Cyber Risks in the Smart Home Ecosystem: Identification, Modeling, and Pricing

FEBRUARY | 2023

SOA Research INSTITUTE

Cyber Risks in the Smart Home Ecosystem: Identification, Modeling, and Pricing

AUTHORS Maochao Xu, PhD

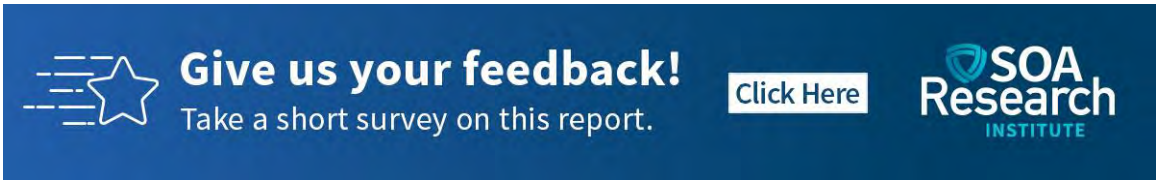
Shouhuai Xu, PhD

SPONSOR General Insurance Research Committee

Actuarial Innovation and Technology Strategic Research Program Steering Committee

Casualty Actuarial Society

Joint Risk Management Section



CONTENTS

Cyber Risks in the Smart Home Ecosystem: Identification, Modeling, and Pricing

Executive Summary

The fourth industrial revolution has emerged (a.k.a., Industry 4.0) in the last few decades and one particular area that it encompasses is the Internet of Things (IoT) and smart manufacturing. The omnipresence of IoT technologies encourages homeowners to transform their homes into smart homes to provide more convenient, comfortable, and secure environments. The global smart home market has exhibited progressive growth in the past few years and has attracted more and more attention from commercial parties. This leads to the emergence of the smart home ecosystem, which consists of various smart devices such as computers, locks, sensors, thermostats, wearables, and various home appliances.

However, many studies in the technological cybersecurity domain have shown that IoT technologies come with their own vulnerabilities, which can be exploited by attackers to compromise smart home devices (e.g., to open garage doors for attackers) and cause damage to homeowners. This new dimension of threats brings an important opportunity for insurance companies to grow their cyber insurance businesses while helping mitigate the risks that are imposed on homeowners by these vulnerabilities.

From an insurer's perspective, state-of-the-art approaches are needed since cyber risks in the smart home ecosystem are little investigated and understood. The present project aims to fill this void. Specifically, we develop a novel and practical quantitative framework for modeling the cyber risks associated with the smart home ecosystem, with an emphasis on cyber insurance pricing. The framework consists of four components: (i) identifying vulnerability-incurred cyber risks; (ii) classifying cyber risks into business lines; (iii) modeling cyber risks; and (iv) determining insurance premiums and coverages. This project lays a sound practical groundwork for launching cyber insurance for the smart home ecosystem, which will provoke further in-depth investigation along this research line.

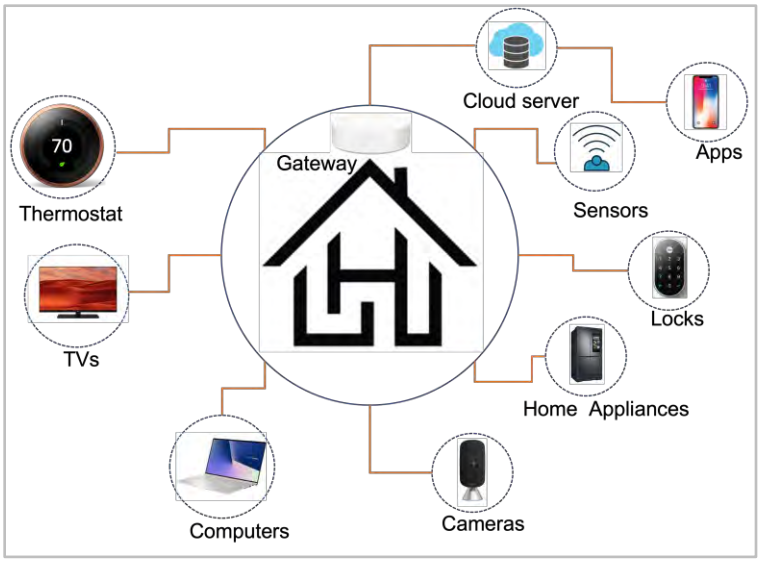
This project enhances the current understanding of cyber risks in the smart home ecosystem from the insurance industry's perspective. In particular, the quantitative framework and pricing strategies developed in this project can be immediately adopted/adapted by actuaries to price the cyber risks for smart homes, a fast-growing insurance market.

Section 1: Introduction

The omnipresence of IoT technology allows average households to transform into smart homes to provide more convenient, comfortable, and secure living environments. According to a report by Zion Market Research, the global smart home market is likely to reach US$137.9 billion by 2026, with a compound growth of slightly more than 1.4% from 2020 to 2026[1]. This explains why the smart home ecosystem is attracting more and more attention from commercial parties such as energy suppliers, third-party software and hardware vendors, and insurance companies.

Figure 1

ILLUSTRATION OF THE COMPLEXITY AND HETEROGENEITY OF THE SMART HOME ECOSYSTEM.



One feature of the smart home ecosystem is that the devices connected to smart homes are highly heterogeneous in their services, ranging from security, healthcare, convenience, and entertainment, to energy efficiency [1]. As illustrated in Figure 1, there are a variety of smart devices in the smart home ecosystem, including thermostats, TVs, computers, cameras, locks, sensors, and smart home appliances. These devices collect and exchange data with each other, allowing homeowners to use smart devices to seamlessly "merge" the physical world and the digital world associated with smart homes into a unified environment. In this ecosystem, the cloud server provides the platform that allows a user to interact with the smart devices at home via mobile applications; each smart device runs independently and communicates via a local mesh network (e.g., Zigbee) with a home gateway acting as the central node. Together, these illustrate the high complexity and heterogeneity of smart devices in smart home ecosystems.

While the smart home ecosystem offers convenience and comfort to our daily lives, smart devices, like many other new technologies, come with many vulnerabilities that can be exploited by attackers to cause damage to homeowners, which leads to an unreliable and insecure ecosystem [2, 3]. For instance, an attacker can wage Denial-of-Service (DoS) attacks to disrupt the use of smart devices; an attacker can wage eavesdropping attacks (e.g., Man-in-the-Middle attack) to steal personal information or corrupt data; an attacker can exploit vulnerabilities in smart cameras for cyber extortions [4]. Indeed, the Avast Smart Home Security Report 2019, which is based on an extensive smart home market analysis, shows that 4.8% of

[1] https://www.zionmarketresearch.com/report/smart-home-market

smart homes have at least one device that is vulnerable to cyber attacks[2]. Notably, among these vulnerabilities, 69.2% are vulnerable because of weak credentials and 31.4% because of software vulnerabilities.

The vulnerabilities such as those mentioned above lead to cyber risks, which represent opportunities for insurance companies to expand their insurance services to help smart homeowners in transferring their cyber risks. The status quo is that the current smart home insurance market mainly provides limited coverage for smart homeowners, which are typically treated as add-ons to standard home insurance policies. For instance, one major insurance company offers smart home coverage, including data breaches, computer and home systems attacks, cyber extortion, and online fraud, but the coverage limits top out at $50,000 with a $500 deductible; another major insurance company offers a personal cyber insurance add-on to its standard home insurance policy by covering some cyber risks in the smart home ecosystem, including cyber-attacks and extortion, with limits up to $15,000. These examples show that the current smart home insurance market is still in the infancy stage, demanding effort and help from the actuarial community.

The insurance coverage demand mentioned above has led to many studies assessing the risks associated with the smart home ecosystem. However, most if not all of these studies are from the perspectives of security, privacy, and users [1, 2, 3, 5, 6]. By contrast, the only study found on quantifying cyber risks in the smart home ecosystem from the insurance perspective is from Zhang et al, which studies cyber risks in a general IoT framework and discusses the associated cyber risks in a small home from a theoretical perspective [7].

This work aims to develop *a practical quantitative framework for identifying, modeling, and pricing cyber risks in a smart home ecosystem*. The present study is unique because of the following characteristics:

- It considers *insurers' perspective*. Cyber risks in the smart home ecosystem are studied from an insurer's perspective, which prompts us to first identify vulnerability-incurred cyber risks in a smart home and then discuss the potential impacts of these cyber risks. The impacts allow us to classify the cyber risks into different lines of insurance risks.
- It makes innovative *technical contributions*. We develop practical probabilistic approaches to pricing cyber risks in the smart home ecosystem. We conduct the following two types of analysis: (i) Conservative analysis, which assumes that any vulnerability in the smart home ecosystem will be exploited by the attackers during a policy period. This analysis is simple but practical but corresponds to the worst-case scenario (i.e., the upper bound for the probabilities that the vulnerabilities will be exploited, respectively). (ii) Bayesian Attack Graph (BAG) analysis [8], which leverages graphical models that represent the information about vulnerabilities and the interactions between the devices in the smart home ecosystem (i.e., the attack paths that can be leveraged to compromise the devices in the smart home ecosystem). Note that (ii) is finer grained than (i) but requires a strong security assessment team.
- It is ready for *real-world adoption*. The present study is more practical and can be easily adopted/adapted by insurance companies to accommodate any relevant scenarios. This is evidenced by our exploration of pricing strategies.

[2] https://cdn2.hubspot.net/hubfs/486579/avast_smart_home_report_feb_2019.pdf

Section 2: A quantitative framework for studying the cyber risks in the smart home ecosystem

For the smart home ecosystem, there are extensive studies in the literature on identifying cyber risks from different perspectives, including risk analysis [8, 9], security [2, 10], privacy [5], industry [6], and scenario-based analysis [3]. However, there is only one prior study on modeling cyber risks in a smart home ecosystem from an insurer's perspective and it provides limited discussion. The present study seeks to fill the void on research from an insurer's perspective. In this study, we discuss a quantitative framework for modeling and pricing cyber risks in a smart home ecosystem. This framework consists of four components: 1) Identifying vulnerability-incurred cyber risks; 2) Classifying cyber risks into business lines; 3) Modeling cyber risks; 4) Determining premiums and coverages.

2.1 IDENTIFYING VULNERABILITY-INCURRED CYBER RISKS

Because of the inherent complexity and heterogeneous structure, the smart home ecosystem has a large attack surface [9, 11]. From an insurer's perspective, it is of utmost importance to identify the risks via a simple but efficient approach. For this purpose, we propose identifying the risk based on vulnerabilities in the smart home network.

A common approach to assessing vulnerability is based primarily on the Common Vulnerability Scoring Systems (CVSS) [12]. The CVSS computes the severity of a vulnerability as a function of its characteristics, and the confidentiality, integrity, and availability impact on an information system. The base score of CVSS is the most commonly used component which produces a score ranging from 0 to 10, where higher scores represent high threat levels. Almost all known vulnerabilities are published on the National Vulnerability Database[3]. The vulnerability information is identified via the Common Vulnerabilities and Exposures (CVE), and each CVE includes the CVE identifier, description, and references discussing the vulnerability. However, it should be noted that the CVSS score does not reflect the probability that a vulnerability will be used to attack a network since only a very small proportion of vulnerabilities are exploited in practice[4]. It is necessary to convert the CVSS into the exploitation probability. Jacobs et al. [13] propose a data-driven framework for assessing the probability that a vulnerability will be exploited within a certain time period after public disclosure. This system is named the Exploit Prediction Scoring System (EPSS)[5]. They show that this system is easy to implement and provides satisfactory estimates of exploitation.

The following steps can be performed to identify the cyber risks in a smart home ecosystem.

1. Scan vulnerabilities. Typically, the vulnerability report generated by vulnerability scanners includes vulnerability dependency details and CVSS scores [14].
2. Create the vulnerability graph. The vulnerability graph abstracts the exploitation relationship among vulnerabilities which can be used for the attack analysis [8]. Based on the vulnerability details, the vulnerability graph can be created.
3. Determine exploitation probabilities. The exploitation probabilities of vulnerabilities can be determined from the vulnerability graph based on vulnerability details and the EPSS.

For illustration, assume that there are three vulnerabilities discovered in a smart home: CVE2021-21736, CVE-2018-3919, and CVE-2022-22667. The vulnerability graph can be created based on the attack scenario: the attacker exploits a vulnerability in a smartphone Operating System (V_1: CVE-2022-22667) over the

[3] https://nvd.nist.gov/vuln-metrics/cvss
[4] https://www.kennasecurity.com/resources/prioritization-to-prediction-report/
[5] https://www.first.org/epss/model

wireless network and compromises the smartphone. This grants the attacker access to the smartphone Operating System, which allows the attacker to pivot into the smart home network to compromise the smart home hub by exploiting the vulnerability (V_3: CVE-2018-3919). Further, the attacker exploits the vulnerability (V_5: CVE-2021-21736) in the smart camera to gain control over it. The vulnerability graph can be represented via a path of two edges, namely

$$V_1 \rightarrow V_3 \rightarrow V_5.$$

The CVSS base scores for those vulnerabilities are 7.8 (V_1), 9.9 (V_3), and 7.2 (V_5). The EPSS probabilities are .02, .3, and .05, respectively.

