Insured Access: An Approach to Ad-hoc Information Sharing for Virtual Organizations

Naoki Tanaka[†,‡,∗], Marianne Winslett[†,∗], Adam J. Lee[◦], David K. Y. Yau[◇,∗], Feng Bao[‡]
† Department of Computer Science, University of Illinois at Urbana-Champaign
‡ Cryptography & Security Department, Institute for Infocomm Research
∗ Advanced Digital Sciences Center
◦ Department of Computer Science, University of Pittsburgh
◇ Department of Computer Science, Purdue University

ABSTRACT

A virtual organization (VO) is a group of organizations that have banded together to achieve a common goal. Often a VO could function more effectively if its members were willing to share certain information. However, a typical VO member will not want to share its own information because the member will not benefit directly from the information's reuse, yet will be blamed if the reuse turns out badly.

In this paper, we present *insured access*, the first economically sustainable system for encouraging appropriate information sharing in VOs. Before accessing information, a VO member must purchase a liability policy from the insurance arm of the VO. Insured access uses actuarial principles to set up and run the VO's insurance arm, and provides the following benefits: VO members who share their information are compensated if the information is misused, and can expect a *positive* benefit from sharing; members who use information well are rewarded and those who misuse it are penalized appropriately; and the level of risk-taking in the system is capped at a certain level. We demonstrate the sustainability of insured sharing through simulations of a map-sharing scenario.

Categories and Subject Descriptors: H.4 [Information Systems Applications]: Miscellaneous; D.4.6 [Operating Systems]: Security and Protection—Access controls

Keywords: Insured Access; Risk-Aware Authorization; Risk-Based Access Control; Actuarial Science

1. INTRODUCTION

A shared goal binds VO members together, such as when a consortium responds to a business opportunity, or agencies work together to respond to a flood and nuclear disaster. Any sufficiently large organization operates as a VO, because its internal divisions have their own vested interests that do not always align with the VO's best interests.

Often, to be successful in achieving the shared goal, VO members need to share information with each other. Information sharing usually requires a VO member to take information that it collected for its own internal purposes, and release it to another for a different purpose. Since the first member – the information *producer* – has set the access control policy to match its original internal use for the information, sharing requires policy changes. Further, the producer will usually be blamed if another member – the information *consumer*[1] – misuses the information, and will not directly benefit if the consumer makes good use of it. Thus, producers are often reluctant to share.

The misaligned incentives for sharing stem from traditional approaches to authorization, which try to *eliminate* risk for individual VO members, rather than *maximize VO productivity while bounding risk*. To fix this, researchers have proposed approaches where a VO allocates "risk tokens" to its members, which they can use to "pay" for risky accesses to information that would not otherwise be allowed. However, it is not clear why producers would want to participate in a risk-token-based economy, or whether the VO would really benefit. Token-based economies also face problems like hoarding and shortages.

In recent decades, the business community has benefited from the use of actuarial methods to manage many kinds of business risks, but information sharing has not been among them. In this paper, we address this gap by introducing *insured access*, an insurance-based approach to incentivize information providers to allow risky accesses that are likely to benefit the VO. Our contributions:

- We propose *sustainable* methods for a VO to determine the price of a particular information access and decide whether to grant a particular access request, given a bound ε on the risk that the VO is willing to tolerate.
- We show that information providers can expect to benefit from sharing and providers will have recourse when information sharing turns out badly for the provider.
- We demonstrate the use of insured access in a simulation, and show that the system behaves as predicted.

The rest of the paper is organized as follows. Section 2 discusses related work, and Section 3 presents the details of insured access. Section 4 presents experimental results, and Section 5 concludes the paper.

CODASPY'13, February 18–20, 2013, San Antonio, Texas, USA.

[1]A member may both produce and consume information.

2. RELATED WORK

To address the problem of encouraging appropriate sharing of sensitive defense-related information in a VO, the MITRE JASON report [1] proposed a *risk-based access control* approach. In their approach, principals use *risk tokens* to purchase access rights to data. The access price is the expected value of damages due to this access. The VO decides how much risk it can handle during the next fiscal period, creates that many tokens, and allocates them to its members. Risk-based access control does not effectively control the worst-case aggregate damages, and does not distinguish between good and bad risk-takers.

The JASON report inspired follow-on papers addressing specific aspects of a risk-based approach, such as how to integrate trust and risk into RBAC [4], consider relative security risks in RBAC [13], balance the risks and benefits [17], combine risk-based access control with fuzzy logic [3], and allocate risk tokens to VO members [12]. But none of these schemes provides reasons for a rational, self-interested VO member to volunteer to provide information. In addition, none of these schemes provides a clear bound on the worst-case damages, or ensures that members who use information badly are treated differently from those who use it well. With no incentive for consumers to try to use information well, they may expose the VO to unnecessary risk.

