.org/10.23919/CYCON.2019.8756990

Parwal, R. (2024). Can performing no testing or minimal testing be a huge risk? [Blog post]. MuukTest Blog. https://muuktest.com/blog/not-testing-software

Pfeffer, J., & Salancik, G. (2015). External control of organizations: Resource dependence perspective. In Organizational Behavior (Vol. 2, pp. 355–370). Routledge. https://ssrn.com/abstract=1496213

Pigni, F., Bartosiak, M., Piccoli, G., & Ives, B. (2018). Targeting Target with a 100-million-dollar data breach. Journal of Information Technology Teaching Cases, 8(1), 9–23. https://doi.org/10.1057/s41266-017-0028-0

Platkin, M. J., Fais, C., & Office of the Attorney General, New Jersey. (2022). NJ to receive roughly $500k from $16M settlements over 2012 and 2015 Experian data breaches [Press release]. New Jersey Office of the Attorney General. https://www.njoag.gov/nj-to-receive-roughly-500k-from-16m-settlements-over-2012-and-2015-experian-data-breaches/

Poireault, K. (2024). DBIR: Vulnerability exploits triple as initial access point for breaches. Infosecurity Magazine. https://www.infosecurity-magazine.com/news/dbir-vulnerability-exploits-triple

Pratt, M. (2022, March 22). 12 steps to building a top-notch vulnerability management program. CSO Online. https://www.csoonline.com/article/3659838

Pugliaresi, L. (2018). US House of Representatives Committee on Oversight and Government Reform [Unpublished manuscript].

Ramasubbu, N., & Kemerer, C. F. (2016). Technical debt and the reliability of enterprise software systems: A competing risks analysis. Management Science, 62(5), 1487–1510. https://doi.org/10.1287/mnsc.2015.2196

Ransbotham, S. (2010). An empirical analysis of exploitation attempts based on vulnerabilities in open source software. In Proceedings of the Workshop on the Economics of Information Security (WEIS).
Ransbotham, S., Sabyaschi, M., & Jon, R. (2012). Are markets for vulnerabilities effective? MIS Quarterly, 36(1), 43–64. https://doi.org/10.2307/41410405
Reyns, B. W., Henson, B., & Fisher, B. S. (2016). Guardians of the cyber galaxy: An empirical and theoretical analysis of the guardianship concept from routine activity theory as it applies to online forms of victimization. Journal of Contemporary Criminal Justice, 32(2), 148-168. https://doi.org/10.1177/1043986215621378
Rightley, C. E., Siegfried-Spellar, K. C., & Rogers, M. K. (2023). Revisiting the Computer Crime Index-Revised (CCI-R): Assessing individual differences and cyber-deviant behavior two decades later. International Journal of Cyber Criminology, 17(1), 102–127.
Roumani, Y. (2021). Patching zero-day vulnerabilities: An empirical analysis. Journal of Cybersecurity, 7(1), tyab023. https://doi.org/10.1093/cybsec/tyab023
Ruohonen, J., & Allodi, L. (2018). A bug bounty perspective on the disclosure of web vulnerabilities. In Proceedings of the 11th International Conference on Cyber Conflict (CyCon U.S.A.) (pp. 1–8). IEEE. https://doi.org/10.48550/arXiv.1805.09850
Salancik, G. R., & Pfeffer, J. (1978). A social information processing approach to job attitudes and task design. Administrative Science Quarterly, 23(2), 224–253. https://doi.org/10.2307/2392563
Schrader, D. (2024). Mitigating the security risks of legacy IT systems. SecurityInfoWatch. https://www.securityinfowatch.com/cybersecurity/article/53081992
Schneier, B., & Vance, A. (2025). "Complexity is the worst enemy of security": Studying cybersecurity through the lens of organizational complexity. MIS Quarterly, 49(1), 205–210. https://doi.org/10.25300/MISQ/2025/49.1.075
Schryen, G. (2009). A comprehensive and comparative analysis of the patching behavior of open source and closed source software vendors. In Proceedings of the 2009 Fifth International Conference on IT Security Incident Management and IT Forensics (pp. 1–8). IEEE. https://doi.org/10.1109/IMF.2009.15
Sen, R., & Borle, S. (2015). Estimating the contextual risk of data breach: An empirical approach. Journal of Management Information Systems, 32(2), 314–341. https://doi.org/10.1080/07421222.2015.1063315
Sen, R., Choobineh, J., & Kumar, S. (2020). Determinants of software vulnerability disclosure timing. Production and Operations Management, 29(11), 2532–2552. https://doi.org/10.1111/poms.13120
Shu, X., Tian, K., Ciambrone, A., & Yao, D. (2017). Breaking the target: An analysis of target data breach and lessons learned [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1701.04940
Spellerberg, I. F., & Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' index. Global Ecology and Biogeography, 12(3), 177–179. https://doi.org/10.1046/j.1466-822X.2003.00015.x
Srinivasan, S., Pitcher, Q., & Goldberg, J. S. (2017). Data breach at Equifax. Harvard Business School Case, 118-031