2.2 CLASSIFYING CYBER RISKS INTO BUSINESS LINES

From the insurer's perspective, the goal is to price the potential loss. It is possible that some vulnerabilities with high CVSS scores and EPSS probabilities only lead to negligible loss. Therefore, it is essential to understand the impacts of cyber-attacks and classify those risks into business lines. Based on the characteristics of the smart home ecosystem, we have identified the following risk categories:

- *Data breach* (L_1). The data breach risk refers to the exposure of personal private user data that can be collected from the smart home ecosystem. The private data includes daily activities, emotions, health conditions, voices, and videos. Data breaches can be caused by the exploitation of vulnerabilities in smart devices or by malware attacks. Data breach coverage pays for the attorney costs, IT professionals, and mitigation of damage caused by the private information leak.
- *Loss of use* (L_2). The loss-of-use risk refers to the data recovery, home applicants repair, and system restoration costs due to cyber attacks such as malware and Denial-of-Service (DoS) attacks. Coverage for this risk pays for those costs and removing a virus or reprogramming of desktops, laptops, smartphones, tablets, Wi-Fi routers, and other internet access points, such as smart home devices and security systems.
- *Ransomware* (L_3). Ransomware refers to that online criminal "lock" (via cryptographic techniques) smart devices such as computers/laptops/tablets, security systems, and/or thermostats to demand ransom. Ransomware coverage pays for the ransom upon the approval of the insurance company or professional assistance to resolve the ransomware event.
- *Cyber extortion* (L_4). Cyber extortion occurs when online criminals threaten to release sensitive personal data, activities, conversations, or videos for financial gain. Cyber extortion coverage reimburses individuals for payments they made under the duress of an extortion threat.
- *Online fraud* (L_5). Online frauds include the direct financial losses caused by cyber attacks, such as stolen account funds, unauthorized use of banking or credit cards, phishing schemes, and other types of fraud. Online fraud coverage pays for the direct financial loss incurred by these attacks.
- *Theft* (L_6). Theft refers to the loss incurred by cyber attacks against security systems. For instance, an attacker unlocks the smart lock and steals wealth from policyholders. The theft coverage reimburses the cost of repurchasing items and property loss.

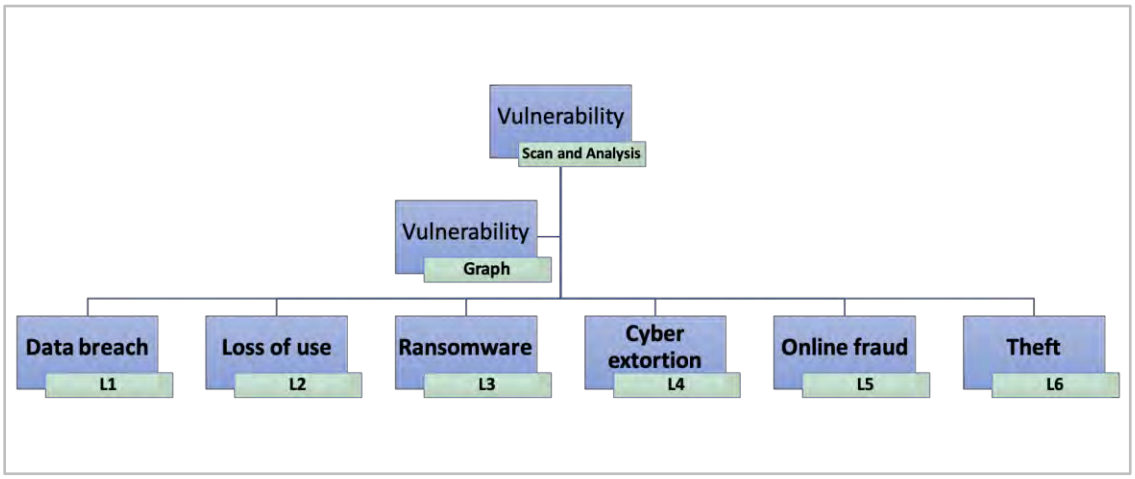
Figure 2 shows those risk categories from the perspectives of insurance companies based on the vulnerability graph.

2.3 MODELING CYBER RISKS IN THE SMART HOME ECOSYSTEM

It is ideal to patch all the vulnerabilities to mitigate risks. However, it is not always possible in practice (e.g., due to lack of manpower). Thus, assessing the risks of the vulnerability network is essential to optimize

resources and the effort required to protect the network. Analyzing the network risks in isolation offers a limited perspective on network security, given the complex interdependencies between vulnerabilities.

Figure 2
VARIOUS INSURANCE RISKS IN A SMART HOME ECOSYSTEM.



Bayesian Attack Graphs (BAGs) [8, 15] provide a powerful framework to represent prior knowledge about vulnerabilities and network connectivity, depicting the paths of an attacker through the system by exploiting successive vulnerabilities. At a high level, BAGs are graphical models representing information about network vulnerabilities and their interactions, and different paths that an attacker can use to compromise a given objective. Along each attack path, vulnerabilities are exploited in sequence, meaning that each successful exploitation allows the attacker to acquire more privileges toward the target.

Figure 3
ILLUSTRATION OF ATTACK PATHS IN THE SMART HOME ECOSYSTEM.

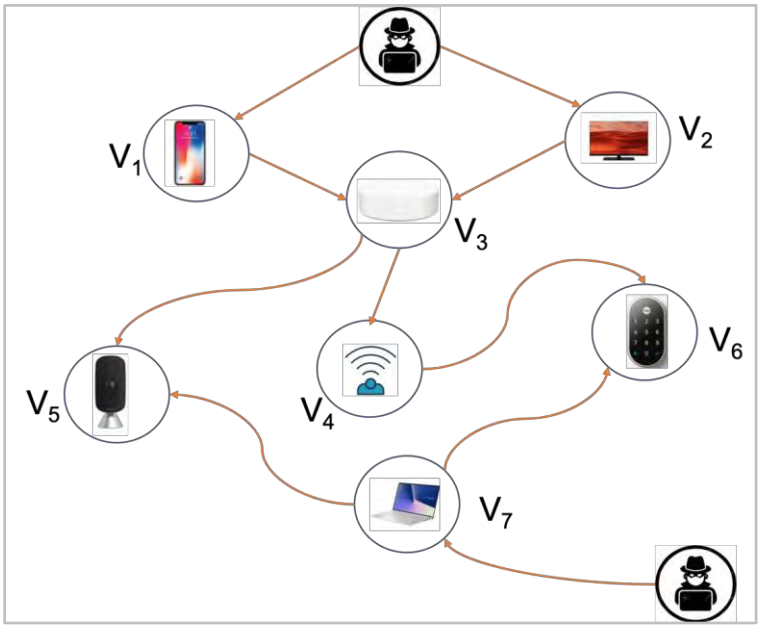


Figure 3 illustrates various attack scenarios, three of which are described in more detail.

- Attack scenario 1: This attack scenario is described in Section 2.1. The attack path is $V_1 \rightarrow V_3 \rightarrow V_5$. This attack can cause risks, including data breach (e.g., data stored in the smart hub) and cyber extortion (e.g., private activities captured by the camera).
- Attack scenario 2: The attacker exploits a vulnerability in the smartphone Operating System (V_1: CVE-2022-22667) over the wireless network, and gains access to the smartphone. The attacker uses this access to compromise the smart home hub in the network by exploiting the vulnerability (V_3: CVE-2018-3919). Further, a smart sensor vulnerability (V_4: CVE-2021-39277) can be exploited, and a smart lock vulnerability (V_6: CVE-2019-12944) can be further exploited to compromise the lock. This attack path can be represented as a path of three edges, namely $V_1 \rightarrow V_3 \rightarrow V_4 \rightarrow V_6$. This attack can cause risks, including data breach and property theft (e.g., unlocking the door).
- Attack scenario 3: The attacker exploits a vulnerability in the laptop (V_7: CVE-2017-8759) over the wireless network, and remotely executes some malicious code. The attacker uses this access to exploit the vulnerability (V_5: CVE-2021-21736) in the smart camera to gain control over it. This attack can cause risks, including data breach and cyber extortion.

Let V_j denote the state that vulnerability j is exploited (i.e., the corresponding smart home device is compromised or not), where $V_j = 1$ means vulnerability j is successfully exploited and $V_j = 0$ the opposite, $j = 1,2,\dots,J$. Let $\mathrm{L_k}$ denote the total loss associated with type k risk, which is determined by successfully exploiting vulnerabilities, $k = 1,\dots,M$. Note that the joint probability of vulnerabilities can be represented via BAG [8]

$$P(V_1 = v_1,\dots,V_N = v_N) = \prod_{i=1}^{N} P(V_i = v_i|\mathbf{pa}_i)\,, \qquad v_i = 1,0, \tag{1}$$

where N represents the total number of vulnerabilities, $\mathbf{pa}_i$ is the parent node sets of vulnerability node i (e.g., vulnerability node V_5 in Figure 3 has the parent node set $\mathbf{pa}_5 = \{V_3, V_7\}$), and

$$v_i = \begin{cases} 1, & \text{the corresponding device is compromised,} \\ 0, & \text{the corresponding device is not compromised.} \end{cases}$$

Note that for type k loss, we have

$$\begin{aligned} P\,(\mathrm{L}_k \le t_k) &= \sum\nolimits_V P\,(\mathrm{L}_k \le t|\mathbf{V} = \mathbf{v})P(\mathbf{V} = \mathbf{v}) \\ &= \sum\nolimits_V P\,(\mathrm{L}_k \le t|\mathbf{V} = \mathbf{v}) \prod_{i=1}^{N} P(V_i = v_i|\mathbf{pa}_i)\,, \end{aligned} \tag{2}$$

where $k = 1,\dots,M$, and $\mathbf{v} = (v_1,\dots,v_N)$. The distribution of total loss (TL) can be represented by

$$\begin{aligned} P\,(\mathrm{TL} \le t) &= P\left(\sum_{k=1}^{M} \mathrm{L}_k \le t\right) \\ &= \sum_{\mathbf{v}} P(\textstyle\sum_{k=1}^{M} L_k \le t\,|\mathbf{V} = \mathbf{v}) \displaystyle\prod_{i=1}^{N} P(V_i = v_i|\mathbf{pa}_i)\,. \end{aligned} \tag{3}$$

From the practical perspective, we study a special case by assuming that any vulnerability that can be exploited will be exploited independently by the attacker. It corresponds to the worst-case scenario, and the joint probability of vulnerabilities can be represented as

$$P(V_1 = v_1, \dots, V_N = v_N) = \prod_{i=1}^{N} P(V_i = v_i | \mathbf{pa}_i), \qquad v_i = 1,0,$$

This approach is straightforward to implement in practice. However, it overestimates the risk since every vulnerability is assumed to be exploitable and hence, can be considered as a conservative case. For ease of reference, we name this **Approach 1**, and the BAG approach **Approach 2**. Please note that the key difference between Approach 1 and Approach 2 is that Approach 1 does not consider exploitable relationships among vulnerabilities while Approach 2 does.

We also pay special attention to the dependence among large risks, i.e., tail dependence. The relevant concept is recalled as follows: Let F_X and F_Y be the distribution of random variable X and Y , then the tail dependence [16, 17] index $\chi(u)$ of X and Y at the level u is

$$\chi(u) = P\big((Y > F_Y^{-1}(u) | X > F_X^{-1}(u)\big).$$

The sample version of $\chi(u)$, based on data $\{(x_i, y_i) | i = 1, \dots, n\}$ is

$$\chi_n(u) = \frac{1}{n(1-u)} \sum_{i \le n} I\{x_i > x_{[nu]:n}, y_i > y_{[nu]:n}\},$$

where $x_{[nu]:n}$ and $y_{[nu]:n}$ represent the $[nu]$th order statistics, respectively.

In the following section, we discuss two scenarios.

2.3.1 SCENARIO WITH FEWER VULNERABILITIES

Assume that a smart home has only three vulnerabilities V_1,V_3 and V_5 as described in attack scenario 1. Table 1 displays the different types of loss caused by successfully exploited vulnerabilities.

Table 1

'x'MEANS VULNERABILITY V_i CAUSES LOSS $L_j, i = 1,3,5, j = 1, \dots, 5$; 'Score' REPRESENTS THE CVSS SCORE; 'Prob.' REPRESENTS THE EPSS PROBABILITY.

CVSS	L_1	L_2	L_3	L_4	L_5	L_6	Score	Prob.
CVE-2022-22667 (V_1)	x				x		7.8	.02
CVE-2018-3919 (V_3)	x	x					9.9	.30
CVE-2021-21736 (V_5)		x		x			7.2	.05

In the following, we discuss different loss distributions.

Gamma loss distributions Let $X_{i,j}$ be the type j loss caused by $V_i, i = 1,3,5; j = 1, \dots, 5$. Assume that $X_{i,j}$ follows a Gamma distribution with scale parameter $\beta = 1$, and shape parameter α_i, and

$$(\alpha_1, \alpha_3, \alpha_5) = (5,1,2).$$

This means that the different vulnerabilities result in various losses. For example, the data breach caused by V_1 is severer than that by V_3. It is further assumed that given the exploration status of vulnerabilities, the different types of losses are independent. We study the loss distribution under the two different approaches in the following.