Token-based approaches face other challenges as well. VO members may hoard tokens in anticipation of future shortages, spend leftover tokens carelessly at the end of a fiscal period, or be unable to obtain tokens when they truly need them. A cash-based scheme will lessen these problems, assuming that VO members behave rationally.

Cash can be coupled with decision theory to decide how much an access might benefit an information consumer or the VO [11]. Decision theory could show whether the likely rewards associated with a particular access exceed its potential risks. In practice, though, a consumer may be unwilling to share detailed information about its planned use of the information, making it hard to do decision-theoretic analysis of risks and rewards at the VO level.

Recent work proposes ways to price access to personal data, such as online behavior and demographic characteristics, using differential privacy and auction mechanisms [7, 14]. Like insured access, these works propose pricing schemes for sensitive information. However, these works assume that upfront payments are sufficient to entice providers to participate. This is appropriate for the low-risk settings these works target, where an individual's aggregate income from sharing will almost surely exceed their aggregate damages. In contrast, insurance is appropriate for situations where there is a very small chance of very high damage.

Conversely, insured access could use auction-based data pricing to bring providers higher profits than the fee-for-service and profit-sharing schemes we propose. However, to reach a fair price, auctions need multiple potential bidders. Insured access is aimed at ad-hoc, non-routine information needs that are too atypical to be institutionalized into a VO's role-based access control system. Single-expected-bid auctions will default to the minimum allowed bid, which is a fee-for-service model. If multiple potential bidders are likely, then the information need is sufficiently routine for the VO to adopt lighter-weight methods than insured access.

Finally, insured access can benefit from ways to reduce the sensitivity of shared information. For example, if a consumer needs only part of a sensitive map, then removing the other parts may greatly reduce the expected damages and hence the price of insurance. A second technique is to generalize the shared information, e.g., releasing the approximate location of a gas pipeline rather than its exact location. Differential privacy is a third technique, but it is practical only when the consumer needs the result of a statistical analysis over many data items. Differential privacy is not effective when the consumer needs the data items themselves, such as a set of maps or phone numbers.

Hoo [9] indicates the potential of using actuarial methods to manage computer security risk, but no concrete scheme for access control was ever proposed. Insured access is the first complete, economically sustainable system for encouraging appropriate information sharing in VOs.

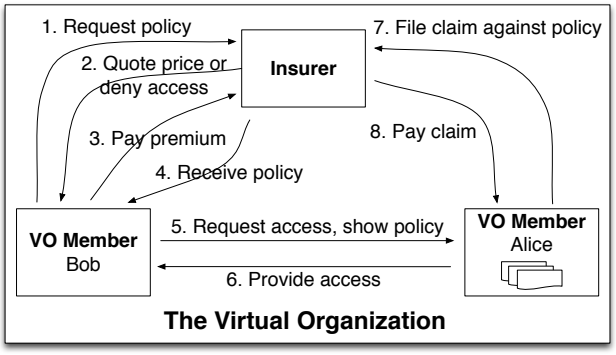


Figure 1: System Model for Insured Access

3. INSURED ACCESS

The VO can either use an external insurer or self-insure, but we consider only the case of self-insurance because the VO will have to expose very sensitive information to the insurer. We begin with a quick overview of insured access, then provide additional detail in the subsections that follow.

Before insured accesses begin, the VO members set up an internal VO insurance group (the *insurer*), agree on what kinds of information to share, and decide how much risk the insurer can assume. The VO members who become information consumers supply the startup capital for the insurer.

In Figure 1, a producer Alice has information that can be shared with others. To obtain access, consumer Bob asks the insurer for a policy covering the specific information he wants to access (1). The policy will insure Alice against damages she might incur because she shared that information with Bob. If the insurer is willing to issue the policy, it gives Bob a price (2), which Bob can choose to pay (3). We assume *consumers are rational, self-interested, and risk-adverse*, so Bob will only buy the policy if its price is less than the benefit he expects to gain from accessing the information. With the policy in hand (4), Bob asks Alice for access (5). Alice examines the policy and can give Bob the information (6). If Bob misuses the information and Alice suffers damages as a result, Alice can submit a claim to the insurer (7) and be reimbursed for her suffering (8). The policy can specify conditions of use, such as secure handling precautions or limits on the purpose of use. If Bob may have violated those rules, then the insurer can request reimbursement from Bob of the amount paid to Alice.

Insurance-based access control is a new idea in the security world and differs from the usual liability insurance, under which asset owners (i.e., providers) buy policies to protect themselves. With insured access, consumers pay for policies that protect providers. This means that consumers must consider potential damages when deciding whether to access information, as well as benefits; and providers need not consider damages when deciding whether to share. Because of this difference, existing actuarial methods cannot be directly applied to purchase decisions in insured access, and we propose appropriate formulas for this task.