Tode, C. (2013). Target is 2013 mobile retailer of the year. Retail Dive. https://www.retaildive.com/ex/mobilecommercedaily/target-is-2013-mobile-retailer-of-the-year
Target. (2011). Target annual report 2011. https://www.annualreports.com/HostedData/AnnualReportArchive/t/NYSE_TGT_2011.pdf
Target. (2012). Target annual report 2012. https://www.annualreports.com/HostedData/AnnualReportArchive/t/NYSE_TGT_2012.pdf
Taştan, H., & Gönel, F. (2020). ICT labor, software usage, and productivity: Firm‐level evidence from Turkey. Journal of Productivity Analysis, 53(2), 265–285. https://doi.org/10.1007/s11123-020-00573-x
Thomson Reuters. (2024). The cost of data breaches. Thomson Reuters. https://legal.thomsonreuters.com/blog/the-cost-of-data-breaches/
Trabelsi, S., Plate, H., Abida, A., Aoun, M. M. B., Zouaoui, A., Missaoui, C., Gharbi, S., & Ayari, A. (2015). Mining social networks for software vulnerabilities monitoring. In Proceedings of the 7th International Conference on New Technologies, Mobility and Security (NTMS) (pp. 1–7). IEEE. https://doi.org/10.1109/NTMS.2015.7266506
Van Wegberg, R., Miedema, F., Akyazi, U., Noroozian, A., Klievink, B., & Van Eeten, M. J. G. (2020). Go see a specialist? Predicting cybercrime sales on online anonymous markets from vendor and product characteristics. In Proceedings of The Web Conference 2020 (pp. 816–826). ACM. https://doi.org/10.1145/3366423.3380162
Wang, J., Gupta, M., & Rao, H. R. (2015). Insider threats in a financial institution: Analysis of attack‐proneness of information systems applications. MIS Quarterly, 39(1), 91‐112. https://doi.org/10.25300/MISQ/2015/39.1.05
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT Press.
Wunder, J., Corona, A., Hammer, A., & Benenson, Z. (2024). On NVD users' attitudes, experiences, hopes, and hurdles. Digital Threats: Research and Practice, 5(3), 1–19. https://doi.org/10.1145/3688806
Xue, L., Mithas, S., & Ray, G. (2021). Commitment to IT investment plans: The interplay of real earnings, management, IT decentralization, and corporate governance. MIS Quarterly, 45(1), 193–216. https://doi.org/10.25300/MISQ/2021/14970
Yeboah‐Ofori, A., & Opoku‐Akyea, D. (2019). Mitigating cyber supply chain risks in cyber‐physical systems organizational landscape. In Proceedings of the 2019 International Conference on Cyber Security and Internet of Things (ICSIoT) (pp. 74‐81). IEEE. https://doi.org/10.1109/ICSIoT47925.2019.00020
Yeh, M.-L., Chu, H.-P., Sher, P. J., & Chiu, Y.-C. (2010). R&D intensity, firm performance and the identification of the threshold: Fresh evidence from the panel threshold regression model. Applied Economics, 42(3), 389–401. https://doi.org/10.1080/00036840701604487
Zhang, X., Tsang, A., Yue, W. T., & Chau, M. (2015). The classification of hackers by knowledge exchange behaviors. Information Systems Frontiers, 17(6), 1239–1251. https://doi.org/10.1007/s10796-015-9567-0
Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. Annals of Translational Medicine, 7(1), 14. https://doi.org/10.21037/atm.2018.12.10

Zhao, M., Laszka, A., & Grossklags, J. (2018). Devising effective policies for bug-bounty platforms and security vulnerability discovery. Journal Name, 7, 372–418. https://doi.org/10.5325/jinfopoli.7.2017.0372

Zhu, K. X., & Zhou, Z. Z. (2012). Research note—Lock-in strategy in software competition: Open-source software vs. proprietary software. Information Systems Research, 23(2), 536–545. https://doi.org/10.1287/isre.1110.0358

Zorz, Z. (2019, May 24). How mainstream media coverage affects vulnerability management. Help Net Security. https://www.helpnetsecurity.com/2019/05/24/media-coverage-vulnerability-management/