Approach 1. The probabilities of exploitation are independent, i.e.,

$$P(V_1 = v_1, V_3 = v_3, V_5 = v_5) = P(V_1 = v_1)P(V_3 = v_3)P(V_5 = v_5).$$

The probabilities of exploration scenarios can be easily computed in Table 2. For example,

$$P(V_1 = 1, V_3 = 0, V_5 = 0) = .02(1 - .3)(1 - .05) = .0133.$$

Table 2 displays different loss scenarios, where $X_{i,j} = \mathrm{L}_j(\mathbf{s}_i)$ represents the type j loss caused by scenario $\mathbf{s}_i, i = 1, \ldots, 8$. Note that for scenarios $\mathbf{s}_i, i = 1, \ldots, 8$, the loss is computed based on all the vulnerabilities in this scenario, i.e.

$$X_{5,j} \overset{d}{=} \mathrm{L}_j(\mathbf{s}_2) + \mathrm{L}_j(\mathbf{s}_3), \quad X_{6,j} \overset{d}{=} \mathrm{L}_j(\mathbf{s}_3) + \mathrm{L}_j(\mathbf{s}_4),$$

$$X_{7,j} \overset{d}{=} \mathrm{L}_j(\mathbf{s}_2) + \mathrm{L}_j(\mathbf{s}_4), \quad X_{8,j} \overset{d}{=} \mathrm{L}_j(\mathbf{s}_2) + \mathrm{L}_j(\mathbf{s}_3) + \mathrm{L}_j(\mathbf{s}_4),$$

where '$\overset{d}{=}$' represents both sides are equal in distribution. For example, we have

$$X_{5,1} \overset{d}{=} X_{2,1} + X_{3,1}$$

$$X_{5,2} \overset{d}{=} X_{3,2}, \quad X_{5,5} \overset{d}{=} X_{2,5}.$$

Since $X_{i,j}$'s are independent Gamma random variables, $i = 1, \ldots, 4$, and $j = 1, \ldots, 5$, $X_{k,j}$'s also Gamma distributions, $k = 5, \ldots, 8$. The distributions of losses (DLs) are shown in Table 2.

Table 2

VARIOUS LOSS SCENARIOS UNDER DIFFERENT SCENARIOS OF (V_1, V_3, V_5). 'Prob.' REPRESENTS THE SCENARIO PROBABILITY, AND 'DL' PRESENTS THE DISTRIBUTION OF LOSS.

Scenario	(V_1, V_3, V_5)	Prob.	L_1	L_2	L_4	L_5	DL_1	DL_2	DL_4	DL_5
1	(0,0,0)	.6517	0	0	0	0	0	0	0	0
2	(1,0,0)	.0133	$X_{2,1}$	0	0	$X_{2,5}$	$\Gamma(5,1)$	0	0	$\Gamma(5,1)$
3	(0,1,0)	.2793	$X_{3,1}$	$X_{3,2}$	0	0	$\Gamma(1,1)$	$\Gamma(1,1)$	0	0
4	(0,0,1)	.0343	0	$X_{4,2}$	$X_{4,4}$	0	0	$\Gamma(2,1)$	$\Gamma(2,1)$	0
5	(1,1,0)	.0057	$X_{5,1}$	$X_{5,2}$	0	$X_{5,5}$	$\Gamma(6,1)$	$\Gamma(1,1)$	0	$\Gamma(5,1)$
6	(0,1,1)	.0147	$X_{6,1}$	$X_{6,2}$	$X_{6,4}$	0	$\Gamma(1,1)$	$\Gamma(3,1)$	$\Gamma(2,1)$	0
7	(1,0,1)	.0007	$X_{7,1}$	$X_{7,2}$	$X_{7,4}$	$X_{7,5}$	$\Gamma(5,1)$	$\Gamma(2,1)$	$\Gamma(2,1)$	$\Gamma(5,1)$
8	(1,1,1)	.0003	$X_{8,1}$	$X_{8,2}$	$X_{8,4}$	$X_{8,5}$	$\Gamma(6,1)$	$\Gamma(3,1)$	$\Gamma(2,1)$	$\Gamma(5,1)$

Next, we compute the distribution for each business line. Since we assume that the scenarios are independent, it holds that

$$P(\mathrm{L}_j \le t_j) = \sum_{\mathbf{v}} (\mathrm{L}_j \le t_j | \mathbf{V} = \mathbf{v}) P(\mathbf{v}) \ . \tag{4}$$

For instance, we have

$$P(\mathrm{L}_1 \le t_1) = .6517 + .0133 \cdot F_5(t_1) + .2793 \cdot F_1(t_1) + .0343 + .0057 \cdot F_6(t_1) + .0147 \cdot F_1(t_1) + .0007F_5(t_1) + .0003F_6(t_1) = .686 + .014 \cdot F_5(t_1) + .294 \cdot F_1(t_1) + .006 \cdot F_6(t_1),$$

where $F_\alpha(\cdot) \sim \Gamma(\alpha, 1)$ represents a Gamma distribution with shape parameter α and scale parameter 1. Similarly, it holds that

$$P(\mathrm{L}_2 \le t_2) = .665 + .285F_1(t_2) + .035F_2(t_2) + .015F_3(t_2),$$
$$P(\mathrm{L}_4 \le t_4) = .95 + .05F_2(t_4),$$
$$P(\mathrm{L}_5 \le t_5) = .98 + .02F_5(t_5).$$

For the total loss $\mathrm{TL} = \mathrm{L}_1 + \mathrm{L}_2 + \mathrm{L}_4 + \mathrm{L}_5$, we have

$$\begin{aligned} &P(L_1 + L_2 + L_4 + L_5 \le t) \\ &= \sum_{\mathbf{V}} \mathrm{L}_1 + \mathrm{L}_2 + \mathrm{L}_4 + \mathrm{L}_5 \le t | \mathbf{V} = \mathbf{v}) \prod_{\mathbf{V}} P(\mathbf{v}) \\ &= .6517 + .0133 \cdot F_{10}(t) + .2793 \cdot F_2(t) + .0343 \cdot F_4(t) \\ &\quad + .0057 \cdot F_{12}(t) + .0147 \cdot F_6(t) + .0007F_{14}(t) + .0003F_{16}(t). \end{aligned} \tag{5}$$

It should be noted that although we assume that given the attack exploration scenario, $(\mathrm{L}_i|\mathbf{v})$'s are independent, L_i's are not independent because of the same vulnerabilities. It is further confirmed in the following discussion.

- Simulation. We perform a simulation study to validate the loss distributions. The simulation algorithm is presented in Algorithm 1. In our case, we have vulnerabilities (V_1, V_3, V_5), and the exploration probabilities
$$(e_1, e_3, e_5) = (p_1, p_3, p_5) = (.02, .3, .05).$$
The number of simulations is R = 100,000.

Algorithm 1: Loss simulation in a smart home ecosystem.

INPUT: Vulnerabilities $(V_1, \dots, V_N)$; EPSS probabilities $(p_1, \dots, p_N)$; Loss distributions of $X_{i,j}$s; Number of simulations R.

OUTPUT: Loss distributions.

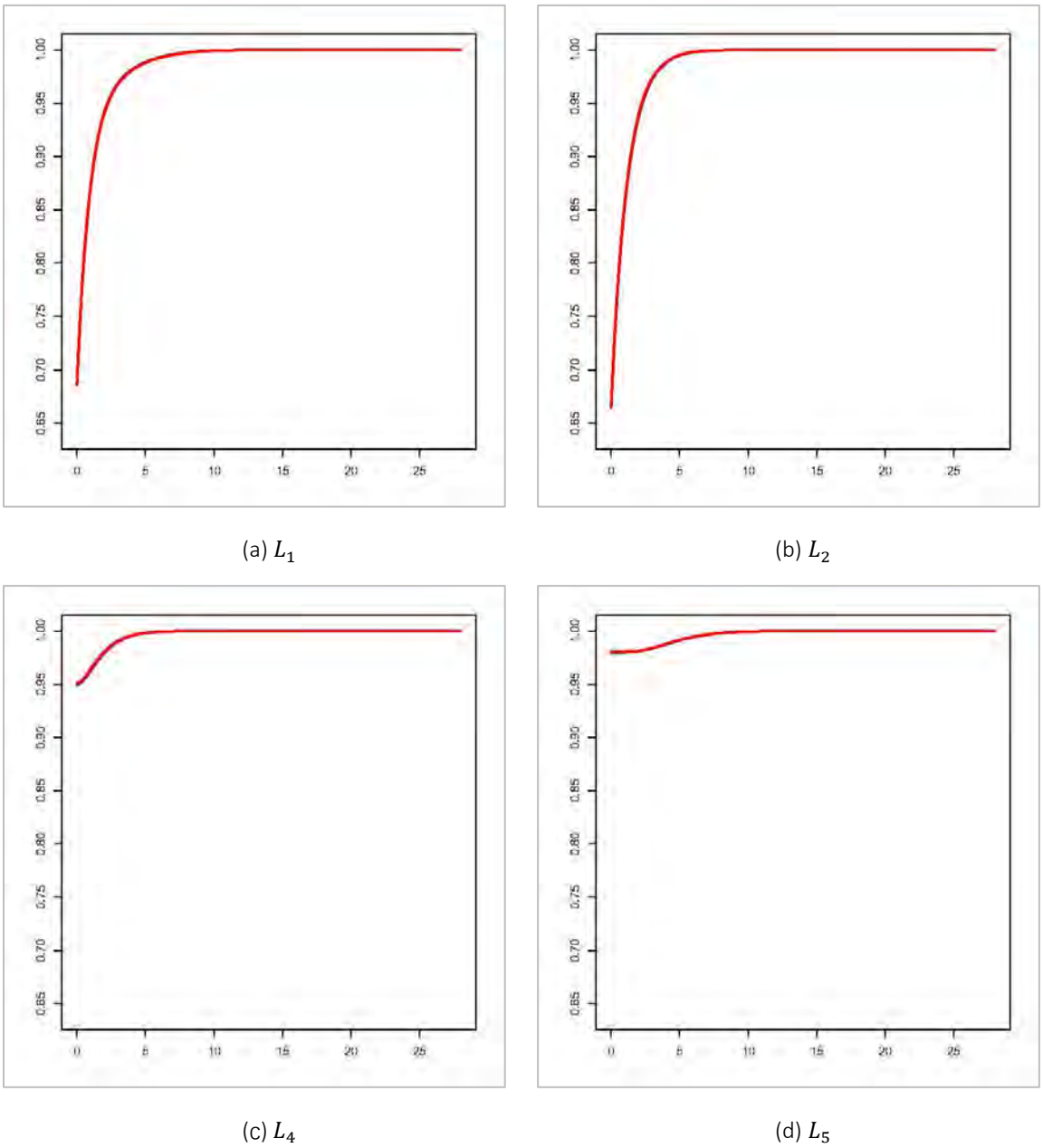
1 Draw the BAG based on vulnerabilities $(V_1, \dots, V_N)$;
2 Determine the exploitation probabilities $(e_1, \dots, e_N)$ based on EPSS probabilities $(p_1, \dots, p_N)$ and the BAG;
3 for $k = 1$ to R do
4 Generate Bernoulli vector $(v_1, \dots, v_N)$ based on $(e_1, \dots, e_N)$;
5 Determine business lines that are impacted by each vulnerability V_i with $v_i = 1$;
6 Randomly generate losses $x_{i,j}$'s from their corresponding distributions for those impacted business lines;
7 end
8 Record losses for each business $L_j, j = 1, \dots, M$.

Figures 4(a), 4(b),4(c), and 4(d) show the distributions of losses based on Eq.(4) (black color) and the simulation (red color) for $\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_4$, and L_5, respectively. We observe that the simulated results coincide with theoretical results very well.

Figure 5(a) displays the theoretical distribution of the total loss based on Eq.(5) (black color) and the simulated distribution of the total loss (red color). Again, we observe that they match very well. Figure 5(b) shows all the losses together for comparison. It is seen that L_4 and L_5 are more likely to have small values since they are impacted by V_5 and V_1, respectively, which have low exploit probabilities. Because V_1 is more difficult to exploit, L_5 is more likely to have a small value than L_4. L_1 and L_2 are more likely to be large values because of the high exploit probability of V_3. Table 3 displays the summary statistics of different types of losses and the total loss based on simulation. It is seen that the 95th percentiles of L_4 and L_5 are both 0s which fits the previous conclusion.

Figure 4

SIMULATED DISTRIBUTIONS (COLORED) AND THEORETICAL DISTRIBUTIONS (BLACK).



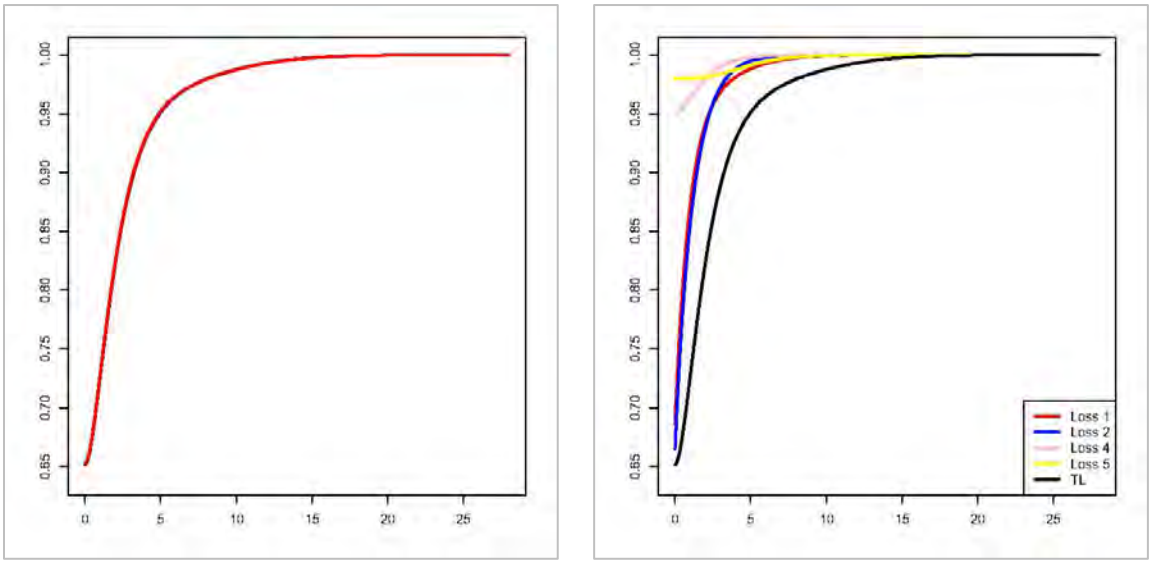
(a) L_1 (b) L_2

(c) L_4 (d) L_5

- **Dependence among losses**. It is interesting to study the dependence among risks of different business lines. Although it is assumed that the exploitations are independent, the dependence can be caused by the same vulnerability. Table 4 displays Kendall's tau correlation coefficients between the losses of different business lines based on simulation. It is seen that L_1 and L_2 have a large correlation .691. This can be explained by vulnerability V_3 that affects both business lines and has a large exploit probability .3. L_4 and L_5 are caused by different vulnerabilities; therefore, the correlation is 0. The other small correlations in the table can be explained similarly. In terms of the correlation between each business line and total loss, it is observed that L_1 and L_2 have large correlations. This can be explained by the fact that both business lines have large risks.

