

Post-Incident Audits on Cyber Insurance Discounts

Sakshyam Panda^a, Daniel W Woods^b, Aron Laszka^c, Andrew Fielder^d, Emmanouil Panaousis^{a,*}

^aUniversity of Surrey

^bUniversity of Oxford

^cUniversity of Houston

^dImperial College London

Abstract

We introduce a game-theoretic model to investigate the strategic interaction between a cyber insurance policyholder whose premium depends on her self-reported security level and an insurer with the power to audit the security level upon receiving an indemnity claim. Audits can reveal fraudulent (or simply careless) policyholders not following reported security procedures, in which case the insurer can refuse to indemnify the policyholder. However, the insurer has to bear an audit cost even when the policyholders have followed the prescribed security procedures. As audits can be expensive, a key problem insurers face is to devise an auditing strategy to deter policyholders from misrepresenting their security levels to gain a premium discount. This decision-making problem was motivated by conducting interviews with underwriters and reviewing regulatory filings in the US; we discovered that premiums are determined by security posture, yet this is often self-reported and insurers are concerned by whether security procedures are practised as reported by the policyholders.

To address this problem, we model this interaction as a Bayesian game of incomplete information and devise optimal auditing strategies for the insurers considering the possibility that the policyholder may misrepresent her security level. To the best of our knowledge, this work is the first theoretical consideration of post-incident claims management in cyber security. Our model captures the trade-off between the incentive to exaggerate security posture during the application process and the possibility of punishment for non-compliance with reported security policies. Simulations demonstrate that common sense techniques are not as efficient at providing effective cyber insurance audit decisions as the ones computed using game theory.

Keywords: Game theory, Cyber insurance, Economics of security, Premium discount, Post-incident audit, Security

1. Introduction

No amount of investment in security eliminates the risk of loss [1]. Driven by the frequency of cyber attacks, risk-averse organizations increasingly transfer residual risk by purchasing cyber insurance. As a result, the cyber-insurance market is predicted to grow to between \$7.5 and \$20 billion by 2020, as identified in [2].

Similar to other types of insurance, cyber-insurance providers pool the risk from multiple policyholders together and charge a premium to cover the underlying risk. Yet cyber risks like data breaches are qualitatively different from traditional lines like property insurance. For instance, buildings are built once according to building regulations, whereas computer systems continually change as mobile devices blur the network perimeter and software evolves with additional product features and security patches. Adversaries

update strategies to exploit vulnerabilities emerging from technological flux.

Further, the problems of moral hazard and adverse selection become more pressing. Adverse selection results from potential clients being more likely to seek insurance if they face a greater risk of loss. Meanwhile, information asymmetry limits insurers in assessing the applicant's risk. The risk depends on computer systems with many devices in different configurations, users with a range of goals, and idiosyncratic organizational security teams, policies, and employed controls. Collecting information is a costly procedure, let alone assessing and quantifying the corresponding risk.

Moral hazard occurs when insureds engage in riskier behaviour in the knowledge that the insurer will indemnify any losses. Even if initial assessment reveals that security policies are in place, it is no guarantee that they will be followed given that “a significant number of security breaches result from employees’ failure to comply with security policies” [3]. Technological compliance suffers too, as evidenced by the Equifax breach resulting from not patching a publicly known vulnerability [4].

*corresponding author

Email addresses: s.panda@surrey.ac.uk (Sakshyam Panda), daniel.woods@cs.ox.ac.uk (Daniel W Woods), alaszka@uh.edu (Aron Laszka), andrew.fielder@imperial.ac.uk (Andrew Fielder), e.panaousis@surrey.ac.uk (Emmanouil Panaousis)

Insurance companies collect risk information about applicants to address adverse selection. We interviewed 9 underwriters in the UK and found that 8 of them use self-reported application forms; 7 of them use telephone calls with the applicant; 3 of them use external audits; and only one uses on-site audits¹. This suggests that the application process relies on accurate self-reporting of risk factors. Cyber insurance application forms collect information about questions ranging from generic business topics to questions related to information security controls [5].

Romanosky et al. [6] introduced a data set resulting from a US law requiring insurers to file documents describing their pricing schemes. Pricing algorithms depended on the applicant’s industry, revenue, past-claims history, and—most relevant to this paper—the security controls employed by the organization. The insurer collects all this information and sets a price according to the formulas described in [6], reducing the premium when security controls are reported to be in place. This was corroborated by interviews with insurance professionals in Sweden [7]. Surprisingly, individual underwriters determine the size of the discount for security controls on a case-by-case basis, even though this can be as large as 25% of the premium.

Moral hazard is generally addressed by including terms in the policy that insureds must follow for their coverage to be valid. An early study found that coverage was excluded for a “failure to take reasonable steps to maintain and upgrade security” [8]. A study from 2017 found few exclusions prescribing security procedures but the majority of policies contained exclusions for “dishonest acts” [6]. One such dishonest act is violating the *principle of up-most good faith* requiring insureds to not intentionally deceive the company offering the insurance [9].

This principle and the corresponding exclusion mitigates moral hazard, which might otherwise drive honest firms to de-prioritize compliance with security procedures. Further, it imposes a cost on fraudulent organizations claiming that entirely fictional security products are in place to receive a lower premium. For example, one insurer refused to pay out on a cyber policy because “security patches were no longer even available, much less implemented” [10] despite the application form reporting otherwise. We do not consider the legality of this case, but include it as evidence that insurers conduct audits to establish whether there are grounds for refusing coverage.

Further, insurers offer discounts for insureds based on security posture and often rely on self-reports that security controls are in place. Interviewing insurers revealed concerns about whether security policies were being complied with in reality. Besides, larger premium discounts increase the incentive to misrepresent security levels potentially necessitating a higher frequency of investigation

which is uneconomical for insurers. To explore how often should insurers audit cyber insurance claims, we develop a game-theoretic model that takes into account relevant parameters from pricing data collected by analyzing 26 cyber insurance pricing schemes filed in California and identifies different optimal auditing strategies for insurers. Our analytical approach relies on *Perfect Bayesian Equilibrium* (PBE). We complement our analysis with simulation results with parameter values from the collected data. We further make “common sense” assumptions regarding auditing strategies and show that in general, insurers are better-off with the game-theoretic strategies. The results will be of interest to policymakers in the United States and the European Union, who believe cyber insurance can improve security levels by offering premium discounts [11].

The remainder of this paper is organized as follows. Section 2 identifies existing approaches to modeling the cyber insurance market. We introduce our game-theoretic model in Section 3 and present the analysis in Section 4. Section 5 details our methodology for data collection which instantiate our simulation results. Finally, we end with concluding remarks in Section 6.

2. Related Work

This paper continues the trend towards rectifying the “substantial discrepancy” [12] between early cyber insurance models and informal claims about the insurance market. Early research considered factors relevant to the viability of a market. Interdependent security occurs when the risk “depends on the actions of others” [13, 14]. Optimists argued that insurers could coordinate the resulting collective action problem [15, 16], leading to a net social welfare gain and a viable market. Skeptics instead focused on the “high correlation in failure of information systems” [17, 18, 19], citing it as a major impediment to the supply of cyber insurance. Recent empirical work [6] analyzing 180 cyber insurance filings shows that the cyber insurance market is viable.

Beyond viability, researchers explored how insurers can intervene by sharing information, assessing the security of service providers, and investing in software quality. The insurer sharing information about claims data was shown to increase social welfare in [20]. Khalili et al. [21] show that underwriting service providers improves both insurer profit and social welfare. Laszka et al. [22] found that the insurer directly investing in software quality can “reduce non-diversifiable risks and can lead to a more profitable cyber insurance market”.