APPENDIX A

Table A1: Correlation among all variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
(1) Year-on-Year Breach Indicator	1																				
(2) Vendor-Originated Vulnerability	-0.027*	1																			
(3) Patch Average	0.002	-0.035*	1																		
(4) Avg Exploitability Score	-0.037*	0.429*	0.068*	1																	
(5) Avg Impact Score	-0.022*	0.821*	-0.153*	0.193*	1																
(6) Avg Access Vector Score	-0.040*	0.360*	-0.009*	0.844*	0.177*	1															
(7) Avg Access Complexity Score	-0.028*	0.295*	-0.109*	0.282*	0.377*	0.361*	1														
(8) Avg Authentication Score	-0.008*	0.186*	-0.008*	0.195*	0.209*	0.169*	0.197*	1													
(9) Avg Obtain All Privileges Score	-0.005	0.181*	-0.032*	0.085*	0.193*	0.070*	0.188*	0.136*	1												
(10) Avg Confidentiality Impact Score	-0.031*	0.528*	-0.171*	0.258*	0.676*	0.285*	0.519*	0.208*	0.155*	1											
(11) Avg Integrity Impact Score	-0.028*	0.556*	-0.113*	0.227*	0.604*	0.249*	0.405*	0.190*	0.158*	0.490*	1										
(12) Avg Availability Impact Score	-0.022*	0.769*	-0.096*	0.190*	0.909*	0.184*	0.313*	0.216*	0.169*	0.556*	0.544*	1									
(13) Log Avg Patch Release Time	-0.007*	0.028*	-0.184*	0.024*	0.065*	0.053*	0.148*	-0.139*	0.045*	0.253*	0.014*	0.016*	1								
(14) Log Media Articles Count	0.000	0.071*	-0.181*	0.056*	0.177*	-0.053*	-0.016*	-0.018*	-0.009*	0.097*	0.038*	0.137*	0.246*	1							
(15) Log Vulnerability Knowledge	-0.005	0.196*	0.349*	0.169*	0.141*	0.079*	0.132*	0.006	0.056*	0.117*	0.211*	0.144*	0.242*	0.106*	1						
(16) Log Num of Records Breached	0.659*	-0.020*	-0.007*	-0.029*	-0.020*	-0.031*	-0.023*	-0.011*	-0.003	-0.025*	-0.022*	-0.021*	0.003	0	-0.009*	1					
(17) Log IT Budget	0.086*	-0.239*	0.099*	-0.227*	-0.236*	-0.268*	-0.258*	-0.079*	-0.028*	-0.296*	-0.283*	-0.224*	-0.106*	-0.032*	-0.031*	0.061*	1				
(18) Log Employees	0.092*	-0.244*	-0.037*	-0.260*	-0.222*	-0.304*	-0.261*	-0.091*	-0.032*	-0.274*	-0.273*	-0.225*	0.002	0.029*	-0.091*	0.076*	0.794*	1			
(19) Log Revenue	0.077*	-0.234*	0.016*	-0.218*	-0.228*	-0.267*	-0.235*	-0.085*	-0.025*	-0.251*	-0.241*	-0.235*	-0.015*	-0.061*	-0.059*	0.064*	0.778*	0.793*	1		
(20) Log Total Sites	0.074*	-0.138*	0.012*	-0.137*	-0.123*	-0.157*	-0.146*	-0.043*	-0.023*	-0.146*	-0.163*	-0.125*	-0.017*	0.031*	-0.042*	0.058*	0.458*	0.483*	0.323*	1	
(21) Log CVE Count	0.044*	-0.238*	0.368*	-0.252*	-0.280*	-0.275*	-0.320*	-0.216*	-0.096*	-0.363*	-0.239*	-0.249*	0.066*	-0.003	0.676*	0.028*	0.356*	0.287*	0.291*	0.157*	1

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A2: Summary of measures for all variables

Variable	Measure	Source
Year-on-Year Breach Indicator	Binary indicator for data breach occurrence recorded by PRC, for a given firm in a given year.	PRC, IDTRC, Verizon DBIR
Num of Records Breached	Total number of data records lost or impacted for a given firm in a given year.	PRC, IDTRC, Verizon DBIR
Vendor-Originated Vulnerabilities	An aggregate CVSS score reflecting a firm's overall vendor-originated vulnerability level. Computed by aggregating the CVSS scores of all relevant CVEs (vulnerabilities) across all vendors and sites used by the firm for each year. A higher value indicates that the firm was more vulnerable overall due to external vendors.	NIST NVD
Patch Average	Aggregate percentage of the number of patches available calculated across all vulnerabilities (CVEs) across all vendors and sites used by the firm for each year. A higher percentage indicates more patches were available for vendor-originated vulnerabilities.	NIST NVD
Avg Patch Release Time	Aggregate score of the days taken to release patches calculated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Larger values indicate longer patch-release times	NIST NVD
Media Coverage	Aggregate score of media articles counts referencing the vulnerabilities (CVEs) contained in vendors adapted across all sites for each firm per year. A higher value indicates broader or more frequent media coverage, suggesting potentially greater public awareness or severity	Computed using Google Web Search API
Vulnerability Knowledge (References Count)	Aggregate score of the number of references available on the NVD website calculated across all vulnerabilities (CVEs) from all vendors adopted across all sites for each firm per year (e.g., links to vendor advisories, security mailing lists, or other references). A higher value indicates a broader or more diverse set of external information sources associated with those vendor-originated vulnerabilities.	NIST NVD
Exploitability[9] Score	Aggregate CVSS "exploitability" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher values suggest that vulnerabilities are easier to exploit.	NIST NVD
Impact Score	Aggregate CVSS "impact" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher values indicate more damaging vulnerabilities (e.g., data loss, system downtime).	NIST NVD
Access Vector	Aggregate CVSS "access vector" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Reflects whether vulnerabilities can be exploited locally, via adjacent network, or over the internet. Higher scores correspond to more remote or network-based exploitability (e.g., over the internet), while	NIST NVD