The price Bob pays for the policy is called the *premium*, which depends on how risk-averse the insurer is, what information Bob wants to access, the history of past claims for that information, and the insurer's *premium principle* (pricing methodology). Based on Bob's track record, the insurer can reward Bob with lower premiums if Bob's other policies have no recent claims, and penalize him otherwise. The insurer can refuse to issue a policy to Bob because of his track record, or because the insurer needs to limit the risk associated with its current portfolio of policies.

If the VO members are very unlucky or the data used to set prices is insufficient, then the insurer might run out of funds to pay claims and be ruined. The insurer guarantees that at the moment it issues a policy, the new policy does not make the chance of eventual ruin cross a threshold ε agreed upon in advance by VO members. To estimate likely claim amounts for rare events that it has never seen, i.e., *tail events*, the insurer can use *extreme value theory*. The insurer can also use stop-loss policies and reinsurance to limit its exposure to tail risk, as discussed later.

If the insurer's chance of eventual ruin ever exceeds ε in spite of these precautions, the VO members have several options. They may choose to supply additional capital reserves. They may choose to take no action, so that the insurer runs out of money if its luck remains poor, necessitating internal loans, unpaid claims, or delayed claim payments. The insurer may raise premiums to increase capital. Or it may transfer some of its risk to a reinsurer.

Each VO member, and the VO as a whole, can expect to benefit from insured access. A consumer expects a net benefit from each insured access, as otherwise it would not buy the policy. Providers' losses attributable to sharing are reimbursed by the insurer's payments on their claims. So that providers directly benefit from sharing, an additional fee that goes directly to the provider can be included in each premium, to cover the provider's additional costs attributable to sharing, plus a profit. A second, complementary, and more conservative option is for the premium to include a fee retained by the insurer and intended as *future* profit for the provider. Then at the end of each fiscal period, the insurer's excess capital reserves can be shared with providers. If desired, the same two fee-based approaches can be used to cover the costs of running the insurer, or give it an expected profit; in this paper, we assume that the insurer is non-profit and cost-free. We also assume that the benefits and damages attributable to sharing accrue only to the provider and consumer, without impact on other VO members, though insured access can also be used in more complex situations with potential collateral benefits and damages. Finally, as discussed later, the VO must set up the system so that each act of sharing is expected to benefit the VO as a whole, i.e., the shared purpose that binds its members together.

3.1 Pricing

Consider a particular insured access. The insurer does not know in advance what the total size of all claims on that policy will be, but it can represent this quantity by a random variable X representing the *risk* associated with a particular consumer accessing a particular piece of information owned by a particular producer. More precisely, X represents the total amount of claims that will eventually be paid to the producer under that policy. For now, let us assume that the insurer knows the probability distribution for X, and discuss how this information can be used to set the price.

The most widely adopted approach to pricing risk in general is the *Principle of Equivalent Utility*, in which the premium P is calculated by solving the following equation (p. 4):[2]

$$E[u_I(w_I + P - X)] = u_I(w_I). \tag{1}$$

Here u_I is the utility function of the insurer I, and w_I is its current capital. This principle says that the premium P should be set to the amount at which the insurer is equally happy whether or not the policy is issued, i.e., indifferent.

The exact formula to calculate P depends on the utility function $u(z)$. Given $u(z)$, we can derive the *risk aversion index* $r(z) = -\frac{u''(z)}{u'(z)}$ of a principal. More risk averse principals have more concave utility functions.

With a linear utility function, $u(z) = z$ and $r(z) = 0$, meaning the principal is risk-neutral. Let π denote a premium pricing principle. If we use a linear utility function in the Principle of Equivalent Utility in Formula 1, we get $\pi[X] = E[X]$, which is called the *Net Premium*. With the Net Premium, the insurer sells a policy for the expected amount of its claims. In the long run, an insurer could break even with the Net Principle. However, in practice insurers usually prefer to set prices higher than the Net Premium, because it requires high capital reserves to avoid ruin.

With an exponential utility function $u(z) = 1 - e^{-\alpha z}$, for $\alpha > 0$, the risk aversion index $r(z) = \alpha$, and we derive the *Exponential Principle* from Formula 1:

$$\pi[X] = \frac{1}{\alpha} \log(m_X(\alpha)), \tag{2}$$

where $m_X(\alpha) = E[e^{\alpha X}]$ is the moment generating function of X around α, and X represents the risk (total claims) associated with a policy. We could add an additional fee to $\pi[X]$ to support the provider or insurer, as discussed previously.

Although there are pros and cons for each principle, the Exponential Principle is particularly widely used in actuarial literature [16]. Among its favorable properties, the following two are especially important:

Additivity for independent risks. $\pi[X+Y] = \pi[X] + \pi[Y]$, where X and Y are independent.

Superadditivity for positively correlated risks. $\pi[X+Y] \geq \pi[X] + \pi[Y]$, where X and Y are positively correlated.

For independent risks from separate acts of sharing, the additivity means that the price for one policy covering all of them is the same as the total price for separate policies for each. Thus, an insurer can price a new policy without having to analyze its aggregate risk across all its policies.