The timing of the insurer’s intervention plays is an important strategic aspect. Ex-ante interventions for the insurer include risk assessments and security investments before the policy term begins. Shetty et al. [23] investigated an insurer who could assess security levels perfectly or not at

¹Note that these are mutually inclusive events.

all, concluding that the latter cannot support a functioning market. Majuca et al. [24] showed that ex-ante assessments in combination with discounts for adopting security controls can lead to an increase in social welfare. A more recent model introduces stochastic uncertainty about the policyholder’s security level [25].

None of these adverse selection studies consider the potential for insureds to misrepresent their security posture. Allowing malicious insureds to “subvert insurer monitoring” in both the application process and over the policy period was studied in [26]. The analysis showed that no cyber insurance market could exist. However, we know from [10] that insurers audit insureds and refuse coverage for fraudulent claims. Our model deviates from [26] by allowing the insurer to audit claims and withdraw coverage if the insured misrepresents information.

Beyond ex-ante assessment, insurers make decisions regarding ex-post claims management. These decisions have received less attention. The impact of secondary losses on the policyholder’s incentive to claim could lead to overpriced products [27]. Further, insurers can aggregate claims information to increase social welfare [20]. Empirically it has been suggested insurers will refuse “claims arising from the insured’s failure to maintain security levels”, but the strategic aspects of insurers investigating the incidents leading to claims has not been considered [8].

The literature on economic theory of insurance fraud has developed two main approaches: *costly state verification* and *costly state falsification* [28]. The costly state falsification approach assesses the client’s behaviour towards a claim. We consider the costly state verification approach, which focuses on the insurer identifying fraudulent claims. The insurer can verify the claims via auditing but has to bear a verification cost. The optimal claim handling usually involves random auditing [29].

Our contribution to the literature is the first theoretical consideration of post-incident claims management. Our model captures the trade-off between the incentive to exaggerate security posture to receive a premium discount and the possibility of punishment for non-compliance with the reported security policies. We consider misrepresenting security posture a strategic choice for the insured and allow the insurer to respond by auditing claims. Not allowing the insurer to do so leads to market collapse [26].

3. Model

We model the interaction between the policyholder \mathcal{P} and insurer \mathcal{I} as a one-shot dynamic game called the *Cyber Insurance Audit Game* (CIAG), which is represented in Figure 1. Each decision node of the tree represents a state where the player with the move has to choose an action. The leaf nodes present the payoffs of the players for the

sequence of chosen actions. The payoffs are represented in the format $\begin{pmatrix} x \\ y \end{pmatrix}$, where x and y are the payoffs of \mathcal{P} and \mathcal{I} , respectively. Table 1 presents the list of symbols used in our model. Note that the initial wealth of the policyholder (W) and the premium for insurance coverage (p) are omitted from the tree for ease of presentation.

Table 1: List of Symbols

Symbol	Description
a	Cost of audit
c	Security investment cost
d	Discount on premium for better security level
l	Loss due to a breach
p	Premium for the insurance coverage
W	Initial wealth of the policyholder
β	Probability of a breach
β^*	Probability of a breach after investment

We assume that the policyholder does not make a decision regarding its security investment in our model, because that decision has been made before seeking insurance. Hence, a particular applicant has a certain fixed type (with respect to security), but the insurer does not know the type of an applicant due to information asymmetry. We can model the insurer’s uncertainty by assuming that it encounters certain types of applicants with certain probabilities. The type of the policyholder is modeled as an outcome of a random event, that is, nature (N) decides the policyholder’s type with respect to additional security investments, i.e., \mathcal{P}_S represents one with additional security investments and \mathcal{P}_N one without. Further, nature also decides whether a security incident occurs for each policyholder, represented as B (breach) and NB (no breach). The probability of an incident depends on the type of the policyholder.

Nature moves first by randomly choosing the policyholder’s type according to a known *a priori* distribution: \mathcal{P}_S with probability $\Pr(\mathcal{P}_S) = \varphi$ and \mathcal{P}_N with probability $\Pr(\mathcal{P}_N) = 1 - \varphi$, $\varphi \in [0, 1]$. The type is private to a policyholder and the insurer knows only the probability distribution over the different types. Hence, the game is of incomplete information. Regardless of the types, the policyholder’s actions are CD (claim premium discount) and NC (no discount claim). Nature then decides the occurrence of the breach on a policyholder, followed by the insurer’s decision to audit (A) or not audit (NA) only in the event of a breach. We assume that in CIAG, an audit investigates the misrepresentation of the cyber security investment and the claim for receiving a premium discount. In particular, it investigates whether the policyholder had indeed invested in cyber security countermeasures before claiming this discount. Our model does not assume that there is a particular type of audit.

Having described the players and actions, in the following we present the interaction between \mathcal{P} and \mathcal{I} . First, \mathcal{P} has signed up for a cyber insurance contract by paying a pre-

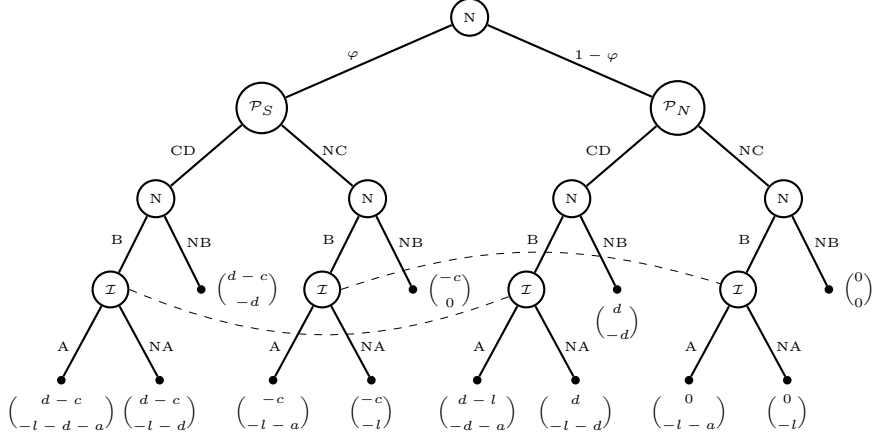


Figure 1: Extensive form representation of the Cyber Insurance Audit Game (CIAG) with the Nature deciding the types of policyholder and the occurrence of an incident.

mium p . The type of \mathcal{P} is decided by the nature based on an additional security investment. We assume that this investment equals c . This investment will decrease the probability of \mathcal{P} being compromised from β to β^* .

At the same time the investment will enable \mathcal{P} to claim a premium discount d . We assume that \mathcal{I} offers d without performing any audit since investigating at this point would mean that \mathcal{I} would have to audit policyholders who might never file an indemnity claim, thereby incurring avoidable losses.

We further assume that if \mathcal{P} decides to claim a discount without making the security investment, she will still receive d but risks having a future claim denied after an audit. After an incident, where \mathcal{P} suffers loss l , insurer \mathcal{I} has to decide whether to conduct an audit (e.g. forensics) to investigate details of the incident including the security level of \mathcal{P} at the time of breach. We assume that this audit costs a to the insurer. This audit will result in:

- Case 1: confirming that \mathcal{P} has indeed invested in security as claimed, in which case \mathcal{I} will pay the indemnity. We assume full coverage so the indemnity payment equals l .
- Case 2: discovering that \mathcal{P} has misrepresented her security level, \mathcal{I} refuses to pay the indemnity and \mathcal{P} has to bear the incident cost l . We assume that this case falls within the contract period during which \mathcal{P} is locked-in by the contract. We define *misrepresentation* as when \mathcal{P} is fraudulent or simply careless in maintaining the prescribed security level in the insurance contract and reports a fabricated security level to get the premium discount.