[9] Some CVSS sub-dimensions (e.g., Access Vector, Access Complexity) were originally categorical. We converted these categories to numerical values for aggregation and analysis. See Appendix C, Table C1 for the specific coding details.

	lower scores indicate more localized exploitation (e.g., direct physical or on-device access)	
Access Complexity	Aggregate CVSS "access complexity" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher scores mean exploit requires more effort/conditions to succeed.	NIST NVD
Authentication	Aggregate CVSS "authentication" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher scores generally indicate fewer (or no) authentication requirements, making exploitation easier, whereas lower scores suggest more stringent authentication needs.	NIST NVD
Obtain All Privileges	Aggregate CVSS "obtain all privileges" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher scores indicate that vulnerabilities more readily enable attackers to gain full (root/administrator) privileges on a system.	NIST NVD
Confidentiality Impact	Aggregate CVSS "confidentiality" component score, aggregated across all vulnerabilities (CVEs) from all vendors adopted across all sites for each firm per year. Higher values suggest that vendor-originated vulnerabilities have a greater effect on data confidentiality.	NIST NVD
Integrity Impact	Aggregate CVSS "integrity" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher values indicate a greater risk of data tampering or unauthorized modification.	NIST NVD
Availability Impact	Aggregate CVSS "availability" component score, aggregated across all vulnerabilities (CVEs) in all vendors adapted across all sites for each firm per year. Higher values indicate a greater potential for system downtime or service disruptions	NIST NVD
CVE Count	Total count of CVEs across all adapted vendors across all sites for each firm per year.	NIST NVD
IT Budget	Total annual IT spending by the firm.	CiTDB
Employees	Total number of employees working at the firm.	CiTDB
Revenue	Total revenue generated from the firm's principal operations.	CiTDB
Total Sites	Total number of operational sites for the firm.	CiTDB

Table A3: Description of all variables

Variable Name	Mean	Std. Dev.	Min	Max
Year-on-Year Breach Indicator	0.01	0.10	0.00	1.00
Vendor-Originated Vulnerability (1-10)	4.91	1.02	1.20	10.00

Avg Exploitability Score (1-10)	6.27	1.39	1.90	10.00
Avg Impact Score (1-10)	4.03	1.46	2.90	10.00
Avg Access Vector Score (1-3)	2.16	0.57	1.00	3.00
Avg Access Complexity Score (1-5)	1.31	0.45	1.00	4.56
Avg Authentication Score (1-2)	0.01	0.04	0.00	1.00
Avg Obtain All Privileges Score (1-2)	0.00	0.01	0.00	1.00
Avg Confidentiality Impact Score (1-5)	1.16	1.25	0.00	5.00
Avg Integrity Impact Score (1-5)	1.14	1.14	0.00	5.00
Avg Availability Impact Score (1-5)	0.97	1.19	0.00	5.00
Log Patch Average	0.12	0.10	0.00	1.00
Log Avg Patch Release Time	5.92	1.42	0.00	8.57
Log Media Articles Count	9.89	1.15	0.00	13.90
Log Vulnerability Knowledge	6.76	0.85	0.00	9.07
Log Num of Records Breached	0.04	0.56	0.00	19.76
Log IT Budget	13.55	2.10	0.00	24.67
Log Employees	5.44	1.69	0.00	15.28
Log Revenue	3.42	1.94	0.00	13.97
Log Total Sites	1.33	1.03	0.00	9.00
Log CVE Count	6.90	0.98	0.69	9.50