[2]All mentions of page numbers in this paper refer to [10].

Superadditivity for positively correlated risks is important because allowing several instances of information sharing can introduce a much larger risk than just the sum of each risk, if they are positively correlated. For example, military phone books are often classified, even if each number in them is unclassified. The Exponential Principle also enjoys subadditivity for negatively correlated risks, i.e., it reflects the fact that diversification can reduce overall risk. Because of these favorable properties, we use the Exponential Principle in the remainder of this paper.

To summarize, given an access request (risk) X, the insurer uses Formula 2 to compute a premium, tacking on a fee to go directly to the provider if desired.

3.2 Tail Events, Ruin, & Reinsurance

The insurer groups similar risks into a *class*, as discussed in detail later, such that all risks X in the class follow the same probability distribution, and assigns the same premium to all risks in the class. The previous section assumes that the insurer knows this distribution, but the insurer may only know its own history for the class, consisting of the details of every policy it issued and every claim it received. From this data, the insurer can produce an approximation to X's distribution. Often, X is known to belong to a particular family of distributions. In that case, the claims history can be used to estimate the parameters of the distributions, using maximum-likelihood estimation [2]. With a long enough history, one might expect a very good approximation.

In the real world, however, damage-causing events have an extremely long-tailed distribution, where the tail includes many highly unlikely catastrophic events. A claims history is a finite random sample from this long-tailed distribution. Thus, if an insurer prices premiums solely based on the history and without considering the unseen long tail, the insurer eventually face ruin (pp. 87-111). *Extreme value theory* can help by providing a basis for statistical modeling of unseen tail events [8]. Still, it only approximates the true risk distribution.

To handle the risk of high-damage events it has never observed, an insurer can buy a stop-loss insurance policy (pp. 8-13) from a reinsurer. The insurer pays claims as usual until the total payout exceeds a threshold d specified in the policy; the reinsurer pays subsequent claims. The stop-loss policy transfers tail risks to the reinsurer and lowers the variance of the insurer's portfolio. Stop-loss is provably optimal for reducing the variance of the insurer's claims (Theorem 1.4.3, p. 11), when the reinsurer uses the Net Principle.

Once the VO members have set a bound ε for their insurer's chance of eventual ruin and decided how risk-averse (α) the insurer is, then *ruin theory* (pp. 87-111) specifies the minimum capital the insurer needs to keep the chance of ruin below ε. The classical ruin model assumes that independent, identically distributed (iid) claims arrive according to a Poisson process and that the insurer's income from premiums holds steady at each time step. Under the Exponential Principle, we have:

$$\alpha = \frac{1}{w_I} |\log \varepsilon|, \text{ i.e., } \varepsilon = e^{-\alpha w_I}, \tag{3}$$

where α is the insurer's risk aversion index, w_I is its initial capital, and ε is the upper bound on ruin probability. Even if the insurer does not experience tail events, the chance of ruin may approach ε due to bad luck. Given its current capital and α, an insurer can apply Formula 3 periodically to determine whether the current upper bound ε' on the chance of ruin is still below ε, and work to increase capital if not. For correlated risks or unsteady premium income, the insurer will need to perform lengthy simulations to determine the chance of ruin, as discussed later.

3.3 Defining Classes of Risks

If the insurer does not know the probability distribution of a new risk X, it cannot use Formula 2 to set the premium. If risks are correlated, then Formula 3 no longer applies, so issuing the policy might push the chance of ruin above ε. To ensure that this does not happen, the insurer can run simulations to compute the probability of ruin, but this is a very lengthy process. Preanalysis offers a solution to these problems. The VO members can identify all the classes of insured access requests that they might like to allow in the future. Similar to traditional access control, each class might identify a type of information, a group of VO members allowed to access this data, and constraints on the context under which such accesses are to be permitted.

The insurer must subdivide classes until all risks (i.e., expected total policy claims) in the same class fit the same probability distribution, and then use that distribution to determine the (identical) premium for all policies to be issued in that class. Subdividing a class can be approached as a mixture model problem [5], where subclasses all belong to the same family of distributions, but with different parameters. Expectation maximization is popular for creating mixture models. Real-world claim sizes for many kinds of policies are exponentially distributed, so subdivision is not as daunting as it might sound. The classes must not become so small that the claims history for a class is too small for statistical significance of tests of goodness of fit, i.e., there is not enough data to compute its distribution's parameters within a desired error bound.

The insurer must periodically check that recent claims history is consistent with what it expected, by rebuilding its probability distribution for historical claims data for a class, and looking for changes and trends that may suggest premium changes. To help with this task, the insurer can employ *actuarial credibility theory* (pp. 203-227), which helps a model-builder extrapolate from a small sample that is highly relevant (recent history), by exploiting a large set of data that is not quite so relevant (the rest of history).