In Figure 1, some decision nodes of \mathcal{I} are connected through dotted lines indicating that the \mathcal{I} cannot distinguish between the connected nodes due to unknown \mathcal{P} type. These sets of decision nodes define the *information sets* of the insurer. An information set is a set of one or more decision nodes of a player that determines the possible subse-

quent moves of the player conditioning on what the player has observed so far in the game. The insurer also has two information sets, one where the breach has occurred to the policyholder who has claimed premium discount $CD = \{(CD|\mathcal{P}_S), (CD|\mathcal{P}_N)\}$ and the one where the breach has occurred to the policyholder who has not claimed premium discount $NC = \{(NC|\mathcal{P}_S), (NC|\mathcal{P}_N)\}$. Each of the insurer's information sets has two separate nodes since the insurer does not know the real type of the policyholder when deciding on whether to audit or not.

In outcome CD, B, A , the expected utility of the \mathcal{P}_S is

$$U_{CD,B,A}^{P_S} = U(W - p + d - c),$$

where U is a utility function, which we assume to be monotonically increasing and concave, W is the policyholder's initial wealth, p is the premium paid to the insurer, d is the premium discount, and c is the cost of the security investment. We assume the utility function to be concave to model the risk aversion of policyholders as defined in [12]. Note that we assume that $W > p > d$ and $W - p + d > c$, and both W and p are exogenous to our model.

$$U_{CD,B,A}^{P_S} = U_{CD,B,NA}^{P_S} = U_{CD,NB}^{P_S} = U(W - p + d - c) \quad (1)$$

$$U_{NC,B,A}^{P_S} = U_{NC,B,NA}^{P_S} = U_{NC,NB}^{P_S} = U(W - p - c) \quad (2)$$

$$U_{CD,B,A}^{P_N} = U(W - p + d - l) \quad (3)$$

$$U_{CD,B,NA}^{P_N} = U_{CD,NB}^{P_N} = U(W - p + d) \quad (4)$$

$$U_{NC,B,A}^{P_N} = U_{NC,B,NA}^{P_N} = U_{NC,NB}^{P_N} = U(W - p) \quad (5)$$

We further assume that the policyholder's goal is to maximize her expected utility. The expected utility of the policyholder is influenced by the possibility of a breach and the insurer's probability to audit. In particular, the expected utility for \mathcal{P}_S will be the same regardless of the insurer's probability to audit and the breach probability due to indemnification. \mathcal{P}_N , however, will need to consider these probabilities.

In the outcome $\mathcal{P}_S, \text{CD}, \text{B}, \text{A}$ the insurer's utility is

$$\mathcal{U}_{\mathcal{P}_S, \text{CD}, \text{B}, \text{A}}^{\mathcal{I}} = p - l - d - a,$$

where p is the premium, d is the premium discount offered, l is the loss claimed by the policyholder, and a is the audit cost.

In other outcomes, the insurer's utility is as follows:

$$\mathcal{U}_{\mathcal{P}_S, \text{CD}, \text{B}, \text{A}}^{\mathcal{I}} = p - l - d - a \quad (6)$$

$$\mathcal{U}_{\mathcal{P}_S, \text{CD}, \text{B}, \text{NA}}^{\mathcal{I}} = \mathcal{U}_{\mathcal{P}_N, \text{CD}, \text{B}, \text{NA}}^{\mathcal{I}} = p - l - d \quad (7)$$

$$\mathcal{U}_{\mathcal{P}_S, \text{CD}, \text{NB}}^{\mathcal{I}} = \mathcal{U}_{\mathcal{P}_N, \text{CD}, \text{NB}}^{\mathcal{I}} = p - d \quad (8)$$

$$\mathcal{U}_{\mathcal{P}_S, \text{NC}, \text{B}, \text{A}}^{\mathcal{I}} = \mathcal{U}_{\mathcal{P}_N, \text{NC}, \text{B}, \text{A}}^{\mathcal{I}} = p - l - a \quad (9)$$

$$\mathcal{U}_{\mathcal{P}_S, \text{NC}, \text{B}, \text{NA}}^{\mathcal{I}} = \mathcal{U}_{\mathcal{P}_N, \text{NC}, \text{B}, \text{NA}}^{\mathcal{I}} = p - l \quad (10)$$

$$\mathcal{U}_{\mathcal{P}_S, \text{NC}, \text{NB}}^{\mathcal{I}} = \mathcal{U}_{\mathcal{P}_N, \text{NC}, \text{NB}}^{\mathcal{I}} = p \quad (11)$$

$$\mathcal{U}_{\mathcal{P}_N, \text{CD}, \text{B}, \text{A}}^{\mathcal{I}} = p - d - a \quad (12)$$

4. Decision Analysis

In this section, we analyze the equilibria of the proposed Cyber Insurance Audit Game (Figure 1), which is a dynamic Bayesian game with incomplete information. The analysis is conducted using the game-theoretic concept of *Perfect Bayesian Equilibrium* (PBE). This provides insights into the strategic behaviour of the policyholder \mathcal{P} concerning discount claims and the insurer \mathcal{I} 's auditing decision.

A PBE, in the context of our game, can be defined by Bayes requirements discussed in [30]:

Requirement 1: The player at the time of play must have a belief about which node of the information set has been reached in the game. The beliefs must be calculated using Bayes' rule, whenever possible, ensuring that they are consistent throughout the analysis.

Requirement 2: Given these beliefs, a player's strategy must be sequentially rational. A strategy profile is said to be sequentially rational if and only if the action taken by the player with the move is optimal against the strategies played by all other opponents given the player's belief at that information set.

Requirement 3: The player must update her beliefs at the PBE to remove any implausible equilibria. These beliefs are determined by Bayes' rule and players' equilibrium strategies.

In the event of a security breach, the insurer's decision to audit or not must be based on beliefs regarding the policyholder's types. More specifically, a belief is defined as a probability distribution over the nodes within the insurer's information set, conditioned that the information node has

been reached. The insurer has two information sets subjected to whether the policyholder has claimed premium discount or not which are $\text{CD} = \{(\text{CD}|\mathcal{P}_S), (\text{CD}|\mathcal{P}_N)\}$ and $\text{NC} = \{(\text{NC}|\mathcal{P}_S), (\text{NC}|\mathcal{P}_N)\}$. The insurer assigns a belief to each of these information sets. Let μ and λ be the insurer's beliefs where

$$\mu = \Pr(\mathcal{P}_S|\text{CD})$$

$$\lambda = \Pr(\mathcal{P}_S|\text{NC})$$

That is, for the first information set, the insurer believes with μ and $1 - \mu$ that the premium discount claim is from \mathcal{P}_S and \mathcal{P}_N , respectively. Similarly, for the second information set, the insurer believes with λ that \mathcal{P}_S has not claimed premium discount and believes with $1 - \lambda$ that \mathcal{P}_N has not claimed premium discount.

The first requirement of PBE dictates that Bayes' rule should be used to determine beliefs. Thus

$$\mu = \frac{\Pr(\mathcal{P}_S) \Pr(\text{CD}|\mathcal{P}_S)}{\Pr(\mathcal{P}_S) \Pr(\text{CD}|\mathcal{P}_S) + \Pr(\mathcal{P}_N) \Pr(\text{CD}|\mathcal{P}_N)} \quad (13)$$

$$\lambda = \frac{\Pr(\mathcal{P}_S) \Pr(\text{NC}|\mathcal{P}_S)}{\Pr(\mathcal{P}_S) \Pr(\text{NC}|\mathcal{P}_S) + \Pr(\mathcal{P}_N) \Pr(\text{NC}|\mathcal{P}_N)} \quad (14)$$

From the payoffs in Figure 1, it can be clearly seen that CD is always a preferred choice for \mathcal{P}_S . Whereas, the insurer always gets a better payoff for choosing NA against NC irrespective of policyholder's type. Having defined the necessary concepts, next, we identify the possible PBEs of the game for the following constraints

$$l > a \text{ and } l > d \quad (15)$$

$$l > a \text{ and } l < d \quad (16)$$

$$l < a \text{ and } l > d \quad (17)$$

$$l < a \text{ and } l < d \quad (18)$$

where the PBEs are strategy profiles and beliefs that satisfies all the three requirements described earlier.