APPENDIX B

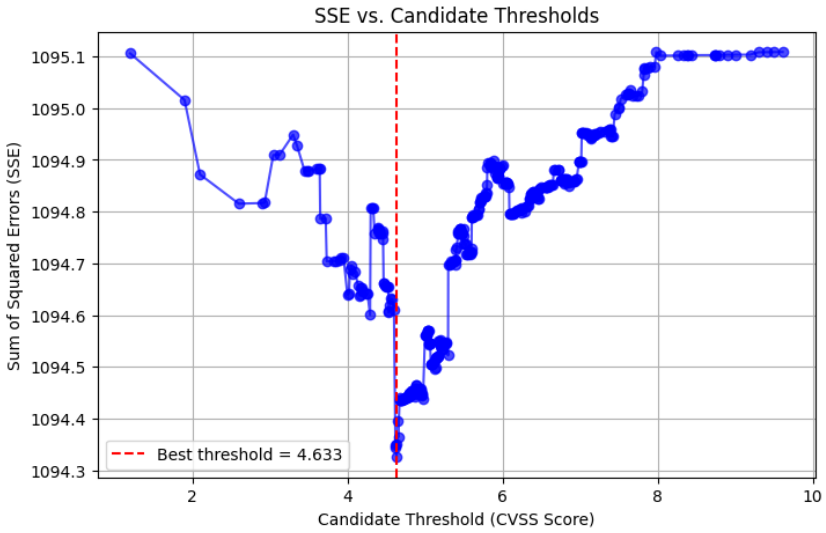


Figure B1: Threshold Identification via SSE

Table B1: Matching statistics PS matching 1:1

	Raw			Matched (ATT)		
Means	Treated	Untreated	StdDif	Treated	Untreated	StdDif
Log IT Budget	15.478	16.881	-0.651	15.479	15.491	-0.006
Log Employees	7.165	8.509	-0.783	7.165	7.178	-0.008
Log Revenue	5.142	6.558	-0.616	5.142	5.162	-0.009
Log Num of Sites	2.296	2.776	-0.309	2.296	2.295	0.001