Preanalysis also helps the insurer with the problem of ensuring that its chance of ruin will not exceed ε once it issues a new policy. With correlated risks, computing the chance of ruin requires lengthy simulations. Positively correlated risks can significantly increase the overall risk, while the aggregate risk of policies with negatively correlated claim sizes can be lower than the sum of their individual risks. Thus the insurer can use portfolio management theory from the financial engineering community to reduce its overall risk by diversifying the types of risk assumed. The insurer can analyze historical data to estimate the correlations between risks of different classes in advance, run simulations to estimate the chance of ruin given particular numbers of policies in each class, and use the simulation results to cap the number of policies sold in each class.

3.4 Purchase Decisions

A consumer can use decision theory to determine its ex-

pected financial benefit from accessing a piece of information. Then the consumer must decide whether the benefit is worth the price of the policy. Because the consumer's benefit is uncertain and the consumer is risk-averse, it is too simplistic to buy the policy as long as the expected benefit exceeds the premium. Instead, consider the following inequality, similar to the Principle of Equivalent Utility:

$$E[u(w + Y - P)] \geq u(w), \tag{4}$$

where u is the consumer's utility function, w is its capital (or wealth) that it can use, and Y is a random variable representing the consumer's expected additional value (or revenue) from accessing the information. Let the consumer have an exponential utility function with parameter α_c. Then from Formula 4, we can derive the maximum premium P^+ the consumer is willing to pay:

$$P^+ = -\frac{1}{\alpha_c} \log(m_Y(-\alpha_c)), \tag{5}$$

where $m_Y(-\alpha_c) = E[e^{-\alpha_c Y}]$ is the moment generating function of Y around $-\alpha_c$. Thus, the consumer buys the policy if the premium is no more than P^+, reflecting the expected benefit and the chance that he might be worse off after using the information. As noted earlier, traditional actuarial methods do not provide this sort of decision theoretic formula to compare policy prices with possible benefits.

3.5 Rewarding Good Risk-takers

In the discussion so far, the insurer sets premiums based solely on the class of risk, i.e., the type of consumer, information, and circumstances of access. The VO can benefit by encouraging good risk-takers, i.e., consumers who do not result in claims, by giving them lower premiums. This is called a *bonus-malus system* (pp. 135-146), a branch of credibility theory. Though much more sophisticated methods are available, we adopt the simple and effective Dutch system (pp. 136-138), still used for auto insurance in the Netherlands.

The system has 14 steps, each with its own weight, which is a discount factor to be multiplied with the policy price obtained by a premium principle. These steps are updated at policy renewal. Consumers with no claims in the previous period ascend one step and get lower premiums, but those with claims filed against their policies descend several steps, resulting in higher prices.

3.6 Bootstrapping the Insurer

In deciding what types of sharing will be allowed, VO members must be careful to align members' individual incentives with the VO's shared purpose so that a consumer's benefit is also a benefit for the VO. If incentives are aligned, then every act of sharing has an expected net positive benefit for the VO. To help align incentives, the VO can offer an incentive scheme that rewards consumers with *wages* when their use of shared information benefits the VO as a whole. Molloy et al. [12] present an abstract model of a VO member's wage as a function of her own profit and the profit of all members, plus a base salary. The key for this scheme is how to choose the function so that making optimal decisions for the VO as a whole is in the member's best interest. With the right function, rational members will try to request insured accesses whose expected outcomes are aligned with the common goal of the VO.

The benefits directly attributable to insured access must be weighed against the opportunity cost to the members who contributed the insurer's capital reserves. However, if insured access is as popular and beneficial among VO members as they expected it to be when they set up the insurer, then the opportunity costs of capital reserves for consumers will be offset by their realized gains due to sharing.

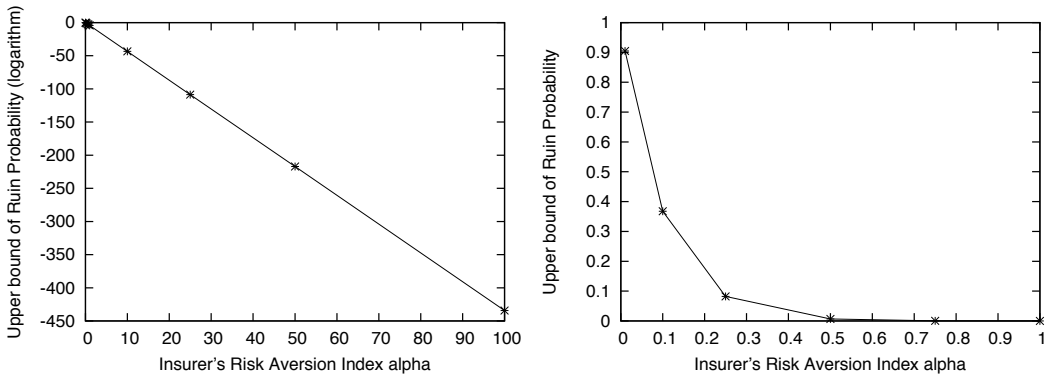
At startup, an insurer may have no historical claim information of its own. It may be able to use historical data from other organizations. If some relevant data is available for modeling a risk, but not enough for statistical significance, the insurer can use actuarial credibility theory to extrapolate from the relevant data plus a large set of slightly related data, to provide better risk estimates. With regard to rare events, the insurer can also use extreme value theory to obtain better estimates of the tail parts of risk distributions from available data. Even without historical data, actuaries manage to estimate future claim amounts for new classes of risks, including such exotic risks as alien abduction and damage to the legs of Heidi Klum, Michael Flatley, and Mariah Carey. Thus we can assume that actuaries will be able to get the system off the ground.