Theorem 1. For $\varphi > \frac{l-a}{l}$, $l > a$ and $l > d$, CIAG has only one pure-strategy PBE $((\text{CD}, \text{CD}), (\text{NA}, \text{NA}))$, in which the policyholder claims premium discount regardless of her type while the insurer does not audit regardless of whether the policyholder claims or not a discount, with $\mu = \varphi$ and arbitrary $\lambda \in [0, 1]$.

Proof. The existence of pure-strategy PBE can be verified by examining the strategy profile (CD, CD) and (NA, NA) with constraint in Equation (15). This represents the case where an incident has occurred on the policyholders who have claimed premium discount.

- a) *Belief consistency:* Due to information asymmetry and as only one of the insurer's information set is in the equilibrium path, she assigns $\Pr(\text{CD}|\mathcal{P}_S) =$

1 and $\Pr(\text{CD}|\mathcal{P}_N) = 1$. Thus, using Bayes' rule in Equation (13) gives

$$\mu = \varphi/(\varphi + 1 - \varphi) = \varphi$$

On the other hand, applying Bayes' rule in Equation (14) to λ yields 0/0 which is an indeterminate result. This implies that if the equilibrium is actually played then the off-equilibrium information set NC should not be reached restricting an update to the insurer's belief with Bayes' rule. Due to an indeterminate result, the insurer specifies an arbitrary $\lambda \in [0, 1]$.

- b) *Insurer's sequentially rational condition given updated beliefs*: The expected payoff for each action of the insurer are

$$\begin{aligned} \mathcal{U}_A &= \varphi \cdot \mathcal{U}_{\mathcal{P}_S, \text{CD}, \text{B}, \text{A}}^I + (1 - \varphi) \cdot \mathcal{U}_{\mathcal{P}_N, \text{CD}, \text{B}, \text{A}}^I \quad (19) \\ &= \varphi(p - d - l - a) + (1 - \varphi)(p - d - a) \\ &= p - \varphi l - d - a \end{aligned}$$

$$\begin{aligned} \mathcal{U}_{\text{NA}} &= \varphi \cdot \mathcal{U}_{\mathcal{P}_S, \text{CD}, \text{B}, \text{NA}}^I + (1 - \varphi) \cdot \mathcal{U}_{\mathcal{P}_N, \text{CD}, \text{B}, \text{NA}}^I \quad (20) \\ &= \varphi(p - d - l) + (1 - \varphi)(p - d - l) \\ &= p - d - l \end{aligned}$$

The condition for A to be sequentially rational is $\mathcal{U}_A > \mathcal{U}_{\text{NA}}$ which gives

$$\begin{aligned} p - \varphi l - d - a &> p - d - l \\ \varphi &\leq \frac{l - a}{l} = \varphi^* \quad (21) \end{aligned}$$

Now considering the off-equilibrium information set NC, the insurer always gets a better payoff by choosing NA. Thus, NA is a dominant strategy of the insurer against the off-equilibrium information set NC. The insurer's belief λ remains arbitrary.

- c) *Policyholder's sequentially rational condition given insurer's best response*: Knowing the best responses of the insurer i.e. (A,NA) for $\varphi \leq \varphi^*$ and (NA,NA) for $\varphi > \varphi^*$ against CD, we derive the best response of the policyholder. For insurer's strategy profile (NA,NA), \mathcal{P}_S gets a payoff $U(W - p + d - c)$ by choosing CD. If she deviates to NC, she will get a payoff $U(W - p - c)$ which is undesirable. Whereas, \mathcal{P}_N receives a payoff $U(W - p + d)$ by choosing CD. If she deviates to NC will get a payoff $U(W - p)$ which is also undesirable. Thus, (CD,CD) and (NA,NA) can be verified as a PBE given $\varphi > \varphi^*$ and $\mu = \varphi$. Note that the PBE includes the updated beliefs of the insurer implicitly satisfying Requirement 3.

□

From the PBE, we can see that if $l > a$, $l > d$ and insurer's belief φ is greater than the threshold value φ^* , not auditing a breach is optimal for the insurer and claiming premium discount is optimal for the policyholder regardless of her type. When the insurer's belief $\varphi \leq \varphi^*$ there exist no pure-strategy PBE. As a result, both players will mix up their strategies. We discuss this mixed-strategy PBE below. Note that, in the following, we use the inner tuple $(x, 1 - x)$ to indicate a mixed strategy where the player chooses the first action with probability x and the second action with probability $1 - x$.

Theorem 2. For $\varphi \leq \frac{l-a}{l}$, $l > a$, $l > d$, CIAG has only one mixed-strategy PBE, in which:

- \mathcal{P}_S will always prefer CD, while \mathcal{P}_N randomizes between CD and NC with probability δ and $1 - \delta$, respectively;
- the insurer randomizes between A and NA with probability θ and $1 - \theta$, respectively, against CD, and she always prefer NA against NC, with her beliefs about \mathcal{P}_S playing CD and NC being $\bar{\mu} = \frac{\varphi}{\varphi + (1-\varphi)\delta}$ and $\bar{\lambda} = 0$, respectively, where

$$\begin{aligned} \delta &= \frac{a}{(1 - \varphi)l} \\ \theta &= \frac{U(W - p + d) - U(W - p)}{\beta \cdot (U(W - p + d) - U(W - p + d - l))} \quad (22) \end{aligned}$$

Proof. The existence of mixed-strategy PBE is outlined below.

- a) *Belief consistency*: Again we apply the Bayes' rule. By assuming that the policyholder sticks to the equilibrium strategy, the insurer can derive that $\Pr(\text{CD}|\mathcal{P}_S) = 1$, $\Pr(\text{NC}|\mathcal{P}_S) = 0$, $\Pr(\mathcal{P}_S) = \varphi$, $\Pr(\mathcal{P}_N) = 1 - \varphi$, $\Pr(\text{CD}|\mathcal{P}_N) = \delta$ and $\Pr(\text{NC}|\mathcal{P}_N) = 1 - \delta$. Using Equations (13) and (14) we obtain $\mu = \bar{\mu}$ and $\lambda = \bar{\lambda}$ in Equation (2).
- b) *Optimal responses given beliefs and opponent's strategy*: Given these beliefs and the mixed strategy of the policyholder, an insurer's optimal strategy would maximize her payoff. The insurer can achieve this by randomizing her actions such that the expected payoffs is equal for all the actions of the policyholder. This is known as the *indifference principle* in game theory. Thus, the expected utility of \mathcal{P}_N for choosing CD is

$$\begin{aligned}
\mathcal{U}_{CD}^{P_N} &= \beta \cdot \left(\theta \cdot \mathcal{U}_{CD,B,A}^{P_N} + (1 - \theta) \cdot \mathcal{U}_{CD,B,NA}^{P_N} \right) \\
&\quad + (1 - \beta) \cdot \mathcal{U}_{CD,NB}^{P_N} \\
&= \beta \cdot \theta \cdot U(W - p + d - l) \\
&\quad + \beta \cdot (1 - \theta) \cdot U(W - p + d) \\
&\quad + (1 - \beta) \cdot U(W - p + d) \\
&= \beta \cdot \theta \cdot \left(U(W - p + d - l) - U(W - p + d) \right) \\
&\quad + U(W - p + d) \tag{23}
\end{aligned}$$

and for choosing NC, where NA is a dominating strategy of the insurer, is

$$\begin{aligned}
\mathcal{U}_{NC}^{P_N} &= \beta \cdot \mathcal{U}_{NC,B,NA}^{P_N} + (1 - \beta) \cdot \mathcal{U}_{NC,NB}^{P_N} \\
&= \beta \cdot U(W - p) + (1 - \beta) \cdot U(W - p) \\
&= U(W - p) \tag{24}
\end{aligned}$$