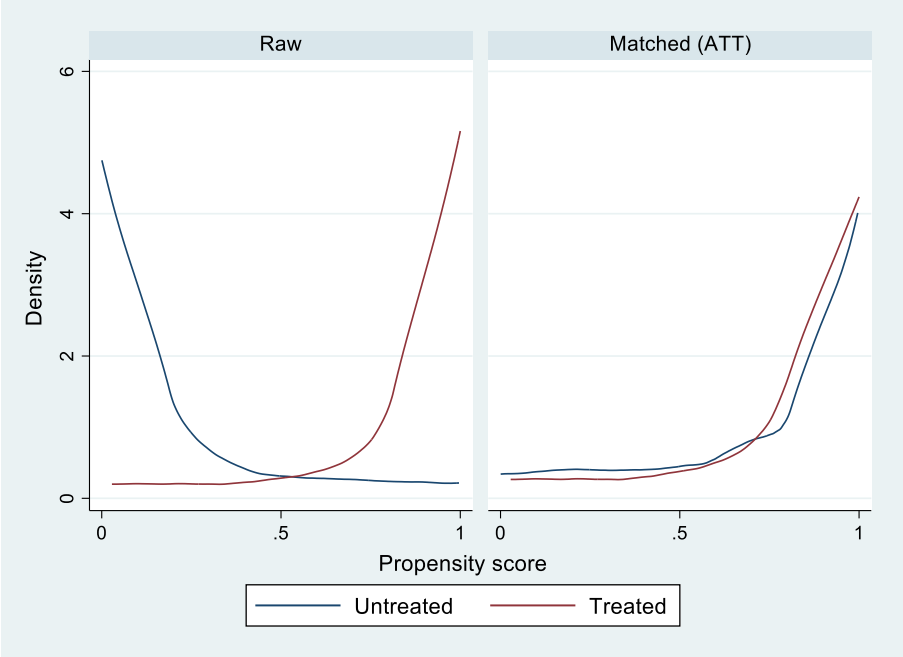


Figure B3: Density plot of propensity scores-Total

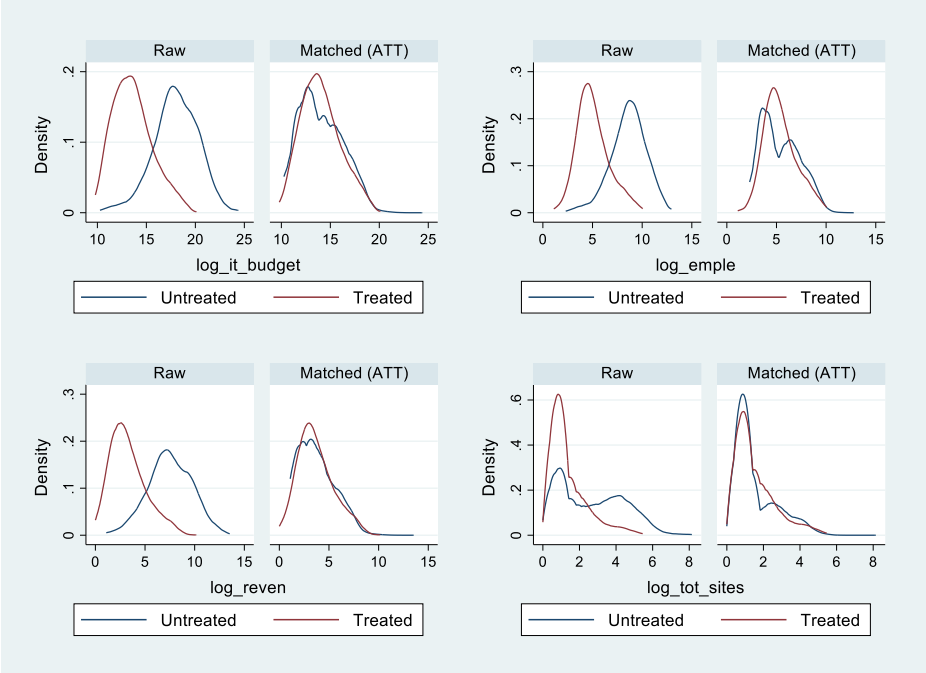


Figure B4: Matching density plot of propensity scores-Individual

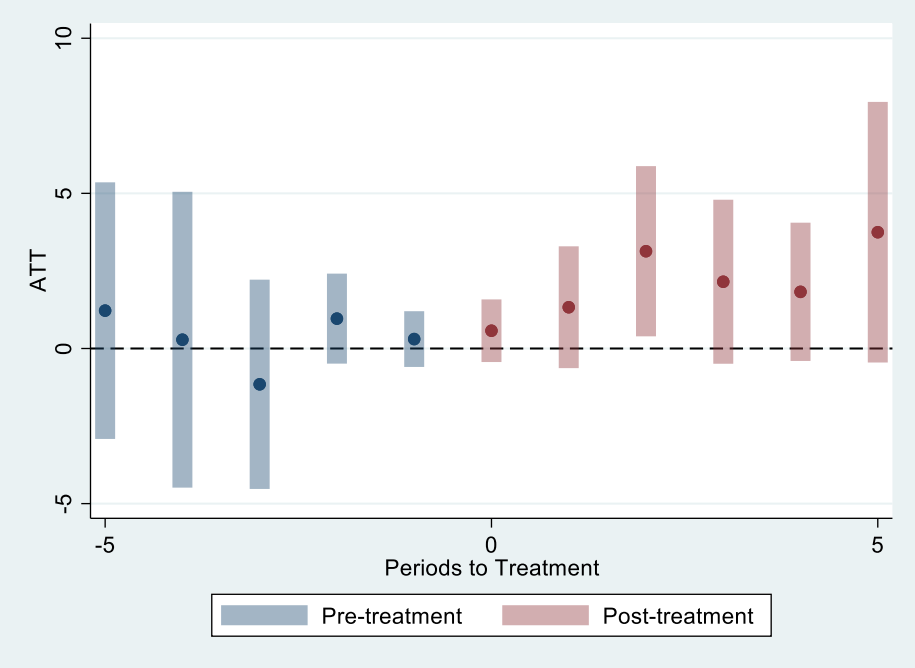


Figure B5: Relative time model for parallel trend analysis

APPENDIX C

Table C1: CVSS Metric Coding Framework

Metric	Categories and Weights
Base Severity	LOW (0), MEDIUM (3), HIGH (5)
Base Score	Range: 0–10
Impact Score	Range: 0–10
Exploitability Score	Range: 0–10
Access Vector	LOCAL (1), ADJACENT NETWORK (3), NETWORK (3)
Access Complexity	LOW (1), MEDIUM (3), HIGH (5)
Authentication	NONE (0), SINGLE (1), MULTIPLE (2)
Confidentiality Impact	NONE (0), PARTIAL (3), COMPLETE (5)
Integrity Impact	NONE (0), PARTIAL (3), COMPLETE (5)
Availability Impact	NONE (0), PARTIAL (3), COMPLETE (5)
Privilege Levels	Obtain All Privilege: TRUE (Low severity); FALSE (Medium/High severity)
	Obtain User Privilege: TRUE (Low severity); FALSE (Medium/High severity)
	Obtain Other Privilege: TRUE (Low severity); FALSE (Medium/High severity)
User Interaction	Required: TRUE (Low severity); FALSE (Medium/High severity)

APPENDIX D

CVSS Background: In 2005, the National Infrastructure Advisory Council (NIAC) introduced the Common Vulnerability Scoring System (CVSS), an open standard allowing quantitative assessment of vulnerabilities' severity (Dobrovoljc et al. 2017). CVSS enables a standardized approach for reporting and evaluating vulnerabilities. Since its introduction, it's been managed by the Forum of Incident Response and Security Teams (FIRST) (Goodman 2024).

The CVSS system generates a score of 0 to 10 based on three main metrics: Base, Temporal, and Environmental. The Base score describes the inherent characteristics of a vulnerability. The temporal score reflects how those characteristics may evolve over time. The Environmental score assesses the vulnerability's impact within a specific environment. A score of 0 indicates a minimal threat, while a score of 10 indicates the highest level of severity. CVSS is widely accepted today and has become an integral part of the automated vulnerability management process (Goodman 2024).