The insurer must determine the fees and/or profit-sharing scheme for the producers at startup. Perhaps the simplest profit-sharing approach is to wait until the end of a fiscal period, calculate the level of capital that the insurer must retain for probability of ruin ε, and distribute the excess capital among the producers. More sophisticated methods distribute the insurer's funds in excess of the *optimal dividend barrier*, which maximizes the total expected present value of the distributions (dividends) before ruin; there is a simple closed-form formula for the optimal dividend barrier under the common assumption that claim sizes follow an exponential distribution [6]. Once the barrier is set, VO members must decide whether to distribute the profit evenly or according to each producer's amount of sharing, risk assumed, benefit derived by consumers, or any other factor.

4. EXPERIMENTS

In this section, we use discrete event simulations to confirm that the theory presented in the previous section correctly predicts the likely outcome of insured access in an example scenario, and to understand the effect of different parameters on the outcome. Our simulator is written in C++, and uses the Boost C++ Libraries and their Math Toolkit. Our results show that on average, the expected capital of each VO member, the insurer, and the VO as a whole does grow over time. We also examine the insurer's probability of ruin as a function of its degree of risk aversion. Testing the techniques for estimation of distributions from claims data is beyond the scope of this paper, as real claims data is not available to us.

For the reasons discussed earlier, we price policies using the Exponential Principle, and use Formula 5 for consumers' policy purchase decisions. We consider a range of values for the Exponential Principle's parameter α, which is the insurer's degree of risk aversion. Figure 2 shows the upper bound ε on the probability of eventual ruin from Formula 3, with a range of risk aversion indexes for the insurer and a fixed initial capital, with and without a log scale. This figure is intended to help the reader visualize the impact of risk aversion on the chance of ruin; note, however, that Formula 3's assumptions do not quite hold in our simulations,



(a) Logarithm of Probability ($0 < \alpha \leq 100$) (b) Probability ($0 < \alpha \leq 1$)

Figure 2: Upper Bounds on Ruin Probabilities

as the insurer's income from premiums will not be constant at each unit of time and the insured accesses will not all have identical claim size distributions.

The simulations model a scenario where the VO is partitioned into producers and consumers. Each producer produces one unique map. Some maps are more sensitive than others, depending on who the consumer is. This sensitivity is reflected in the parameters of their distributions of claim sizes, and thus in their premiums. We model each insured access as an independent (uncorrelated) risk. We model the arrival of requests for insured access using a separate Poisson process for each consumer. For purchased policies, we model the arrival of claims using one Poisson process for each issued policy. The process starts only after the policy is issued, and we limit it to at most one claim per policy.

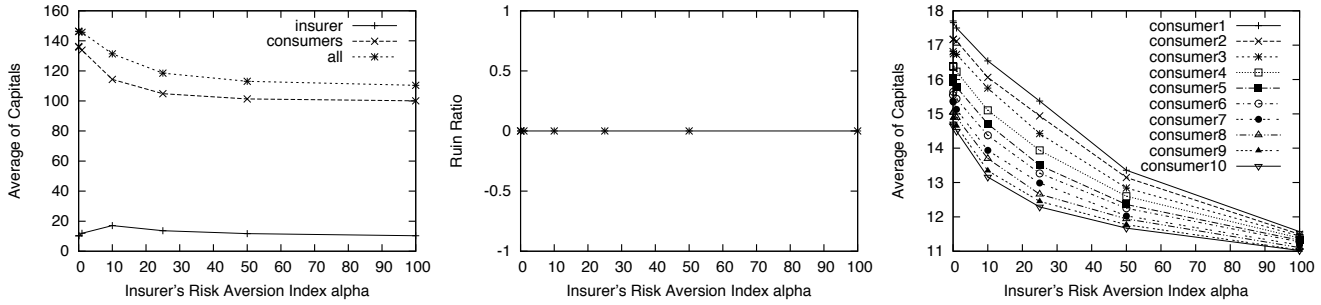
For each potential insured access, the consumer expects a certain benefit, modeled as a random variable, against which the premium must be weighed. The consumer's *profit* is its actual benefit minus the premium it paid. The exponential family of distributions is widely used in many disciplines for modeling outcomes of various kinds of transactions. Among the many members of the exponential family, normal distributions are common and easy to visualize, so we use normally distributed benefits. We model the receipt of benefit from insured accesses with a separate Poisson process for each access. The benefit's process starts only after the policy is issued, and the benefit arrives at most once per policy. For any Poisson process, the time between each pair of consecutive events is exponentially distributed.