The indifference principle requires that $\mathcal{U}_{NC}^{P_N} = \mathcal{U}_{CD}^{P_N}$, which gives

$$\begin{aligned}
U(W - p) &= \beta \cdot \theta \cdot \left(U(W - p + d - l) - U(W - p + d) \right) \\
&\quad + U(W - p + d) \\
\theta &= \frac{U(W - p + d) - U(W - p)}{\beta \cdot \left(U(W - p + d) - U(W - p + d - l) \right)}
\end{aligned}$$

as in Equation (22). Similarly, the policyholder will also mix her strategy with an aim to make the insurer indifferent between choosing A and NA. Thus,

$$\begin{aligned}
\mathcal{U}_A^I &= \varphi \cdot \mathcal{U}_{P_S,CD,B,A}^I \\
&\quad + (1 - \varphi) \cdot \left(\mathcal{U}_{P_N,CD,B,A}^I + \mathcal{U}_{P_N,NC,B,A}^I \right) \\
&= \varphi \left((1)(p - d - l - a) + (0)(p - l - a) \right) \\
&\quad + (1 - \varphi) \left(\delta(p - d - a) + (1 - \delta)(p - l - a) \right) \\
&= p - l - a - \varphi d - \delta d + \delta l + \varphi \delta d - \varphi \delta l \tag{25}
\end{aligned}$$

$$\begin{aligned}
\mathcal{U}_{NA} &= \varphi \cdot \mathcal{U}_{P_S,CD,B,NA}^I \\
&\quad + (1 - \varphi) \cdot \left(\mathcal{U}_{P_N,CD,B,NA}^I + \mathcal{U}_{P_N,NC,B,NA}^I \right) \\
&= \varphi \left((1)(p - d - l) + (0)(p - l) \right) \\
&\quad + (1 - \varphi) \left(\delta(p - d - l) + (1 - \delta)(p - l) \right) \\
&= p - l - \varphi d - \delta d + \varphi \delta d \tag{26}
\end{aligned}$$

and $\mathcal{U}_A = \mathcal{U}_{NA}$ gives

$$\begin{aligned}
p - l - a - \varphi d - \delta d + \delta l + \varphi \delta d - \varphi \delta l &= p - l - \varphi d \\
&\quad - \delta d + \varphi \delta d \\
\delta &= \frac{a}{(1 - \varphi)l}
\end{aligned}$$

as in Equation (22). \square

We conceive all the possible PBEs for CIAG by exhaustively applying this methodology over all combinations of the players' strategy profiles for the four constraints described in Equations (15) to (18). Figure 2 presents the solution space of CIAG. It further shows how the equilibrium strategies of the players depends on the premium discount (d), audit cost (a), and loss (l).

5. Model Evaluation

Our analysis in Section 4 provides a framework for insurers to determine optimal auditing strategy against policyholders who can misrepresent their security levels to avail premium discounts. This section illustrates the methodology used to obtain values for various parameters of our model and simulation results using these values to determine the best strategy for the insurer.

5.1. Methodology and Data Collection

A diverse set of data sources is needed to study the interaction between insurance pricing, the effectiveness of security controls, and the cost of auditing claims. To this end, we combine the following data sources: a US law requiring insurers to report pricing algorithms [6], analysis of a data set of over 12,000 cyber events [2], a study of the cost and effectiveness of security controls [31], and a range of informal estimates regarding the cost of an information security audit.

The model assumes that nature determines incidents according to a Bernoulli distribution with loss amount l and probability of loss β . Analysis of the data set of 12,000 cyber incidents reveals data breach incidents occur with a median loss \$170K and frequency of around 0.015 for information firms [2], which we use as l and β respectively.

We adopt the security control model used in [31]. Both fixed and operational costs are estimated using industry reports, which correspond to c in our model. The effectiveness of a control is represented as a percentage decrease in the size or frequency of losses. For example, operating a firewall (\$2,960) is said to reduce losses by 80% [31]—leading to a probability of breach after investment (β^*) of 0.2β .

We downloaded all of the cyber insurance filings in the state of California and discarded off-the-shelf policies that do not change the price based on revenue, industry or security controls. This left 26 different pricing algorithms and corresponding rate tables, the contents of which are described in [6].

Data breach coverage with a \$1 million limit was selected because it is the default coverage and it comfortably covers

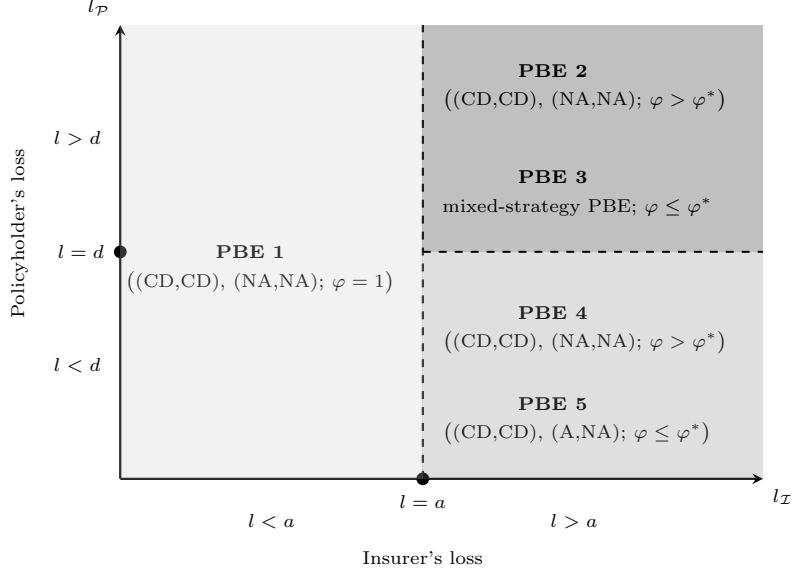


Figure 2: Solution space of Cyber Insurance Audit Game (CIAG).

the loss value l for a data breach on SMEs. The premium p and discount d varies based on the insurer. We chose a filing explicitly mentioning discounts for firewalls. For an information firm with \$40M of revenue, the premium p is equal to \$3,630 and the filings provide a range of discounts up to 25%. The exact value depends on an underwriter's subjective judgment. To comprise this we consider multiple discounts in this range.

Estimating the insurer's cost of audit (a) is difficult because they could be conducted by loss adjusters within the firm or contracted out to IT specialists. With the latter in mind, we explored the cost of an information security audit. The cost depends on the depth of the assessment and the expertise of the assessor. However, collating the quoted figures suggests a range from \$5,000 up to \$100,000.

5.2. Numerical Analysis

We simulate the interaction between the cyber insurance policyholder and the insurer based on our game-theoretic model with parameter values described above. First, we compare the expected payoffs of the insurer for different strategic models:

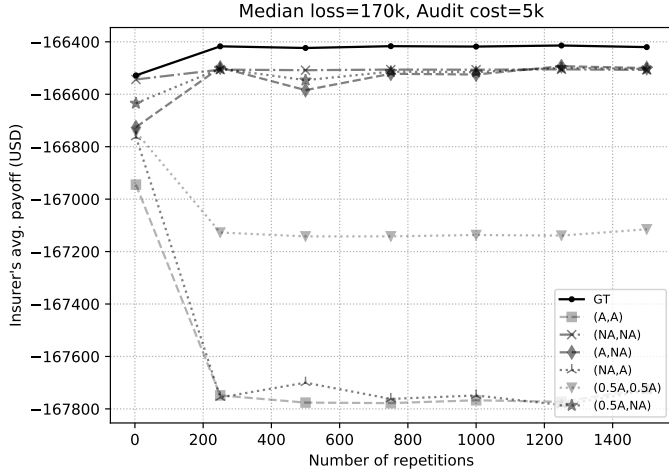
1. the game-theoretic approach (GT) where the insurer chooses an appropriate strategy according to our analysis (refer to Figure 2) and can either audit or not audit;
2. always auditing (A,A) regardless of whether the policyholder has claimed discount or not;
3. always not auditing (NA,NA) regardless of whether the policyholder has claimed discount or not;
4. auditing if the policyholder has claimed discount and not auditing if there is no discount claimed (A,NA);

5. not auditing if the policyholder has claimed discount and auditing if there is no discount claimed (NA,A);
6. auditing half the times regardless of whether the policyholder has claimed discount or not (0.5A,0.5A);
7. auditing half the times when the policyholder has claimed discount and not auditing if there is no discount claimed (0.5A,NA).