Despite its importance, CVSS has been criticized for several reasons. Some argue that the distribution of Base score is highly bimodal, and many combinations of attributes produce the same final score (Mell et al. 2006). There is also doubt about the accuracy of scores, as the severity of vulnerability observed in real-world scenarios does not always align with the CVSS score (Townsend 2018). Despite these debates, there is a lack of empirical investigation into whether CVSS scores-based characteristics of vulnerabilities are key determinants of cyber risks. To address this gap, our study is the first to systematically examine the relationship between CVSS metrics and organizational cyber risks, focusing on how these scores influence the likelihood and impact of data breaches associated with vendor-originated vulnerabilities.

NVD process

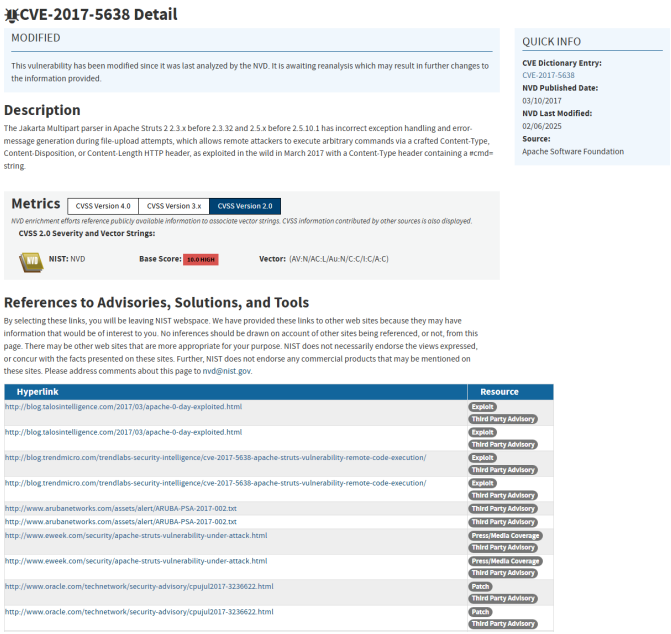
CVE Creation: In the ever-evolving landscape of cybersecurity, where new vulnerabilities emerge daily, the NVD is considered as a centralized, trusted repository that organizations worldwide rely on to identify, assess, and remediate security flaws effectively. When a vulnerability is discovered, researchers or organizations report it to a CVE Numbering Authority (CNA), a global network of vendors and cybersecurity experts. Once validated, the vulnerability is assigned a unique CVE ID, forming the first link in the chain of standardized information. This ID not only identifies the vulnerability but also connects stakeholders to essential details such as the affected software, descriptions of the flaw, and references for further exploration.

CVE Enrichment process: Once a CVE is published, the NVD steps in to enrich the data. NVD adds critical layers of information, including CVSS metrics, which quantify severity; This enrichment transforms raw vulnerability data into actionable intelligence, empowering organizations to prioritize risks, streamline patching efforts, and stay compliant with security standards. The References section particularly is updated during the enrichment phase of the CVE lifecycle. In this process, a dedicated team meticulously reviews the reference materials provided with the CVE record and assigns appropriate reference tags. This effort often includes conducting manual internet searches to identify any additional relevant and publicly available information. Once the enrichment is complete, the results undergo a quality assurance review by a senior team member to ensure accuracy and consistency before being published to the NVD website and data feeds.

CVE Knowledge (References): The "References to Advisories, Solutions, and Tools" section of the NVD plays a critical role in managing vendor-originated vulnerabilities. Enhanced through manual internet research, references are meticulously categorized with labels such as

Patches, Advisory, and others. These references are among the most widely used sections of a CVE entry, with studies showing that 85 percent of practitioners rely on advisory and patch information for their security workflows (Miranda, 2023). The URL links provided in the references section allow organizations to extract additional information about vulnerabilities, including detailed remediation guidance. This makes references an indispensable resource for firms striving to effectively manage vendor-originated vulnerabilities and reduce risks in their environments (Li, 2017; Anwar, 2018).

CVE Patch availability: In the NVD, each CVE entry (See figure D1) the "References to Advisories, Solutions, and Tools" section has critical information regarding the CVE, often tagged as "Patch," "Vendor Advisory," or "Exploit." A "Patch" tag indicates that the software vendor (or another authority) has released a documented fix or update. However, the absence of a "Patch" tag does not always mean no fix exists: sometimes the vendor's only remedy is to upgrade to a newer version, or a patch may be released but not yet reflected in the NVD. In other cases, the vendor may choose not to fix an issue at all (e.g., for end-of-life products). Despite these nuances, the NVD remains a critical reference point, especially for small and medium - sized organizations that lack dedicated security teams and rely on the NVD as their primary source for managing vendor-originated vulnerabilities and obtaining remediation guidance. Recent research by (Wunder et al., 2024) highlights the NVD's central role in practitioners' workflows, showing that it is widely used as an initial checkpoint for identifying and validating vulnerabilities. While not a perfect indicator of patch availability, the NVD offers one of the closest approximations many firms have for managing software exposures in real-world settings.