Figure 5

THE SIMULATED DISTRIBUTION (COLORED) OF TOTAL LOSS AND THE THEORETICAL DISTRIBUTION (BLACK) OF TOTAL LOSS.



(a) Distributions of total loss

(b) Simulated distributions of all losses

Table 3

SUMMARY STATISTICS OF DIFFERENT TYPES OF LOSSES AND THE TOTAL LOSS BASED ON SIMULATION, WHERE 'SD' REPRESENTS STANDARD DEVIATION.

	Min	Q_{25}	Median	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
L_1	.000	.000	.000	.241	1.293	2.261	5.429	6.936	9.650	16.624	.401	1.061
L_2	.000	.000	.000	.338	1.429	2.245	4.234	5.000	7.042	13.168	.397	.892
L_4	.000	.000	.000	.000	.000	.000	2.926	3.855	5.792	11.753	.097	.529
L_5	.000	.000	.000	.000	.000	.000	4.592	6.213	9.304	15.595	.099	.768
TL	.000	.000	.000	1.208	3.205	4.900	1.844	13.191	17.473	27.579	.995	2.142

Table 4 also shows the tail indexes χ_us with $u = .999$ between different losses. It is interesting to see that L_1 and L_5 have the largest tail dependence .060 although L_1 and L_2 have the largest Kendall's tau correlation. This can be explained by the fact that V_1 affects both L_1 and L_5, which

can result in large losses. This can also be seen in Table 3 that both L_1 and L_5 have large values of high quantiles (namely, $Q_{99.9}$).
We draw the following insight based on the previous discussion:

Insight 1. *It is practical to assume that any vulnerability that can be exploited will be exploited independently by the attacker in a smart home system. However, there exists dependence among different business risks if a common vulnerability exists, and the tail dependence between business lines can also exist.*

Table 4
KENDALL'S TAU CORRELATION COEFFICIENTS AND TAIL INDEXES $\chi(.999)$S BETWEEN DIFFERENT LOSSES.

	Kendall's tau				Tail index $\chi(.999)$			
	L_2	L_4	L_5	TL	L_2	L_4	L_5	TL
L_1	.691	-.007	.266	.817	.000	.000	.060	.420
L_2		.352	.002	.861		.030	.000	.040
L_4			.000	.352			.000	.020
L_5				.258				.380

Approach 2 In this case, the exploitations of vulnerabilities are not independent. Based on the BAG: $V_1 \rightarrow V_3 \rightarrow V_5$, the joint distribution of (V_1, V_3, V_5) can be represented as

$$P(V_1, V_3, V_5) = P(V_1)P(V_3|V_1)P(V_5|V_3).$$

For comparison, we assume that $P(V_1 = 1) = .02$, $P(V_3 = 1|V_1 = 1) = .3$, and $P(V_5 = 1|V_3 = 1) = .05$. The possible scenarios are less compared to Table 2, i.e., 1, 2, 5, and 8. Table 5 shows the probabilities of different scenarios, and also the loss distributions.

Table 5
VARIOUS LOSS SCENARIOS UNDER DIFFERENT SCENARIOS OF (V_1, V_3, V_5). 'Prob.' REPRESENTS THE SCENARIO PROBABILITY, AND 'DL' PRESENTS THE DISTRIBUTION OF LOSS.

Scenario	(V_1,V_3,V_5)	Prob.	L_1	L_2	L_4	L_5	DL_1	DL_2	DL_4	DL_5
1	(0,0,0)	.980	0	0	0	0	0	0	0	0
2	(1,0,0)	.014	$X_{2,1}$	0	0	$X_{2,5}$	$\Gamma(5,1)$	0	0	$\Gamma(5,1)$
5	(1,1,0)	.0057	$X_{5,1}$	$X_{5,2}$	0	$X_{5,5}$	$\Gamma(6,1)$	$\Gamma(1,1)$	0	$\Gamma(5,1)$
8	(1,1,1)	.0003	$X_{8,1}$	$X_{8,2}$	$X_{8,4}$	$X_{8,5}$	$\Gamma(6,1)$	$\Gamma(3,1)$	$\Gamma(2,1)$	$\Gamma(5,1)$

We compute the distribution for each business line based on Eq.(2). For example, we have

$$\begin{aligned} & P(\mathrm{L}_1 \leq t_1) \\ &= \sum_{\mathbf{v}} P(\mathrm{L}_1 \leq t_1 | V_1 = v_1, V_3 = v_3, V_5 = v_5) P(V_1 = v_1) P(V_3 = v_3 | V_1 = v_1) P(V_5 = v_5 | V_3 = v_3) \\ &= .98 + .014 \cdot F_5(t_1) + .006 \cdot F_6(t_1), \end{aligned}$$

Similarly, it holds that

$$P(\mathrm{L}_2 \leq t_2) = .994 + .0057 F_1(t_2) + .0003 F_3(t_2),$$

$$P(\mathrm{L}_4 \le t_4) = .9997 + .0003F_2(t_4),$$
$$P(\mathrm{L}_5 \le t_5) = .98 + .02F_5(t_5).$$

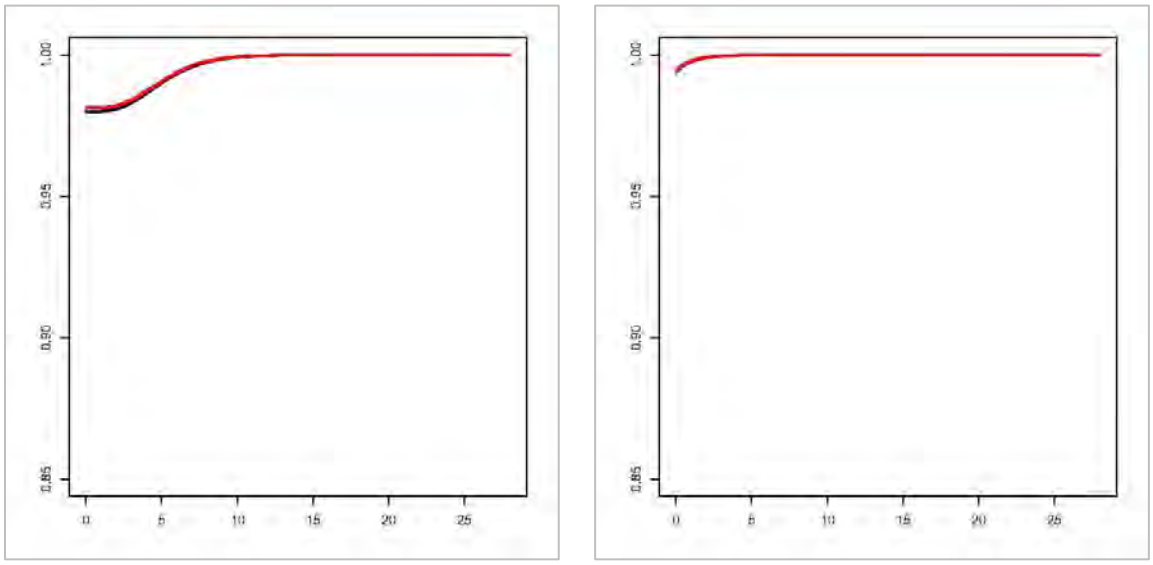
The distribution of total loss can be presented as

$$\begin{aligned} &P(\mathrm{L}_1 + \mathrm{L}_2 + \mathrm{L}_4 + \mathrm{L}_5 \le t) \\ &= \sum_{\mathbf{V}} P(\mathrm{L}_1 + \mathrm{L}_2 + \mathrm{L}_4 + \mathrm{L}_5 \le t | \mathbf{V} = \mathbf{v}) P(V_1 = v_1) P(V_3 = v_3 | V_1 = v_1) P(V_5 = v_5 | V_3 = v_3) \\ &= .98 + .014 \cdot F_{10}(t) + .0057 \cdot F_{12}(t) + .0003F_{16}(t). \end{aligned} \tag{6}$$

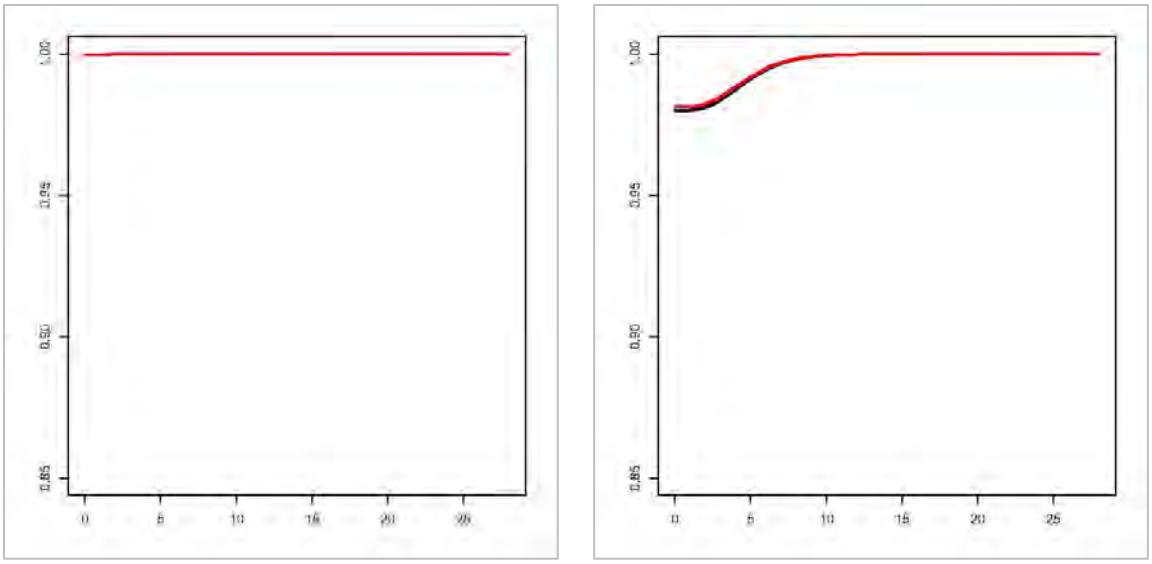
- **Simulation**. We also perform a simulation study to verify the loss distributions using Algorithm 1. In our case, the exploitation probabilities are
$$(e_1, e_3, e_5) = (.02, .3, .05).$$
The number of simulations is also set to be R = 10,000.

Figure 6

SIMULATED DISTRIBUTIONS (COLORED) AND THEORETICAL DISTRIBUTIONS (BLACK).



(a) L_1 (b) L_2

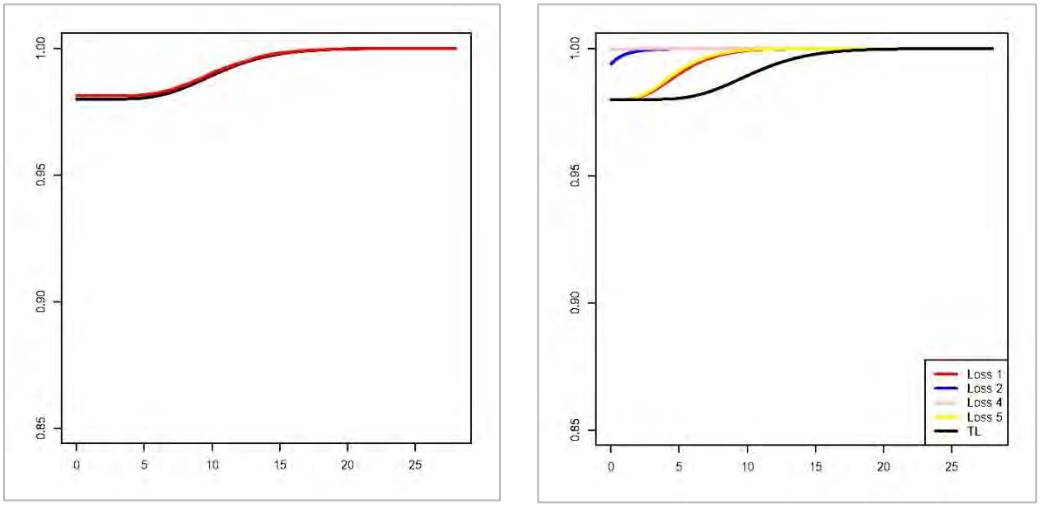


(c) L_4 (d) L_5

The distributions of different losses based on the BAG approach are displayed in Figure 6. It can be observed that the simulated distributions (black) coincide with the theoretical results (red). Figure 7(a) displays the theoretical distribution of the total loss based on Eq.(6) (black color) and the simulated distribution of the total loss (red color). We again observe that they match very well. Figure 7(b) shows all the losses together for comparison. L_4 is more likely to have smaller values than L_2. This can be explained by the fact that V_5, which leads to L_4, is the most difficult vulnerability to exploit since it can only be exploited after both V_1 and V_3 are successfully exploited. Since V_3 is relatively easier to exploit than V_5, L_2 affected by V_3 and V_5, it is more likely to have larger values compared to L_4. Compared with L_4 and L_2, L_1 and L_5 are less likely to have small values since V_1 and V_3 are exploited first.