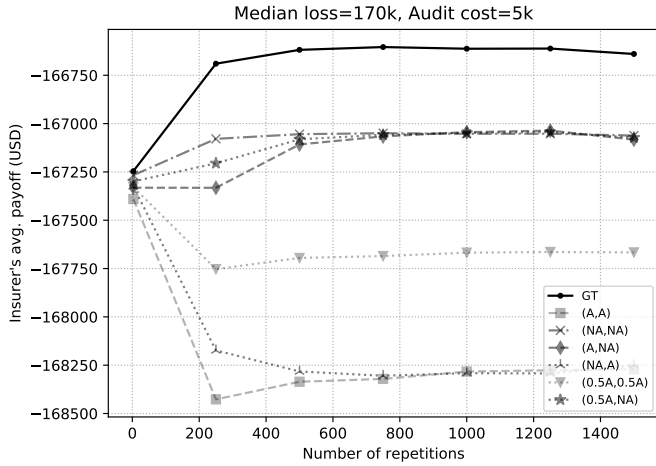
In the following simulation figures, the insurer's average payoffs with each strategic model are calculated against a policyholder who plays the PBE strategy obtained through our analysis. This policyholder is also the most challenging one for the insurer as it claims for a discount even in the case of non investment. The term " x repetitions of the game" reflects that CIAG is played x number of independent runs for a set of parameter values.

From Figures 3a and 3b we observe that the payoff of the insurer when choosing the GT model is always better than rest of the strategic models irrespective of the premium discount. The reason for this is that the model (A,NA), where the insurer audits only policyholders who have claimed the discount, is susceptible to auditing clients who have implemented additional security level bearing the auditing cost as a pure loss. Thus, the larger the number of honest policyholders, the higher the insurer's loss is. Additionally, the insurer's loss as expected increases with the increasing cost of audit. With the (NA,NA) model, the insurer chooses to reimburse the loss without confirming the policyholder's actual security level. Here, the insurer indemnifies even for cases where the policyholder has misrepresented her security level suffering heavy losses. Another non strategic approach would be to randomize over the choice of auditing or not auditing a policyholder who has claimed a premium discount. This strategy, represented by the model (0.5A,NA), gives a payoff within the

range of payoffs from models (A,NA) and (NA,NA). The results exhibit that models (A,A), (NA,A), and (0.5A,0.5A) consistently performs poorly compared to other models.



(a) Premium discount 5%

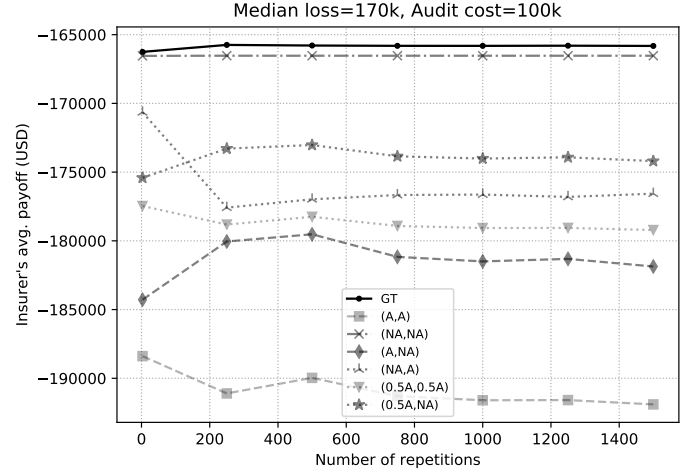


(b) Premium discount 25%

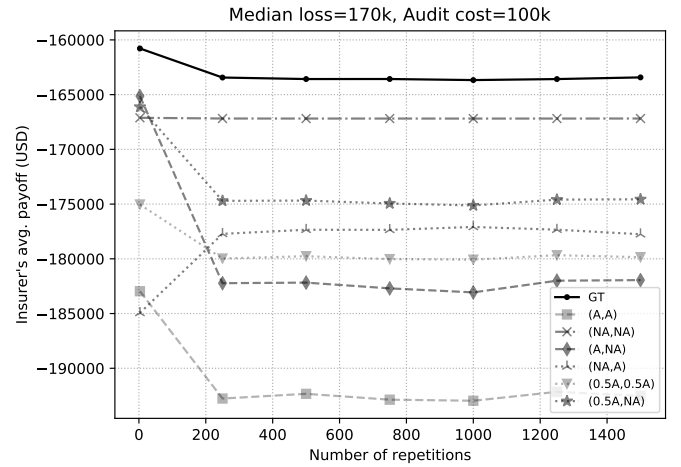
Figure 3: Insurer's average payoff (in US dollars) with different strategic models across various repetitions of the game for a median loss \$170k, audit cost \$5k with (a) premium discount 5% and (b) premium discount 25%.

The GT model presents an optimal mix of (A,NA) and (NA,NA) where the insurer's decision to audit is based on a *prior* belief regarding the policyholder's security investment. For the median loss of \$170k which is greater than both the audit cost and premium discount, the game solution is derived from the upper-right section of the solution space in Figure 2. In particular, when the insurer's belief (ϕ) regarding the policyholder's security investment is greater than a threshold (ϕ^*), she prefers (NA,NA) i.e., **PBE 2**: ((CD,CD),(NA,NA); $\phi > \phi^*$). When the belief is lower than ϕ^* , she prefers a mixed approach (**PBE 3**) by simultaneously relying on (A,NA) and (NA,NA) and choosing whichever is more profitable. The GT model,

thus, enables the insurer to take into account a prior belief regarding the policyholder's security investment under the condition of information asymmetry and maximize her payoffs given this belief. The figures further show that regardless of how many times the game has been played model GT performs better than the non-game-theoretic models.



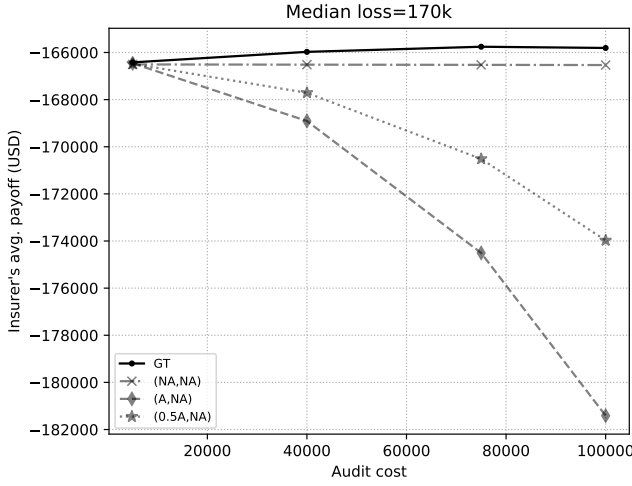
(a) Premium discount 5%



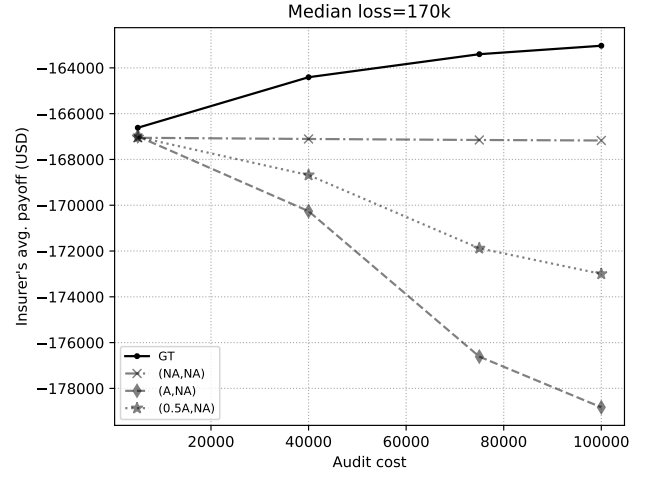
(b) Premium discount 25%

Figure 4: Insurer's average payoff (in US dollars) with different strategic models across various repetitions of the game for a median loss \$170k, audit cost \$100k with (a) premium discount 5% and (b) premium discount 25%.