For total claim sizes, which we refer to simply as *claims*, we adopt two distributions from the exponential family, which is widely used for modeling claims. The first is a normal distribution. However, a normal distribution can overestimate the total claims on a policy, because if one waits long enough when using a Poisson process for claim arrival, a claim will eventually arrive for any given policy. In contrast, in real life many policies never have any claims at all. The zero-adjusted gamma distribution (ZAGA) [15] is very effective at modeling this situation, because it explicitly models the chance of there being no claim at all. Thus when a claim arrives under the Poisson process, the ZAGA distribution explicitly models the chance that the "claim" is \$0. The ZAGA distributions we used have fatter tails than the normal distributions, thus illustrating the impact of rarer events.

When a consumer requests an insured access, we choose its producer uniformly at random from the producers it has not purchased from previously.

When a consumer requests a policy, the insurer uses the Exponential Principle to set the premium for the policy. The consumer computes its expected benefit from the insured access, then uses Formula 5 to determine the maximum premium it is willing to pay. If this is less than the quoted premium, the consumer buys the policy. Each consumer has its own parameter α_c for risk aversion, chosen uniformly at random from [.1, 10], and hence its own maximum premium for a particular map.

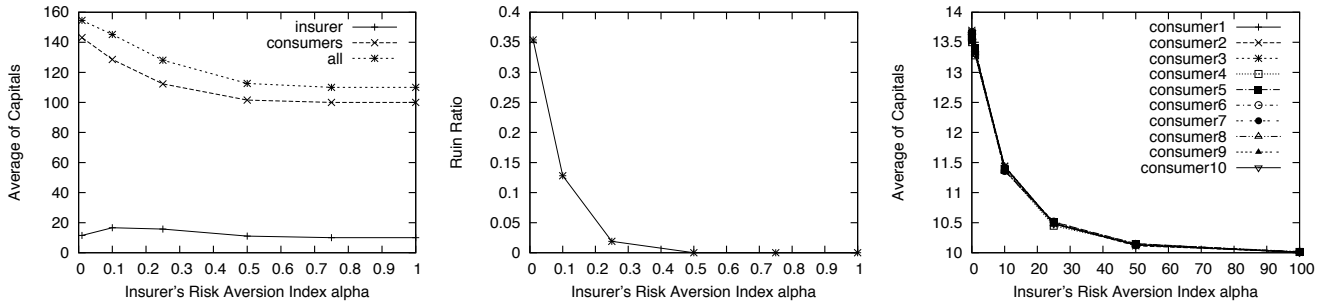
The simulation has to use concrete numbers for the benefit of maps and for claim sizes. For each map, we choose an average benefit uniformly at random in [1.0, 1.5], with its average claims drawn uniformly at random from [.5, 1] for the normal distribution. For the normal distributions, the range of possible means of claims and benefits is narrow and relatively close, as otherwise the outcome of the simulation will be dominated by the larger values. For the normal distributions, standard deviations are set so that three standard deviations from the mean (a reasonable threshold for tail events) is at most twice the mean, so that the tail starts by 3 for benefits and by 2 for claims. The ZAGA distributions are chosen so ZAGA claims have the same average amount as normally distributed claims. This means that when there is a non-zero ZAGA claim with probability 0.1 on average, its average amount is ten times higher than the average value under the normal distributions. To avoid dull simulations where the quoted premiums are usually larger than the maximum premium the consumer will pay, the mean of the distributions for benefits is generally larger than that for claims. The average claim size is equal to the premium under the Net Principle, which in turn is less than the premium under the Exponential Principle, which governs what the consumer will pay. In the simulation, premiums do not include a fee for the insurer or producer, and we do not share profits, so producers break even. We simulate the behavior of 10 consumers and 10 providers and track the wealth (capital) of each consumer plus the insurer, each of whom has an initial capital of 10 for sharing. The benefit from sharing comes from outside the VO, i.e., it is not taken from other VO members' capital. The insurer's ini-



(a) Average Capital (Normal Claim Distribution)

(b) Ruin Ratio (Normal Claim Distribution)

(c) Average Capital with Bonus-Malus (Normal Claim Distribution)



(d) Average Capital (ZAGA Claim Distribution)

(e) Ruin Ratio (ZAGA Claim Distribution)

(f) Average Capital without Bonus-Malus (Normal Claim Distribution)

Figure 3: Simulation Results

tial capital is rather low, a deliberate choice to allow us to investigate ruin empirically.