Figure 7

THE SIMULATED DISTRIBUTION (COLORED) OF TOTAL LOSS AND THE THEORETICAL DISTRIBUTION (BLACK) OF TOTAL LOSS.



(a) Distributions of total loss (b) Simulated distributions of all losses

Table 6 displays the summary statistics of different types of losses and the total loss. Compared to Table 3, it is seen that L_1 and L_2, and TL are more likely to have small values. The means of losses are less than the corresponding ones by Approach 1. A similar conclusion can also be drawn for the standard deviations. It should also be mentioned that the large losses of TL (namely larger than Q_{99}) for both approaches are close. This can be explained by the fact that L_1 and L_5 both have V_1 that has a large impact on the total loss.

- Dependence among losses. Table 7 displays Kendall's tau correlation coefficients. It is observed L_1 and L_5 has the largest correlation .991, which is different from that in Table 4. This is mainly because V_1, which needs to be exploited first affects both L_1 and L_5. It is interesting to see that L_2 and L_5 have a large correlation .551 although they do not have a shared vulnerability. This strong dependence is caused by the chain relation in the BAG. We also observe that L_1 and L_5 are strongly correlated with TL as they are related to V_1. Since V_5 is the last to exploit, this results in less dependence of L_4 on TL.
  The tail indexes χ_us with $u = .999$ between different losses are also shown in Table 7. We again observe that L_1 and L_5 have the largest tail dependence, and both have large tail dependence with the total loss.

Table 6
SUMMARY STATISTICS OF DIFFERENT TYPES OF LOSSES AND THE TOTAL LOSS BASED ON SIMULATION, WHERE 'SD' REPRESENTS STANDARD DEVIATION.

	Min	Q_{25}	Median	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
L_1	.000	.000	.000	.000	.000	.000	4.844	6.487	9.314	17.176	.099	.780
L_2	.000	.000	.000	.000	.000	.000	.000	.144	1.811	7.652	.006	.113
L_4	.000	.000	.000	.000	.000	.000	.000	.000	.000	4.738	.001	.037
L_5	.000	.000	.000	.000	.000	.000	4.463	6.098	9.011	18.827	.093	.737
TL	.000	.000	.000	.000	.000	.000	9.912	12.501	16.440	27.260	.198	1.505

Table 7
KENDALL'S TAU CORRELATION COEFFICIENTS AND TAIL INDEXES $\chi(.999)$S BETWEEN DIFFERENT LOSSES.

	Kendall's tau				Tail index $\chi(.999)$			
	L_2	L_4	L_5	TL	L_2	L_4	L_5	TL
L_1	.552	.123	.991	.996	.040	.010	.060	.420
L_2		.223	.551	.554		.019	.000	.017
L_4			.123	.124			.020	.012
L_5				.995				.350

Insight 2. *The BAG has a significant impact on the loss distributions and tail dependence. It is recommended to use the BAG approach to assessing the risks. Approach 1 can be useful for the tail risk assessment of total loss if the total loss is mainly affected by some risks with vulnerabilities that can be easily exploited.*

General loss distributions Now, we study the impacts of different loss distributions. Assume that $\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_4, \mathrm{L}_5$ have different distributions:

$$\mathrm{L}_1 \sim \Gamma(\alpha_1 V_1 + \alpha_3 V_3, \beta), \tag{7}$$

$$\mathrm{L}_2 \sim \text{Lognormal}\,(\mu_3 \mathrm{V}_3 + \mu_5 \mathrm{V}_5, \sigma^2), \tag{8}$$

$$\mathrm{L}_4 \sim \exp(\lambda_5 V_5), \mathrm{L5} \sim \exp(\lambda_1 V_1). \tag{9}$$

Let $(\alpha_1, \alpha_3, \beta) = (5,1,1)$, $(\mu_3, \mu_5, \sigma) = (1,2,1)$, and $(\lambda_1, \lambda_5) = (.2, .5)$. Under this setting, L_2 has a heavy tail distribution that can lead to a significant loss. We perform a simulation study based on Algorithm 1 by using the above loss distributions, and the other settings are kept the same.

Table 8 shows the summary statistics of losses of each business line and TL. It is seen that L_2 has significantly larger values compared to the others for both Approach 1 and Approach 2.

Further, the Q_{99} of L_2 by Approach 1 is much larger than that by Approach 2 (31.027 vs 0). This is because L_2 having vulnerabilities V_3 and V_5 can only be exploited after V_1 is successfully exploited for Approach 2. Since the exploitation probability for V_1 is small (.02), this leads to the small values in Approach 2. Similarly, we observe that the total loss of Approach 2 is also much smaller.

Table 8

SUMMARY STATISTICS OF SIMULATED LOSSES BASED ON APPROACH 1 AND APPROACH 2.

	Min	Q_{25}	Median	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
	Approach 1											
L_1	.000	.000	.000	.243	1.310	2.255	5.398	6.884	9.657	17.148	.403	1.056
L_2	.000	.000	.000	1.569	5.791	1.526	31.027	45.978	101.570	587.541	2.223	8.319
L_4	.000	.000	.000	.000	.000	.046	3.241	4.542	7.780	15.023	.100	.619
L_5	.000	.000	.000	.000	.000	.000	3.463	7.119	14.583	41.677	.102	1.001
TL	.000	.000	.000	2.821	7.908	13.200	33.451	48.243	103.804	588.319	2.828	8.848
	Approach 2											
L_1	.000	.000	.000	.000	.000	.000	4.934	6.731	9.374	18.485	.105	.811
L_2	.000	.000	.000	.000	.000	.000	.000	.968	7.696	328.126	.036	1.388
L_4	.000	.000	.000	.000	.000	.000	.000	.000	.000	8.852	.000	.042
L_5	.000	.000	.000	.000	.000	.000	3.315	6.841	14.932	33.404	.099	.993
TL	.000	.000	.000	.000	.000	.000	1.197	14.825	24.618	334.562	.240	2.319

Table 9 shows Kendall's tau correlation coefficients and tail indexes. We again observe that L_1 and L_2 have the largest Kendall's tau correlation (.685) for Approach 1, and L_1 and L_5 have the largest Kendall's tau correlation (.990) for Approach 2. For the tail dependence, it is seen that L_2 dominates the tail dependence with the total loss for Approach 1, and the others are negligible. This is because the tail loss of L_1 is much larger than the others. However, it is not true for Approach 2 since we also observe that L_5 has a large tail dependence with the total loss. This is because that i) V_1 is the first vulnerability to be exploited which determines L_5; ii) Once V_1 is exploited, L_5 with an exponential distribution with a rate .2 that can produce a relatively large value (e.g., $Q_{99.9} = 14.932$).

Table 9

KENDALL'S TAU CORRELATION COEFFICIENTS AND TAIL INDEXES $\chi(.999)$S BETWEEN DIFFERENT LOSSES.

	L_2	L_4	L_5	TL	L_2	L_4	L_5	TL
	Kendall's tau							
	Approach 1				Approach 2			
L_1	.685	0	.269	.764	.542	.104	.990	.993
L_2		.366	-.001	.919		.193	.540	.545
L_4			-.002	.362			.104	.106
L_5				.225				.995
	Tail index $\chi(.999)$							
L_1	.000	.000	.050	.000	.050	.020	.030	.090
L_2		.010	.000	.990		.019	.007	.510
L_4			.000	.010			.000	.016
L_5				.000				.520

We draw the following insight based on the discussion:

Insight 3. *A business line with heavy tail distribution dominates the total loss.*

2.3.2 SCENARIO WITH MORE VULNERABILITIES

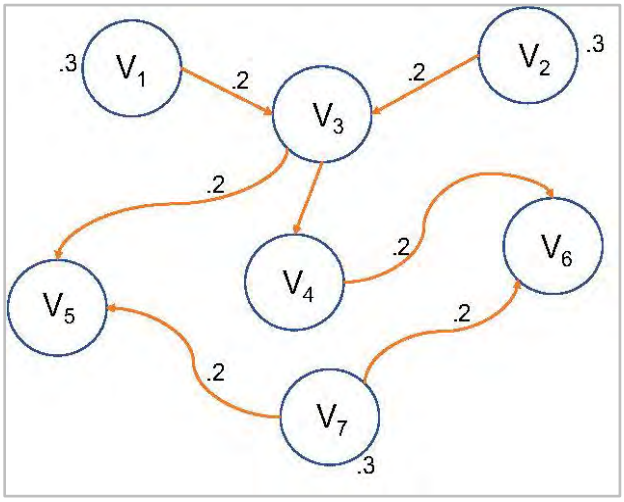
We consider the scenario with more vulnerabilities in Figure 3. Table 10 shows the impacts of each vulnerability. For simplicity, we assume that the outside exploitation probability for V_1, V_2, V_7 is .3, and the inside exploration is .2. The BAG is displayed in Figure 8 according to vulnerabilities.

Table 10

'X'MEANS VULNERABILITY V_i CAUSES LOSS $L_j, i = 1, \dots, 7, j = 1, \dots, 6$;

CVSS	L_1	L_2	L_3	L_4	L_5	L_6
CVE-2022-22667 (V_1)	x				x	
CVE-2021-27943 (V_2)	x					
CVE-2018-3919 (V_3)	x	x				
CVE-2021-39277(V_4)	x					
CVE-2021-21736 (V_5)		x		x		
CVE-2019-12944 (V_6)						x
CVE-2017-8759 (V_7)	x		x			

Figure 8
BAG IN THE SMART HOME ECOSYSTEM.



For the loss distributions, we assume that L_1 and L_2 follow exponential distributions, i.e., $\mathrm{L}_1 \sim \exp(\lambda_1)$, and $\mathrm{L}_2 \sim \exp(\lambda_2)$, where

$$\lambda_1 = \sum_{i \in \mathcal{L}_A} I(V_i), \quad \lambda_2 = \sum_{j \in \mathcal{L}_B} I(V_j)$$

and $\mathcal{L}_A = \{1,2,3,4,7\}$, and $\mathcal{L}_B = \{3,5\}$. L_3 and L_4 follow a lognormal distribution, i.e., $\mathrm{Lognorm}(\mu, \sigma)$, where $(\mu, \sigma) = (1,1)$. L_5 and L_6 follow a Gamma distribution, i.e., $\Gamma(\alpha, \beta)$, where $(\alpha, \beta) = (1,1)$.

The joint probabilities of vulnerabilities can be obtained from Figure 8,

$$\begin{aligned} P(\mathbf{V} = \mathbf{v}) = {} & P(V_1 = v_1)P(V_2 = v_2)P(V_7 = v_7) \\ & \cdot P(V_3 = v_3 | V_2 = v_2, V_1 = v_1)P(V_5 = v_5 | V_3 = v_3, V_7 = v_7) \\ & \cdot P(V_4 = v_4 | V_3 = v_3)P(V_6 = v_6 | V_4 = v_4, V_7 = v_7). \end{aligned}$$

Note that the exploitation probability of V_j can be represented as

$$e_j = P(V_j = 1 | \mathbf{pa}_j) = \begin{cases} 0, & \forall\, V_i \in \mathbf{pa}_j, V_i = 0; \\ 1 - \prod_{V_i \in \mathbf{pa}_j, V_i = 1} (1 - e_{ij}), & \text{Otherwise} \end{cases}, \tag{10}$$

where $e_{ij} = P(V_j = 1 | V_i = 1)$. For example, given $V_2 = 1$ and $V_3 = 1$, we have

$$e_3 = P(V_3 = 1 | V_2 = 1, V_1 = 1) = 1 - (1 - e_{23})(1 - e_{13}),$$

where $e_{23} = P(V_3 = 1 | V_2 = 1)$, and $e_{13} = P(V_3 = 1 | V_1 = 1)$. Since we have 7 vulnerabilities in the network, the total number of vulnerability scenarios is 2^7, i.e., 128[6]. Certainly, some of the scenarios have 0

[6] R script is available upon request to compute each scenario's probability.

probabilities due to the BAG. For example, Table 11 presents 5 scenarios with probabilities computed based on Eq. (10). It is seen that scenarios 3 and 5 both have probabilities of 0s.

Table 11
'Prob. ' REPRESENTS THE JOINT PROBABILITIES OF VULNERABILITIES IN DIFFERENT SCENARIOS.

Scenario	V_1	V_2	V_3	V_4	V_5	V_6	V_7	Prob.
1	0	0	0	0	0	0	0	.343
2	0	0	0	0	0	0	1	.094
3	0	0	0	0	0	1	0	.000
4	0	0	0	0	0	1	1	.024
5	0	0	0	0	1	0	0	.000

Algorithm 1 is used to simulate the losses, where the exploit probability e_j is computed based on Eq. (10), $j = 1, \ldots, 7$. The number of simulations is set to be R= 10,000. Table 12 displays the summary statistics of losses based on the simulation. It is seen that L_3 has the largest mean 1.363 and large values of high quantiles. This is because L_3 depends only on V_7 that has a large probability of being exploited from the outside, and further, it follows a lognormal distribution. Although L_4 also follows the same lognormal distribution, it has less probability of being exploited, as shown in the BAG. Therefore, L_4 has a smaller loss than L_3. Similar conclusions can be drawn for the other business lines.