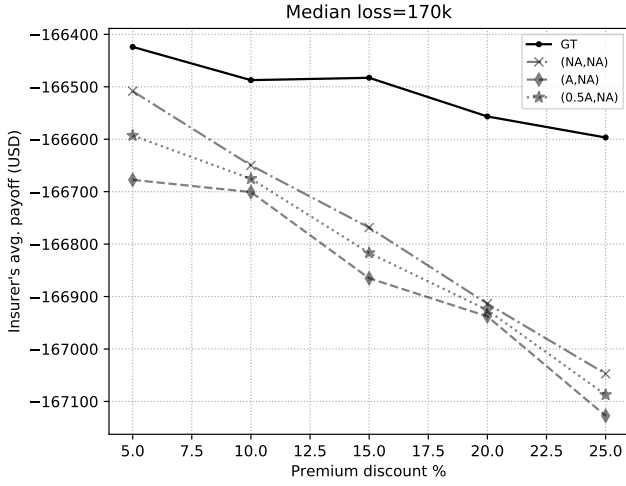
With higher audit cost i.e., \$100k in Figures 4a and 4b, we observe that the insurer's average payoff with the model (A,NA) decreases drastically confirming its shortcomings as discussed above. In the case of 1500 independent repetitions for the highest values of audit cost and premium discount, the insurer gains, on average, a higher payoff when choosing GT as opposed to (NA,NA) model. The increased difference in the payoff is equivalent to 98% of the annual premium charged to policyholder.



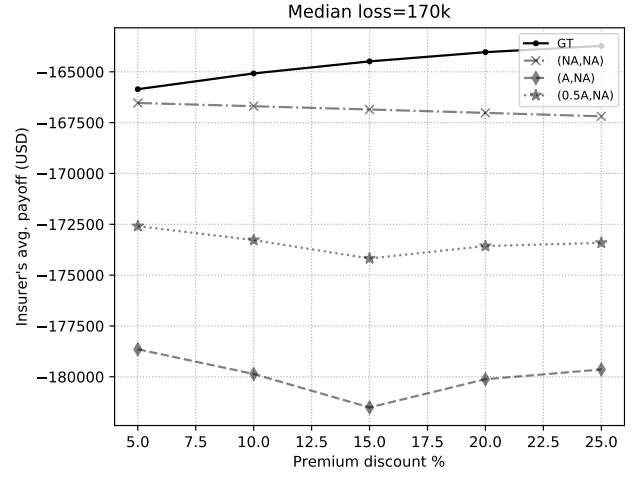
(a) Premium discount 5%



(b) Premium discount 25%



(c) Audit cost 5k



(d) Audit cost 100k

Figure 5: Insurer's average payoff (in US dollars) with different strategic models for median loss \$170k against audit costs with premium discount (a) 5% and (b) 25% and against premium discount with audit cost (c) 5k and (d) 25k.

Remark 1: For constant loss, as premium discount increases, GT consistently outperforms all other strategic models for various repetitions of the game in both sets of experiments with minimum and maximum values of audit cost.

Next, the simulation results are obtained over 100 repetitions with a median loss of \$170k against a range of audit cost, premium discount and loss. Note that the models (A,A), (NA,A), and (0.5A,0.5A) are omitted from the figures as they perform worse than others, and for ease of presentation.

Figures 5a and 5b show that there is a point of convergence where the strategy largely doesn't matter, but then as the audit cost increases, there is motivation for playing the game-theoretic solution as any other solution is worse. As discount increases, a policyholder might be highly stim-

ulated to receive premium discount given that the insurer will grant this without auditing her before an incident occurs. This escalates the possibilities of the policyholder misrepresenting her actual security level. Given this possibility, GT noticeably dominates other strategic models as seen in Figures 5c and 5d. Further, in the case of 100 independent repetitions with the highest values of premium discount and audit cost, deploying GT gives the insurer on average a higher payoff compared to the next best model which is (NA,NA). The increased difference in the payoff is equivalent to 60% of the annual premium charged to the policyholder.

Remark 2: For a constant loss, as premium discount and audit cost increase, GT outperforms all other strategic models.

Figure 6 shows that there is essentially nothing special

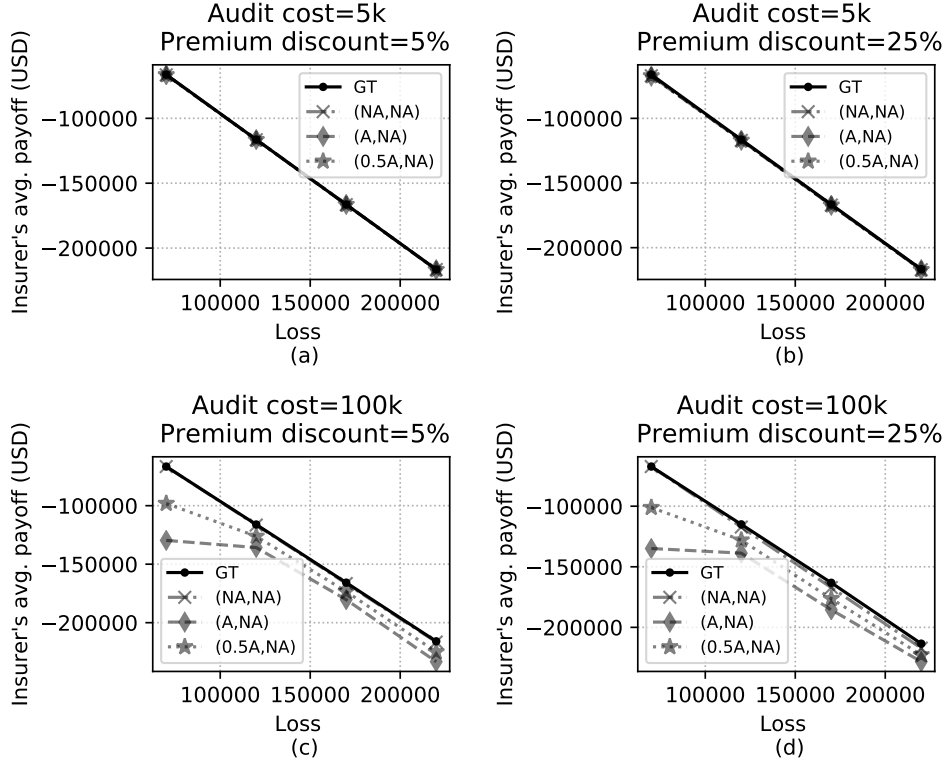


Figure 6: Insurer’s average payoff (in US dollars) with different strategic models against loss for various audit costs and premium discounts.

about the loss as a contributing factor with low audit cost, but become discriminatory as the audit cost approaches the loss. GT and (NA,NA) performs equally well until this condition, but as the discount increases with the audit cost, GT exceeds (NA,NA). In this case, for 100 independent repetitions, insurers gain on average a higher payoff with model GT, compared to model (NA,NA), the next best. The increased difference in payoff is equivalent to 66% of the annual premium charged to the policyholder.

Remark 3: As premium discount, audit cost, and loss increase, GT consistently outperforms all other strategic models.