CVE-2017-5638 Detail

MODIFIED

This vulnerability has been modified since it was last analyzed by the NVD. It is awaiting reanalysis which may result in further changes to the information provided.

Description

The Jakarta Multipart parser in Apache Struts 2 2.3.x before 2.3.32 and 2.5.x before 2.5.10.1 has incorrect exception handling and error-message generation during file-upload attempts, which allows remote attackers to execute arbitrary commands via a crafted Content-Type, Content-Disposition, or Content-Length HTTP header, as exploited in the wild in March 2017 with a Content-Type header containing a #cmd= string.

QUICK INFO

CVE Dictionary Entry: CVE-2017-5638
NVD Published Date: 03/10/2017
NVD Last Modified: 02/06/2025
Source: Apache Software Foundation

Metrics CVSS Version 4.0 | CVSS Version 3.x | CVSS Version 2.0

NVD enrichment efforts reference publicly available information to associate vector strings. CVSS information contributed by other sources is also displayed.

CVSS 2.0 Severity and Vector Strings:

NIST: NVD **Base Score:** 10.0 HIGH **Vector:** (AV:N/AC:L/Au:N/C:C/I:C/A:C)

References to Advisories, Solutions, and Tools

By selecting these links, you will be leaving NIST webspace. We have provided these links to other web sites because they may have information that would be of interest to you. No inferences should be drawn on account of other sites being referenced, or not, from this page. There may be other web sites that are more appropriate for your purpose. NIST does not necessarily endorse the views expressed, or concur with the facts presented on these sites. Further, NIST does not endorse any commercial products that may be mentioned on these sites. Please address comments about this page to nvd@nist.gov.

Hyperlink	Resource
http://blog.talosintelligence.com/2017/03/apache-0-day-exploited.html	Exploit, Third Party Advisory
http://blog.talosintelligence.com/2017/03/apache-0-day-exploited.html	Exploit, Third Party Advisory
http://blog.trendmicro.com/trendlabs-security-intelligence/cve-2017-5638-apache-struts-vulnerability-remote-code-execution/	Exploit, Third Party Advisory
http://blog.trendmicro.com/trendlabs-security-intelligence/cve-2017-5638-apache-struts-vulnerability-remote-code-execution/	Exploit, Third Party Advisory
http://www.arubanetworks.com/assets/alert/ARUBA-PSA-2017-002.txt	Third Party Advisory
http://www.arubanetworks.com/assets/alert/ARUBA-PSA-2017-002.txt	Third Party Advisory
http://www.eweek.com/security/apache-struts-vulnerability-under-attack.html	Press/Media Coverage, Third Party Advisory
http://www.eweek.com/security/apache-struts-vulnerability-under-attack.html	Press/Media Coverage, Third Party Advisory
http://www.oracle.com/technetwork/security-advisory/cpujul2017-3236622.html	Patch, Third Party Advisory
http://www.oracle.com/technetwork/security-advisory/cpujul2017-3236622.html	Patch, Third Party Advisory

Figure D1: NVD CVE entry

CVE Patch Release Time: One of the crucial dimensions of remediating vendor-originated vulnerabilities is the speed with which a patch is documented after a CVE's initial disclosure. While the primary NVD web interface does not display the exact date a patch reference is added, the NVD Change History API provides timestamps for every update done by the CVE enrichment team. By comparing each CVE's publish date to the date on which a "Patch" reference appears in change history, we derive a measure of time taken for patch to be available, i.e the number of days that elapse before a fix is formally recognized in the NVD record. This approach has the same practical limitations as that for patch availability as described in previous section on patch availability. However, despite these nuances, for researchers, this timestamp-based measure is likely one of the most feasible methods to gauge how quickly a disclosed vendor-originated vulnerability transitions to an officially documented remediation.

NVD Importance: The importance of the NVD extends far beyond its technical details. For firms around the globe, it acts as a centralized source of truth, enabling consistent

communication about vendor-originated vulnerabilities and their impact. Security teams use NVD data to correlate vulnerabilities with their assets, prioritize patches for critical systems, and respond swiftly to incidents. Automated tools such as vulnerability scanners and asset management systems often rely on the NVD to validate results and provide precise remediation guidance. This centralized repository ensures that firms, regardless of size or sector, have access to reliable and standardized information, leveling the playing field in the fight against cyber threats associated with vendor-originated vulnerabilities.