Table 13 displays Kendall's tau correlations. It is seen that L_1 has small correlations with L_4 and L_6 since they do not have a common vulnerability; L_1 has common vulnerabilities with $\mathrm{L}_3(V_7)$ and L_5 (V_1), and they have relatively large correlations since both vulnerabilities can be directly exploited from the outside. Although L_1 has a common (V_3) with L_2, it has a relatively small correlation since V_3 cannot be exploited directly.

Table 12
SUMMARY STATISTICS OF DIFFERENT TYPES OF LOSSES AND THE TOTAL LOSS BASED ON THE SIMULATION.

	Min	Q_{25}	Median	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
L_1	.000	.000	.171	.671	1.432	2.041	3.512	4.137	5.627	8.029	.492	.765
L_2	.000	.000	.000	.000	.472	1.149	2.678	3.276	4.667	7.430	.158	.522
L_3	.000	.000	.000	1.038	4.199	7.284	17.169	23.212	39.256	78.983	1.363	3.839
L_4	.000	.000	.000	.000	.000	1.959	8.299	12.209	22.344	52.349	.353	1.893
L_5	.000	.000	.000	.180	1.102	1.844	3.350	4.066	5.629	7.693	.302	.715
L_6	.000	.000	.000	.000	.000	.235	1.830	2.594	4.527	9.025	.064	.369
TL	.000	.000	1.060	3.383	7.289	11.169	22.742	28.717	47.767	79.752	2.732	4.867

The other correlations can be interpreted similarly. For the tail dependence, it is seen that the tail dependence between different business lines is small. L_3 and L_4 have significant tail dependence with the total loss. This is mainly due to the lognormal loss, which coincides with the observation from Table 12.

Table 13
KENDALL'S TAU CORRELATION COEFFICIENTS AND TAIL INDEXES $\chi(.99)$S BETWEEN DIFFERENT LOSSES.

	Kendall's tau					
	L_2	L_3	L_4	L_5	L_6	TL
L_1	.137	.290	.127	.259	.120	.592
L_2		.184	.605	.200	.089	.355
L_3			.296	.010	.350	.620
L_4				.068	.112	.348
L_5					.011	.357
L_6						.268
	Tail index $\chi(.99)$					
L_1	0	0	.040	.010	0	.020
L_2		.020	.060	.020	.020	.050
L_3			.010	.010	.030	.660
L_4				.030	.000	.330
L_5					.010	.010
L_6						.020

Section 3: Systemic risk

In this section, we study a particular kind of risk, namely, *systemic risk*, in a smart home ecosystem which occurs when common vulnerabilities exist in many smart home networks. If an attacker successfully exploits the common vulnerabilities, this can cause catastrophic financial loss for the insurer. For instance, in 2017, the WannaCry ransomware attack targeted the computers running the Microsoft Windows Operating System by locking data and demanding ransom payments in the Bitcoin cryptocurrency. It was estimated that more than 200,000 computers were affected with a financial loss of up to billions of dollars [18]. The infamous ScarePackage ransomware attack reached over 900,000 Android cell phone users in just 30 days. If the user wanted to regain control of the device, they had to pay the criminal several hundred dollars in MoneyPak voucher(s).

Assume there are S smart homes in an insurance portfolio with common vulnerabilities $V_1, \dots, V_m$, and those vulnerabilities can cause a systemic risk. We assume that those vulnerabilities can be independently exploited from the outside. In the following section, we discuss two scenarios of systemic risks.

3.1 RANSOMWARE

A ransomware attack is possible in a smart home ecosystem. For example, the hacker can hack the smart lock to demand ransomware to unlock the door, or a smart thermostat can be hacked and display a message to pay a ransom. Since the ransomware attack typically demands the same amount of money for the systemic risk, the distribution of the total loss can be represented as

$$P(SL_3 \leq t) = \sum_{\mathbf{V}} P(L_3 \leq t/s|\ \mathbf{V}=\mathbf{v}) \prod_{i=1}^{m} P(V_i = v_i)\,.$$

This scenario becomes a single-line risk in a smart home, which can be analyzed similarly to those in Section 2.3.1. Note that the loss is enlarged S times compared to the loss in a single smart home in this scenario.

3.2 OTHER ATTACKS

We assume that once a common vulnerability is successfully exploited in one smart home, the other smart homes are subject to the same risk, and the same type of losses are independent and identically distributed. Then, the total loss of line i can be represented as

$$P(L_i^1 + L_i^2 + \cdots + L_i^S \leq t_1) = \sum_{\mathbf{V}} P(L_i^1 + L_i^2 + \cdots + L_i^S \leq t_1 \big| \mathbf{V} = \mathbf{v}) \prod_{i=1}^{m} P(V_i = v_i),$$

where L_i^j represents the loss of line i in smart home j, $j = 1, \ldots, S$.

For illustration, we consider the scenario in Section 2.3.1, i.e., assuming that V_1, V_3, V_5 are common vulnerabilities. The loss distributions are the same as those in Section 2.3.1, and $S = 100$.

Table 14 displays summary statistics of loss with different types and total loss. Compared to Table 3, it is seen that the values are much larger. For example, the mean of total loss increases from .995 to 10.296, and particularly, the standard deviation increases from 2.142 to 196.216. The high quantile of $Q_{99.5}$ increases from 13.191 to 1191.65. This indicates that the systemic risk can indeed cause a huge loss.

Insight 4. *The systemic risk in a smart home portfolio can be caused by one or more common vulnerabilities. It can result in a considerable loss. We suggest the insurer assess the systemic risk constantly.*

Table 14
SUMMARY STATISTICS OF LOSS WITH DIFFERENT TYPES AND TOTAL LOSS WITH GAMMA DISTRIBUTION BASED ON SIMULATION.

	Min	Q_{25}	Median	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
L_1	.000	.000	.000	91.419	105.888	113.108	521.267	59.029	629.788	655.514	4.734	88.769
L_2	.000	.000	.000	93.722	108.875	123.842	286.630	306.474	322.289	334.733	38.387	62.190
L_4	.000	.000	.000	.000	.000	.000	212.159	218.776	231.933	243.592	9.405	42.480
L_5	.000	.000	.000	.000	.000	.000	504.582	519.100	539.208	57.106	11.769	75.954
TL	.000	.000	.000	193.645	217.556	404.460	1039.804	1191.650	1405.556	1672.561	10.296	196.216

Section 4: Pricing cyber risks in a smart home ecosystem

In the current market, only a few companies provide insurance for smart homes, and the majority provide coverage via a personal cyber insurance policy[7]. According to a survey from Security.org[8], 83 percent of

[7] https://www.valuepenguin.com/personal-cyber-home-insurance
[8] https://www.security.org/insurance/cyber/cost/

respondents thought annual premiums of $25,000 coverage would cost under $150, while 93 percent thought they would cost less than $20. That explains why most insurers offer cyber insurance at the cost of less than $100 in the current market, and some charge even as little as $10 a month for $25,000 coverage. There are a limited number of companies that provide coverage for high-value homeowners, and the premiums for those higher-value endorsement range from a few hundred dollars to over a thousand per year, depending on the limit and coverage. For example, a major insurance company provides customized coverage with a limit of up to $250,000 with a premium of $1,652 per year. Since the smart home insurance market continues to expand, proper limits and affordable premiums need to be further studied. To address this, in the following, we discuss the pricing strategies for cyber risks in a smart home ecosystem.

Based on the loss analysis in the previous sections, we consider the following actuarial premium principles:

- Expectation principle: $\rho_1(L) = (1+\theta)E[L]$, where $\theta > 0$ is the loading parameter that reflects the risk preferences of the insurer.
- Standard deviation principle: $\rho_2(L) = E[L] + \theta\sqrt{Var(L)}$.
- Gini mean difference (GMD) principle: $\rho_3(L) = E[L] + \theta GMD(L)$ where
$$GMD(L) = E\,[|L_1 - L_2|],$$
is a statistical measure of variability, and L_1 and L_2 be a pair of independent copies of L; see [19, 20].
- Conditional tail expectation:
$$\rho_4(L) = E[L|L \geq VaR_\beta],$$
where VaR_β is the value-at-risk at level $\beta \in (0,1)$
$$VaR_\beta = \min_\gamma\{\gamma : P(L \leq \gamma) \geq \beta\},$$
For more details on the conditional tail expectation, please refer to [21, 22].

In the following, we consider the attack scenario in Figure 3. Similar to Section 2.3.2, we use the BAG approach. To mimic the risk in practice, we assume that the outside exploitation probability for V_1, V_2, and V_7 are .01, .02, and .9, respectively. Further, the inside exploration probability is .01. For the loss distributions, we assume that L_1 and L_2 follow exponential distributions, i.e., $\mathrm{L}_1 \sim \exp(\lambda_1)$, and $\mathrm{L}_2 \sim \exp(\lambda_2)$, where

$$\lambda_1 = \sum_{i\in\mathcal{L}_A} \boldsymbol{\alpha}_1 I(V_i) \quad , \lambda_2 = \sum_{j\in\mathcal{L}_B} \boldsymbol{\alpha}_2 I(V_j)$$

and $\mathcal{L}_A = \{1,2,3,4,7\}$, $\mathcal{L}_B = \{3,5\}$, and

$$\boldsymbol{\alpha}_1 = (1/160,1/32,1/80,0,0,1/160), \boldsymbol{\alpha}_2 = (0,0,1/640,0,1/320,0,0).$$

L_3 and L_4 follow a lognormal distribution, $\mathrm{Lognorm}(\mu,\sigma)$, where $(\mu,\sigma) = (7,1)$. L_5 and L_6 follow Gamma distributions, i.e., $\mathrm{L}_5 \sim \Gamma(\alpha_1,\beta)$, and $\mathrm{L}_6 \sim \Gamma(\alpha_2,\beta)$, where $(\alpha_1,\alpha_2,\beta) = (1000,2000,1)$.

Algorithm 1 is again used to simulate the losses, and the exploit probability e_j is also computed based on Eq. (9), $j = 1,\ldots,7$. The number of simulations is set to be R=10,000. Table 15 displays the summary statistics of losses based on the simulation. We observe huge losses under some scenarios (e.g., high quantiles of L_4 and L_6).

Table 15
SUMMARY STATISTICS OF DIFFERENT TYPES OF LOSSES AND THE TOTAL LOSS BASED ON THE SIMULATION.

	Min	Q_{25}	Median	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
L_1	.00	28.31	92.35	204.10	354.61	476.86	749.93	867.01	1143.49	1875.83	144.81	165.50
L_2	.00	.00	.00	.00	.00	.00	.00	199.73	704.06	1388.21	3.02	43.15
L_3	.00	21.26	48.13	99.59	191.60	273.29	57.05	727.44	1047.45	4597.62	83.16	123.43
L_4	.00	.00	.00	.00	.00	.00	.00	868.25	4867.59	15563.73	18.80	319.33
L_5	.00	.00	.00	.00	.00	.00	.00	998.13	1029.31	1069.61	9.46	96.69
L_6	.00	.00	.00	.00	.00	.00	.00	2004.76	2072.27	2150.52	19.70	198.09
TL	.00	87.22	181.37	332.07	539.64	771.11	2142.00	2458.68	5326.47	16521.96	278.95	465.62

4.1 PRICING BASED ON THE MARKET DATA

In this section, we study the pricing strategies based on market data. For comparison purposes, we consider the premium charged by a company (A) with $1,000 deductible and $50,000 coverage limit for one year policy period. Company A provides coverage for the following risks: cyber extortion, data restoration, crisis management, and cyber bullying. The yearly premiums are displayed in Table 16. Since we have the same business line L_4 (e.g. cyber extortion), it is used as a standard to determine the parameters in our pricing formulas. Specifically, we fix the premium for L_4 as 28, and determine the parameters in ρ_1 to ρ_4 based on the simulated losses. The parameters are .5, .03, .25, and .34, respectively. The deductible and coverage limit remain the same as those of company A.

Table 16
CYBER INSURANCE OFFERED BY COMPANY A WITH $1,000 DEDUCTIBLE AND $50,000 COVERAGE LIMIT FOR ONE YEAR POLICY PERIOD.

Coverage type	Cyber extortion	Data restoration	Crisis management	Cyber bullying	Total Premium
Premium	28	151	231	28	438

Table 17 shows the premium for each business line under different premium principles. The total premium is the sum of all premiums. In addition, we compute the premium based on the aggregated loss.

Table 17
PREMIUMS UNDER DIFFERENT PRICING PRINCIPLES, WHERE 'Total premium' IS THE SUM OF ALL PREMIUMS, AND 'Premium*' REPRESENTS THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

Premium	ρ_1	ρ_2	ρ_3	ρ_4
L_1	217	150	185	211
L_2	5	4	5	5
L_3	125	87	107	120
L_4	**28**	**28**	**28**	**28**
L_5	14	12	14	14
L_6	30	26	29	30

Total premium	418	307	368	408
Premium*	418	293	355	422

It is seen that for total premium, ρ_1 is the largest (418) while ρ_2 is the smallest (293). For the premium determined from the aggregated loss, we observe similar results. In the following, we assess the performance of premium principles based on the profit and loss ratio (LR):

$$\text{Profit} = \text{Premium} - \text{Claim},$$
$$\text{LR} = \frac{\text{Claim}}{\text{Premium}}$$

We assume that the permissible loss ratio is 40%.