In summary, we have demonstrated how an insurer may use our framework in practice to determine the best auditing strategy against a policyholder. We have illustrated how the insurer’s payoff is maximized by strategically choosing to audit or not in the event of a breach. Such strategic behaviour also allows the insurer to maximize her payoff against policyholders who can misrepresent their security levels to avail premium discount.

6. Conclusion

Speaking to cyber insurance providers reveals concerns about the discrepancy between the security policies applicants report that they follow, in the application process, and the applicant’s compliance with these policies once

coverage is in place. To address this, we developed a game-theoretic framework investigating audits as a mechanism to disincentivize misrepresentation of security level by policyholders. Thus far, we know of one instance [10] denying cyber insurance coverage due to non-compliance with the security practices as defined in the insurance contract. Although there could have been denials settled in private, this suggests that most cyber insurance providers follow the *never audit* strategy. Our analysis derived a game-theoretic strategy that outperforms naïve strategies, such as never audit. By considering the post-incident claims management process, we demonstrated how a cyber insurance market can avoid collapse (contradicting [26]) when the policyholder can fraudulently report their security level.

To extend this paper, future work could consider modelling uncertainty about the effectiveness of the implemented security measure. In the current model, the policyholder’s type is chosen by Nature according to some probability distribution. It could be extended such that the policyholder maximizes expected payoff by selecting an investment strategy based on the beliefs about her type. This consideration would extend, for example, our analysis to consider the overall utility function of the policyholder, that is considering both the investment and no investment types simultaneously, and maximizing the expected payoff.

Another interesting direction is investigating how the potential loss l changes as a function of the security investment. In this case, we will be looking into different types of

risk profiles of the policyholders. We could also investigate the trade-off between the additional investment, discount, and residual risk.

Finally, a future extension could make investment in security a strategic choice for the policyholder in a multi-round game with a *no claims bonus*, as our data set describes the size of these discounts. We could also allow belief updates to influence insurer choices on each iteration.

Acknowledgements

We thank the reviewers for their valuable feedback and comments.

Sakshyam Panda and Emmanouil Panaousis have been partially funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie SECONDO grant agreement No 823997.

Daniel Woods could participate thanks to a Fulbright Cyber Security Award.

References

- [1] R. J. Anderson, Security engineering: a guide to building dependable distributed systems, John Wiley & Sons, 2010.
- [2] S. Romanosky, Examining the costs and causes of cyber incidents, *Journal of Cybersecurity* 2 (2) (2016) 121–135.
- [3] A. Beauteament, M. A. Sasse, M. Wonham, The compliance budget: managing security behaviour in organisations, in: *Proceedings of the 2008 New Security Paradigms Workshop*, ACM, 2009, pp. 47–58.
- [4] T. Moore, On the harms arising from the equifax data breach of 2017, *International Journal of Critical Infrastructure Protection* 19 (C) (2017) 47–48.
- [5] D. W. Woods, I. Agrafiotis, J. R. Nurse, S. Creese, Mapping the coverage of security controls in cyber insurance proposal forms, *Journal of Internet Services and Applications* 8 (1) (2017) 8.
- [6] S. Romanosky, L. Ablon, A. Kuehn, T. Jones, Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk?, in: *Proceedings of The 16th Workshop on the Economics of Information Security (WEIS 2017)*, 2017.
- [7] U. Franke, The cyber insurance market in Sweden, *Computers & Security* 68 (2017) 130–144.
- [8] J. Kesan, R. Majuca, W. Yurcik, Cyberinsurance as a market-based solution to the problem of cybersecurity: a case study, in: *Proc. WEIS*, 2005, pp. 1–46.
- [9] R. Thoyts, *Insurance theory and practice*, Routledge, 2010.
- [10] *Complaint in columbia cas. co. v. cottage health sys.*, no. 2:16-cv-03759 (c.d. cal.), <https://www.insideprivacy.com/wp-content/uploads/sites/6/2016/06/CNA-v-Cottage-Health-2016-complaint.pdf> (2016).
- [11] D. W. Woods, A. C. Simpson, Policy measures and cyber insurance: A framework, *Journal of Cyber Policy* 2 (2) (2017) 209–226.
- [12] R. Böhme, G. Schwartz, et al., Modeling cyber-insurance: Towards a unifying framework., in: *WEIS*, 2010.
- [13] H. Kunreuther, G. Heal, Interdependent security, *Journal of risk and uncertainty* 26 (2-3) (2003) 231–249.
- [14] A. Laszka, M. Felegyhazi, L. Buttyan, A survey of interdependent information security games, *ACM Computing Surveys* 47 (2) (2015) 23:1–23:38.
- [15] H. Ogut, N. Menon, S. Raghunathan, Cyber insurance and IT security investment: Impact of interdependence risk., in: *Proceedings of The 4th Workshop on the Economics of Information Security (WEIS 2005)*, 2005.
- [16] J.-C. Bolot, M. Lelarge, A new perspective on internet security using insurance, in: *INFOCOM 2008. The 27th Conference on Computer Communications*, IEEE, 2008, pp. 1948–1956.
- [17] R. Böhme, G. Kataria, Models and measures for correlation in cyber-insurance., in: *Proceedings of The 5th Workshop on the Economics of Information Security (WEIS 2006)*, 2006.
- [18] W. S. Baer, A. Parkinson, Cyberinsurance in it security management, *IEEE Security & Privacy* 5 (3).
- [19] A. Laszka, B. Johnson, J. Grossklags, M. Felegyhazi, Estimating systematic risk in real-world networks, in: *Proceedings of the 18th International Conference on Financial Cryptography and Data Security (FC)*, 2014, pp. 417–435.
- [20] D. W. Woods, A. C. Simpson, Monte carlo methods to investigate how aggregated cyber insurance claims data impacts security investments, *Workshop on the Economics of Information Security*, 2018.
- [21] M. M. Khalili, M. Liu, S. Romanosky, Embracing and controlling risk dependency in cyber-insurance policy underwriting, in: *Proceedings of The 17th Workshop on the Economics of Information Security (WEIS 2018)*, 2018.
- [22] A. Laszka, J. Grossklags, Should cyber-insurance providers invest in software security?, in: *European Symposium on Research in Computer Security*, Springer, 2015, pp. 483–502.
- [23] N. Shetty, G. Schwartz, M. Felegyhazi, J. Walrand, Competitive cyber-insurance and internet security, in: *Economics of information security and privacy*, Springer, 2010, pp. 229–247.
- [24] R. P. Majuca, W. Yurcik, J. P. Kesan, The evolution of cyberinsurance, *arXiv preprint cs/0601020*.
- [25] A. Laszka, E. Panaousis, J. Grossklags, Cyber-insurance as a signaling game: Self-reporting and external security audits, in: *Proceedings of the 9th Conference on Decision and Game Theory for Security (GameSec 2018)*, Springer, 2018.
- [26] G. Schwartz, N. Shetty, J. Walrand, Why cyber-insurance contracts fail to reflect cyber-risks, in: *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IEEE, 2013, pp. 781–787.
- [27] T. Bandyopadhyay, V. S. Mookerjee, R. C. Rao, Why it managers don’t go for cyber-insurance products, *Communications of the ACM* 52 (11) (2009) 68–73.
- [28] P. Picard, *Economic analysis of insurance fraud*, in: *Handbook of Insurance*, Springer, 2013, pp. 349–395.
- [29] P. Picard, Auditing claims in the insurance market with fraud: The credibility issue, *Journal of Public Economics* 63 (1) (1996) 27–56.
- [30] R. Gibbons, *A primer in game theory*, Harvester Wheatsheaf, 1992.
- [31] C. D. Heitzenrater, A. C. Simpson, Policy, statistics and questions: Reflections on uk cyber security disclosures, *Journal of Cybersecurity* 2 (1) (2016) 43–56.