Consider a portfolio with 500 policyholders who purchase smart home insurance policies. The premiums are charged according to Table 17. The loss scenarios of the portfolio are simulated 10,000 times. Table 18 shows the summary statistics of portfolio profit based on the loss of each business line and aggregated loss for different pricing principles. It is observed that, in this case, all the profits are positive. ρ_1 leads to the largest profit, and ρ_2 leads to the minimum profit. All the standard deviations are the same because the coverage limits and deducible are the same. The summary statistics and loss ratios are shown in Table 19. It is seen that the mean loss ratios are very low (namely less than .1) for all premium principles, and the high quantiles of loss ratios (e.g., $Q_{99.9}$) are still less than 40%. But the worst-case scenarios for ρ_2 and ρ_3 are beyond the permissible loss ratio of 40%.

Table 18

SUMMARY STATISTICS OF PROFITS UNDER DIFFERENT PRICING PRINCIPLES WITH $1,000 DEDUCTIBLE AND $50,000 COVERAGE LIMIT, WHERE 'PROFIT' REPRESENTS THE PROFIT BASED ON THE LOSS OF EACH BUSINESS LINE AND 'PROFIT*' REPRESENTS THE PROFIT BASED ON THE AGGREGATED LOSS.

		Min	Q_1	Q_5	Q_{10}	Q_{15}	Q_{50}	Q_{75}	Max	Mean	SD
ρ_1	Profit	127,549	174,542	183,192	186,869	189,152	196,215	199,612	207,628	195,089	6,429
	Profit*	127,549	174,542	183,192	186,869	189,152	196,215	199,612	207,628	195,089	6,429
ρ_2	Profit	72,049	119,042	127,692	131,369	133,652	140,715	144,112	152,128	139,589	6,429
	Profit*	65,049	112,042	120,692	124,369	126,652	133,715	137,112	145,128	132,589	6,429
ρ_3	Profit	102,549	149,542	158,192	161,869	164,152	171,215	174,612	182,628	170,089	6,429
	Profit*	96,049	143,042	151,692	155,369	157,652	164,715	168,112	176,128	163,589	6,429
ρ_4	Profit	122,549	169,542	178,192	181,869	184,152	191,215	194,612	202,628	190,089	6,429
	Profit*	116,549	163,542	172,192	175,869	178,152	185,215	188,612	196,628	184,089	6,429

Table 19

SUMMARY STATISTICS OF PROFITS UNDER DIFFERENT PRICING PRINCIPLES WITH $1,000 DEDUCTIBLE AND $50,000 COVERAGE LIMIT, WHERE 'LR' REPRESENTS THE LR BASED ON THE LOSS OF EACH BUSINESS LINE AND 'LR*' REPRESENTS THE LR BASED ON THE AGGREGATED LOSS.

		Min	Q_{25}	Q_{50}	Q_{75}	Q_{80}	Q_{90}	Q_{95}	$Q_{99.5}$	$Q_{9.99}$	Max	Mean	SD
ρ_1	LR	.01	.04	.06	.08	.09	.11	.12	.19	.24	.39	.07	.03
	LR*	.01	.04	.06	.08	.09	.11	.12	.19	.24	.39	.07	.03
ρ_2	LR	.01	.06	.08	.11	.12	.14	.17	.26	.32	.53	.09	.04
	LR*	.01	.06	.09	.12	.12	.15	.18	.27	.34	.56	.09	.04
ρ_3	LR	.01	.05	.07	.09	.10	.12	.14	.22	.27	.44	.08	.03
	LR*	.01	.05	.07	.10	.10	.12	.15	.22	.28	.46	.08	.04
ρ_4	LR	.01	.05	.06	.08	.09	.11	.13	.19	.24	.40	.07	.03
	LR*	.01	.05	.06	.09	.09	.11	13	.20	.25	.41	.07	.03

It should be pointed out that some companies are charging $0 deductibles in practice. For example, Company B offers a smart home policy covering cyber extortion and ransomware, cyber financial loss, and cyber personal protection. The coverage and limits are displayed in Table 20. We apply the premium strategy of company B to our simulated portfolio losses. The summary statistics of the profit and loss ratio are shown in Table 21. It can be seen that under this premium strategy, company B can't make a profit, and the mean loss ratio is 1.35, much larger than the permissible loss ratio.

Table 20

SMART HOME INSURANCE POLICY OFFERED BY COMPANY B WITH $0 DEDUCTIBLE AND $50,000 COVERAGE LIMIT FOR ALL COVERED EVENTS.

Coverage limit				Premium
Cyber extortion	Cyber financial loss	Cyber personal protection	All covered events	
10,000	50,000	50,000	50,000	200

Table 21

SUMMARY STATISTICS OF PROFIT AND LOSS RATIO UNDER THE PREMIUM STRATEGY OF COMPANY B.

	Min	Q_1	Q_5	Q10	Q50	Q95	Q995	Q999	Max	Mean	SD
Profit	-106,801	-61,674	-51,710	-46,830	-34,010	-20,912	-13,866	-10,635	-4,540	-34,764	9,416
LR	1.05	1.16	1.21	1.24	1.34	1.52	1.65	1.74	2.07	1.35	.09

4.2 PRICING STRATEGIES

In this section, we discuss adjusting the premiums under the permissible loss ratio based on the scenario of Table 17.

Deducible. We consider different deductibles other than \$1,000 while the premiums and coverage remain the same. In Appendix A, the profits and loss ratios are shown in Tables 24 to 33 under deductibles \$500, \$250, \$200, \$150, and \$100.

We consider two strategies: i) the permissible mean loss ratio is 40%; ii) the permissible high quantile of loss ratios $Q_{99.5}$ is 40%. Based on Tables 24 to 33, we recommend various deductibles for different pricing principles and show their mean profits in Table 22.

Table 22

RECOMMENDED DEDUCTIBLES UNDER DIFFERENT PRICING PRINCIPLES. DEDUCTIBLES/MEAN PROFIT 1 AND 2 ARE DETERMINED BASED ON THE MEAN LOSS RATIO AND 99.5TH LOSS RATIO OF 40%, RESPECTIVELY. 'Premium*' REPRESENTS THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

	Total premium		Coverage limit	Deductible 1	Mean Profit 1	Deductible 2	Mean Profit 2
ρ_1	Premium	418	50,000	150	131,809	250	154,670
	Premium*	418	50,000	150	131,809	250	154,670
ρ_2	Premium	307	50,000	250	99,170	500	125,380
	Premium*	293	50,000	250	92,170	500	118,380
ρ_3	Premium	368	50,000	200	119,583	500	155,880
	Premium*	355	50,000	200	113,083	500	149,380
ρ_4	Premium	408	50,000	150	126,809	500	175,880
	Premium*	396	50,000	150	120,809	500	169,880

Compared to Table 18, although the mean profits are reduced, the deductibles can be significantly smaller.

Premiums. We fix the coverage limit \$50,000 and the deductible \$1,000 to discuss different premium strategies. Because of the coverage limit and deductible, the premiums based on the four pricing principles are larger than the mean of total loss (i.e., 279). We set different premiums in Table 23, which correspond to 75%, 71%, 70%, 50%, 25%, and 20% of the mean of the total loss, respectively.

It is seen that if the permissible loss ratio of $Q_{99.5}$ is set to be 40%, the premium is \$198 and the mean profit is \$85,089; if the permissible mean loss ratio is to be 40%, the premium is \$70 and mean profit is \$21,089. Therefore, it is possible for insurers to offer a smart home insurance policy with a low premium (e.g., \$70) but decent coverage (i.e., \$50,000).

Table 23

SUMMARY STATISTICS OF PROFITS AND LOSS RATIOS WITH $1,000 DEDUCTIBLE AND $50,000 COVERAGE LIMIT.

Premium		Min	Q_1	Q_5	Q_{10}	Q_{50}	Q_{95}	$Q_{99.5}$	$Q_{99.9}$	Max	Mean	SD
210	Profit	23,549	70,542	79,192	82,869	92,215	99,089	101,761	102,699	103,628	91,089	6,429
	LR	.01	.04	.06	.07	.12	.25	.38	.47	.78	.13	.06
198	Profit	17,549	64,542	73,192	76,869	86,215	93,089	95,761	96,699	97,628	85,089	6,429
	LR	.01	.04	.06	.07	.13	.26	.40	.50	.82	.14	.06
195	Profit	16,049	63,042	71,692	75,369	84,715	91,589	94,261	95,199	96,128	83,589	6,429
	LR	.01	.04	.06	.07	.13	.26	.41	.50	.84	.14	.07
140	Profit	-11,451	35,542	44,192	47,869	57,215	64,089	66,761	67,699	68,628	56,089	6,429
	LR	.02	.06	.08	.10	.18	.37	.57	.70	1.16	.20	.09
70	Profit	-46,451	542	9,192	12,869	22,215	29,089	31,761	32,699	33,628	21,089	6,429
	LR	.04	.11	.17	.20	.37	.74	1.13	1.41	2.33	.40	.18
56	Profit	-53,451	-6,458	2,192	5,869	15,215	22,089	24,761	25,699	26,628	14,089	6,429
	LR	.05	.14	.21	.25	.46	.92	1.42	1.76	2.91	.50	.23

Section 5: Conclusion and Discussion

We have presented a practical quantitative framework for insurers to assess and price smart home cyber risks. The framework consists of four components: (i) identifying vulnerability-incurred cyber risks; (ii) classifying cyber risks into business lines; (iii) modeling cyber risks; and (iv) determining insurance premiums and coverages. We discover that common vulnerabilities can cause dependence among different business risks and tail dependence. We recommend the BAG approach for assessing cyber risks in the smart home ecosystem because it can provide more accurate results. It is worth mentioning that a business line with heavy tail distribution dominates the total loss. The systemic risk in a smart home portfolio is caused by one or more common vulnerabilities that can result in a considerable loss. We also recommend different pricing strategies with various deductibles and premiums for smart home policies, which can make the products more attractive in the market.

The current study has some limitations which need to be addressed in future studies. First, the loss distributions and parameters are set based on experience and limited market data. They can be further calibrated when the actual claim data are available. Second, smart homes may have different network structures and protocols, requiring an individualized risk assessment (i.e., individual BAG analysis). This can further enhance the premium strategy by creating individualized smart home policies. However, it requires much effort from insurers, such as security assessment and loss estimation. Third, the coverage of ransomware or extortion may need further attention in practice since the U.S. government strongly discourages all private companies and citizens from paying ransom or extortion demands, and the payment of ransomware claims may lead to some legal issues[9]. Nevertheless, the present study provides practical guidance for insurers to price cyber risks at the initial stage of the development of the smart home insurance market.

[9] https://www.fbi.gov/how-we-can-help-you/safety-resources/scams-and-safety/common-scams-and-crimes/ransomware

Section 6: Acknowledgments

The authors' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the volunteers who generously shared their wisdom, insights, advice, guidance, and arm's-length review of this study prior to publication. Any opinions expressed may not reflect their opinions nor those of their employers. Any errors belong to the authors alone.

Project Oversight Group members:

Laura Bass
Ron Harasym
Jill Harper
Shariq Sikander
Jianxi Su
Xinping Yuan

At the Society of Actuaries Research Institute:

Scott Lennox, FSA, FCAS, FCIA
Rob Montgomery, ASA, MAAA, FLMI

The Society of Actuaries Research Institute would like to acknowledge the generous contribution of the Casualty Actuarial Society and the Joint Risk Management Section to the funding of this research.

Appendix A: Various deductibles

Table 24

SUMMARY STATISTICS OF PROFITS UNDER $500 DEDUCTIBLE AND $50,000 COVERAGE. 'Profit*' REPRESENTS THE PROFIT BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_1	Q_5	Q_{10}	Q_{15}	Q_{50}	Q_{75}	Max	Mean	SD
ρ_1	Profit	112,217	157,355	166,918	170,996	173,536	181,815	186,350	199,217	180,880	7,743
	Profit*	112,217	157,355	166,918	170,996	173,536	181,815	186,350	199,217	180,880	7,743
ρ_2	Profit	56,717	101,855	111,418	115,496	118,036	126,315	130,850	143,717	125,380	7,743
	Profit*	49,717	94,855	104,418	108,496	111,036	119,315	123,850	136,717	118,380	7,743
ρ_3	Profit	87,217	132,355	141,918	145,996	148,536	156,815	161,350	174,217	155,880	7,743
	Profit*	80,717	125,855	135,418	139,496	142,036	150,315	154,850	167,717	149,380	7,743
ρ_4	Profit	107,217	152,355	161,918	165,996	168,536	176,815	181,350	194,217	175,880	7,743
	Profit*	101,217	146,355	155,918	159,996	162,536	170,815	175,350	188,217	169,880	7,743

Table 25

SUMMARY STATISTICS OF LOSS RATIOS UNDER $500 DEDUCTIBLE AND $50,000 COVERAGE. 'LR*' REPRESENTS THE LOSS RATIO BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_{25}	Q_{50}	Q_{75}	Q_{80}	Q_{90}	Q_{95}	$Q_{99.5}$	$Q_{9.99}$	Max	Mean	SD
ρ_1	LR	.05	.11	.13	.15	.16	.18	.20	.27	.31	.46	.13	.04
	LR*	.05	.11	.13	.15	.16	.18	.20	.27	.31	.46	.13	.04
ρ_2	LR	.06	.15	.18	.21	.22	.25	.27	.37	.43	.63	.18	.05
	LR*	.07	.15	.19	.22	.23	.26	.29	.38	.45	.66	.19	.05
ρ_3	LR	.05	.12	.15	.18	.18	.21	.23	.31	.36	.53	.15	.04
	LR*	.06	.13	.15	.18	.19	.21	.24	.32	.37	.55	.16	.04
ρ_4	LR	.05	.11	.13	.16	.16	.19	.21	.28	.32	.47	.14	.04
	LR*	.05	.11	.14	.16	.17	.19	.21	.28	.33	.49	.14	.04

Table 26

SUMMARY STATISTICS OF PROFITS UNDER $250 DEDUCTIBLE AND $50,000 COVERAGE. 'Profit*' REPRESENTS THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_1	Q_5	Q_{10}	Q_{15}	Q_{50}	Q_{75}	Max	Mean	SD
ρ_1	Profit	85,759	129,401	139,186	143,625	146,351	155,495	160,688	178,804	154,670	8,672
	Profit*	85,759	129,401	139,186	143,625	146,351	155,495	160,688	178,804	154,670	8,672
ρ_2	Profit	30,259	73,901	83,686	88,125	90,851	99,995	105,188	123,304	99,170	8,672
	Profit*	23,259	66,901	76,686	81,125	83,851	92,995	98,188	116,304	92,170	8,672
ρ_3	Profit	60,759	104,401	114,186	118,625	12,1351	130,495	135,688	153,804	129,670	8,672
	Profit*	54,259	97,901	107,686	112,125	114,851	123,995	129,188	147,304	123,170	8,672
ρ_4	Profit	80,759	124,401	134,186	138,625	141,351	150,495	155,688	173,804	149,670	8,672
	Profit*	74,759	118,401	128,186	132,625	135,351	144,495	149,688	167,804	143,670	8,672

Table 27

SUMMARY STATISTICS OF LOSS RATIOS UNDER $250 DEDUCTIBLE AND $50,000 COVERAGE. 'LR*' REPRESENTS THE LOSS RATIO BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_{25}	Q_{50}	Q_{75}	Q_{80}	Q_{90}	Q_{95}	$Q_{99.5}$	$Q_{9.99}$	Max	Mean	SD
ρ_1	LR	.14	.23	.26	.28	.29	.31	.33	.40	.45	.59	.26	.04
	LR*	.14	.23	.26	.28	.29	.31	.33	.40	.45	.59	.26	.04
ρ_2	LR	.20	.31	.35	.39	.40	.43	.45	.55	.61	.80	.35	.06
	LR*	.21	.33	.37	.40	.41	.45	.48	.58	.64	.84	.37	.06
ρ_3	LR	.16	.26	.29	.32	.33	.36	.38	.46	.51	.67	.30	.05
	LR*	.17	.27	.30	.33	.34	.37	.39	.48	.53	.69	.31	.05
ρ_4	LR	.15	.24	.26	.29	.30	.32	.34	.41	.46	.60	.27	.04
	LR*	.15	.24	.27	.30	.31	.33	.35	.43	.47	.62	.27	.04

Table 28

SUMMARY STATISTICS OF PROFITS UNDER $200 DEDUCTIBLE AND $50,000 COVERAGE. 'Profit*' REPRESENTS THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_1	Q_5	Q_{10}	Q_{15}	Q_{50}	Q_{75}	Max	Mean	SD
ρ_1	Profit	75,128	118,921	128,691	133,275	136,076	145,387	150,741	169,937	144,583	8,870
	Profit *	75,128	118,921	128,691	133,275	136,076	145,387	150,741	169,937	144,583	8,870
ρ_2	Profit	19,628	63,421	73,191	77,775	80,576	89,887	95,241	114,437	89,083	8,870
	Profit*	12,628	56,421	66,191	70,775	73,576	82,887	88,241	107,437	82,083	8,870
ρ_3	Profit	50,128	93,921	103,691	108,275	111,076	120,387	125,741	144,937	119,583	8,870
	Profit*	43,628	87,421	97,191	101,775	104,576	113,887	119,241	138,437	113,083	8,870
ρ_4	Profit	70,128	113,921	123,691	128,275	131,076	140,387	145,741	164,937	139,583	8,870
	Profit*	64,128	107,921	117,691	122,275	125,076	134,387	139,741	158,937	133,583	8,870

Table 29

SUMMARY STATISTICS OF LOSS RATIOS UNDER $200 DEDUCTIBLE AND $50,000 COVERAGE. 'LR*' REPRESENTS THE LOSS RATIO BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_{25}	Q_{50}	Q_{75}	Q_{80}	Q_{90}	Q_{95}	$Q_{99.5}$	$Q_{9.99}$	Max	Mean	SD
ρ_1	LR	.19	.28	.30	.33	.34	.36	.38	.45	.50	.64	.31	.04
	LR*	.19	.28	.30	.33	.34	.36	.38	.45	.50	.64	.31	.04
ρ_2	LR	.25	.38	.41	.45	.46	.49	.52	.62	.68	.87	.42	.06
	LR*	.27	.40	.43	.47	.48	.52	.55	.65	.71	.91	.44	.06
ρ_3	LR	.21	.32	.35	.38	.39	.41	.44	.52	.57	.73	.35	.05
	LR*	.22	.33	.36	.39	.40	.43	.45	.53	.59	.75	.36	.05
ρ_4	LR	.19	.29	.31	.34	.35	.37	.39	.47	.51	.66	.32	.04
	LR*	.20	.29	.32	.35	.36	.38	.41	.48	.53	.68	.33	.04

Table 30

SUMMARY STATISTICS OF PROFITS UNDER $150 DEDUCTIBLE AND $50,000 COVERAGE. 'Profit*' REPRESENTS THE PROFIT BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_1	Q_5	Q_{10}	Q_{15}	Q_{50}	Q_{75}	Max	Mean	SD
ρ_1	Profit	61,749	105,618	115,508	120,208	123,125	132,594	138,089	158,763	131,809	9,059
	Profit*	61,749	105,618	115,508	120,208	123,125	132,594	138,089	158,763	131,809	9,059
ρ_2	Profit	6,249	50,118	60,008	64,708	67,625	77,094	82,589	103,263	76,309	9,059
	Profit*	-751	43,118	53,008	57,708	60,625	70,094	75,589	96,263	69,309	9,059
ρ_3	Profit	36,749	80,618	90,508	95,208	98,125	107,594	113,089	133,763	106,809	9,059
	Profit*	30,249	74,118	84,008	88,708	91,625	101,094	106,589	127,263	100,309	9,059
ρ_4	Profit	56,749	100,618	110,508	115,208	118,125	127,594	133,089	153,763	126,809	9,059
	Profit*	50,749	94,618	104,508	109,208	112,125	121,594	127,089	147,763	120,809	9,059

Table 31

SUMMARY STATISTICS OF LOSS RATIOS UNDER $150 DEDUCTIBLE AND $50,000 COVERAGE. 'LR*' REPRESENTS THE LOSS RATIO BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_{25}	Q_{50}	Q_{75}	Q_{80}	Q_{90}	Q_{95}	$Q_{99.5}$	$Q_{9.99}$	Max	Mean	SD
ρ_1	LR	.24	.34	.37	.39	.40	.42	.45	.52	.56	.70	.37	.04
	LR*	.24	.34	.37	.39	.40	.42	.45	.52	.56	.70	.37	.04
ρ_2	LR	.33	.46	.50	.54	.55	.58	.61	.70	.76	.96	.50	.06
	LR*	.34	.48	.52	.56	.57	.61	.64	.74	.80	1.01	.53	.06
ρ_3	LR	.27	.39	.42	.45	.46	.48	.51	.59	.64	.80	.42	.05
	LR*	.28	.40	.43	.46	.47	.50	.53	.61	.66	.83	.43	.05
ρ_4	LR	.25	.35	.37	.40	.41	.44	.46	.53	.58	.72	.38	.04
	LR*	.25	.36	.39	.42	.42	.45	.47	.54	.59	.74	.39	.05

Table 32

SUMMARY STATISTICS OF PROFITS UNDER $100 DEDUCTIBLE AND $50,000 COVERAGE. 'Profit*' REPRESENTS THE PROFIT BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_1	Q_5	Q_{10}	Q_{15}	Q_{50}	Q_{75}	Max	Mean	SD
ρ_1	Profit	44,965	89,249	99,248	104,001	107,026	116,589	122,186	144,683	115,821	9,223
	Profit*	44,965	89,249	99,248	104,001	107,026	116,589	122,186	144,683	115,821	9,223
ρ_2	Profit	-10,535	33,749	43,748	48,501	51,526	61,089	66,686	89,183	60,321	9,223
	Profit*	-17,535	26,749	36,748	41,501	44,526	54,089	59,686	82,183	53,321	9,223
ρ_3	Profit	19,965	64,249	74,248	79,001	82,026	91,589	97,186	119,683	90,821	9,223
	Profit*	13,465	57,749	67,748	72,501	75,526	85,089	90,686	113,183	84,321	9,223
ρ_4	Profit	39,965	84,249	94,248	99,001	102,026	111,589	117,186	139,683	110,821	9,223
	Profit*	33,965	78,249	88,248	93,001	96,026	105,589	111,186	133,683	104,821	9,223

Table 33

SUMMARY STATISTICS OF LOSS RATIOS UNDER $100 DEDUCTIBLE AND $50,000 COVERAGE. 'LR*' REPRESENTS THE LOSS RATIO BASED ON THE PREMIUM DETERMINED FROM THE AGGREGATED LOSS.

		Min	Q_{25}	Q_{50}	Q_{75}	Q_{80}	Q_{90}	Q_{95}	$Q_{99.5}$	$Q_{9.99}$	Max	Mean	SD
ρ_1	LR	.31	.42	.44	.47	.48	.50	.53	.59	.63	.78	.45	.04
	LR*	.31	.42	.44	.47	.48	.50	.53	.59	.63	.78	.45	.04
ρ_2	LR	.42	.57	.60	.64	.65	.68	.71	.81	.86	1.07	.61	.06
	LR*	.44	.59	.63	.67	.68	.72	.75	.84	.90	1.12	.64	.06
ρ_3	LR	.35	.47	.50	.54	.54	.57	.60	.67	.72	.89	.51	.05
	LR*	.36	.49	.52	.56	.56	.59	.62	.70	.75	.92	.52	.05
ρ_4	LR	.32	.43	.45	.48	.49	.51	.54	.61	.65	.80	.46	.05
	LR*	.32	.44	.47	.50	.51	.53	.55	.62	.67	.83	.47	.05

References

[1] Davit Marikyan, Savvas Papagiannidis, and Eleftherios Alamanos. A systematic review of the smart home literature: A user perspective. *Technological Forecasting and Social Change*, 138:139–154, 2019.

[2] Omar Alrawi, Chaz Lever, Manos Antonakakis, and Fabian Monrose. Sok: Security evaluation of home-based iot deployments. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1362–1380, 2019.

[3] Andreas Jacobsson, Martin Boldt, and Bengt Carlsson. A risk analysis of a smart home automation system. *Future Generation Computer Systems*, 56:719–733, 2016.

[4] Joseph Bugeja, Andreas Jacobsson, and Paul Davidsson. Prash: A framework for privacy risk analysis of smart homes. *Sensors*, 21(19):6399, 2021.

[5] Denis Kozlov, Jari Veijalainen, and Yasir Ali. Security and privacy threats in iot architectures. In *Proceedings of the 7th International Conference on Body Area Networks*, pages 256–262, 2012.

[6] Changmin Lee, Luca Zappaterra, Kwanghee Choi, and Hyeong-Ah Choi. Securing smart home: Technologies, security challenges, and security requirements. In *2014 IEEE Conference on Communications and Network Security*, pages 67–72. IEEE, 2014.

[7] Xiaoyu Zhang, Maochao Xu, Jianxi Su, and Peng Zhao. Structural models for fog computing based internet of things architectures with insurance and risk management applications. *European Journal of Operational Research*, 305(3):1273–1291, 2023.

[8] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[9] Tamara Denning, Tadayoshi Kohno, and Henry M Levy. Computer security and the modern home. *Communications of the ACM*, 56(1):94–103, 2013.

[10] Somak R Das, Silvia Chita, Nina Peterson, Behrooz A Shirazi, and Medha Bhadkamkar. Home automation and security for mobile devices. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 141– 146. IEEE, 2011.

[11] Brittany D Davis, Janelle C Mason, and Mohd Anwar. Vulnerability studies and security postures of iot devices: a smart home case study. *IEEE Internet of Things Journal*, 7(10):10102– 10110, 2020.

[12] Common vulnerability scoring system. http://www.rst.org/cvss/cvss-guide.html. Accessed: 2021-12-30.

[13] Jay Jacobs, Sasha Romanosky, Benjamin Edwards, Idris Adjerid, and Michael Roytman. Exploit prediction scoring system (epss). *Digital Threats: Research and Practice*, 2(3):1–17, 2021.

[14] Michal Walkowski, Maciej Krakowiak, Jacek Oko, and Sl awomir Sujecki. Efficient algorithm for providing live vulnerability assessment in corporate network environment. *Applied Sciences*, 10(21):7926, 2020.

[15] Nayot Poolsappasit, Rinku Dewri, and Indrajit Ray. Dynamic security risk management using bayesian attack graphs. *IEEE Transactions on Dependable and Secure Computing*, 9(1):61–74, 2011.

[16] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. *Statistics of extremes: theory and applications*, volume 558. John Wiley & Sons, 2004.

[17] Rolf-Dieter Reiss and Michael Thomas. *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*. Springer Science & Business Media, 2007.

[18] Soheil Askarifar, NAA Rahman, and Hasbullah Osman. A review of latest wannacry ransomware: Actions and preventions. *J. Eng. Sci. Technol*, 13:24–33, 2018.

[19] Edward Furman, Ruodu Wang, and Riˇcardas Zitikis. Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking & Finance*, 83:70– 84, 2017.

[20] Edward Furman, Yisub Kye, and Jianxi Su. Computing the gini index: A note. *Economics Letters*, 185:108753, 2019.

[21] Mary R Hardy. An introduction to risk measures for actuarial applications. *SOA Syllabus Study Note*, 19, 2006.

[22] Dirk Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533, 2002.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org