The simulation needs concrete λ parameters for the Poisson processes' exponentially-distributed inter-arrival times. We choose the λ parameters randomly from a normal distribution, with the means of the distributions chosen so that benefits typically arrive before claims, and consumers usually make their next insured access request after the previous request's benefits and claims are known. That translates to an average of five time steps between requests from the same consumer, two time steps for the claims on a new policy to arrive, and one time step to learn the benefits of an insured access. We run the simulation 100 time steps, which is about twice as much time as consumers usually need to get a chance to buy all 10 maps, with this range of Poisson parameters. We repeat the simulation 1,000 times and report the averages across all simulations. We computed the standard error for each reported average, as the standard deviation divided by the square root of the number of runs. The resulting error bars were too small to be observed, so we do not include them in the figures.

In addition to the capital of the insurer and the VO consumers, we present the *ruin ratio*, which gives the chance of insurer insolvency for a given level of insurer risk aversion. The ruin ratio is the fraction of runs where the insurer' capital became negative. In runs where ruin occurs, we assume that the VO loans the insurer enough funds to continue to pay claims until it is back in the black. Thus the simulation continues even after ruin, on borrowed funds.

Figures 3a and 3d present the average capital at the end of the simulation. The three lines in these graphs give the average capital of the insurer, the average sum of capital across all consumers, and the average sum of the capital of the insurer and the consumers. Figures 3b and 3e present the ruin ratio, which is rather large when the risk aversion index α is small and claims follow ZAGA distributions, even though the ruin ratio is always 0 when claims follow normal distributions. This is because claim sizes can be quite large under a ZAGA distribution, even though the average claim is generally smaller than the average benefit. More concretely, if we ignore the chance of a \$0 claim with ZAGA and instead assume that claims are always positive, then for ZAGA's parent family of gamma distributions, the average claim size is $k\theta = (5 + 10)/2 \times 0.99 \simeq 7.5$, while the mean claim size once \$0 claims are taken into account as in ZAGA is just $(0 + 0.2)/2 \times 7.5 = 0.75$. In contrast, the average claim size when claims follow a normal distribution is always $(0.5 + 1.0)/2 = 0.75$.

These figures show that the more risk averse insurer (i.e., larger α) has less capital but a lower ruin ratio, because the number of insured accesses decreases as α gets larger. Although the upper bounds on ruin probability in Figure 2 rely on assumptions not satisfied by our experiments, Figure 3e shows that these upper bounds are actually very good approximations. The ruin ratio for ZAGA claim distributions is rather large, due to the insurer's low initial capital and the high \$7.5 average claim size. In practice, as discussed earlier, the VO must start with sufficient capital for its planned portfolio size and take corrective action if the ruin probability approaches the VO's cap.

We evaluated how bonus-malus systems affect the capital of principals, using the Dutch system explained earlier. The

steps of consumers are updated every five time units according to the transition table, based on their number of claims in the previous period. To differentiate between good and bad risk takers, we set the probability of the ith consumer causing a claim to $i/10$.

Figures 3c and 3f show the results with normal claim distributions, with and without a bonus-malus system. These figures show the average capital of each consumer at the end of the run. These graphs show that consumers who cause fewer claims (i.e., those who have smaller ID numbers) have more capital when the bonus-malus system is enforced[3]. Without bonus-malus, there are no such differences among consumers. These results confirm that the bonus-malus system can reward good risk takers and punish bad ones.

5. CONCLUSION

We have presented *insured access*, the first demonstrably sustainable system for encouraging appropriate information sharing in a VO. Before insured sharing starts, VO members agree on the VO's degree α of risk aversion and its maximum tolerable level of risk, i.e., the chance ε that eventually the VO might not be able to compensate an information provider for damages attributable to sharing. The VO finds or sets up an insurer whose actions are governed by α and ε. To obtain access to a piece of information owned by VO member Alice, VO member Bob must purchase a liability policy from the insurer. The insurer will not issue the policy if the VO would be exposed to more than its maximum tolerable aggregate level of risk as a result. Otherwise, the price of the policy is determined by the type of information, the insurer's current capital reserves, Bob's track record, the insurer's bonus-malus scheme, and the insurer's premium pricing principle. If Bob misuses Alice's information and Alice suffers damages as a result, then Alice can submit a claim and be reimbursed for her suffering.

We showed how to estimate the risk associated with an insured access, i.e., the probability distribution of future damages to the provider. We showed how reinsurance can cap the risk associated with rare events, and provided two schemes to ensure that information providers directly benefit from sharing. Our simulations of a map-sharing scenario showed that each participating VO member, and the VO as a whole, can expect to benefit from insured access, while the risk of failure of the system is limited by ε.

Acknowledgements: This work was supported in part by National Science Foundation awards CNS-0963943, CNS-0964295 and CNS-0963715.

6. REFERENCES

[1] Anonymous. Horizontal Integration: Broader Access Models for Realizing Information Dominance. Technical Report JSR-04-132, MITRE Corporation JASON Program Office, December 2004.

[2] L. L. Cam. Maximum Likelihood: An Introduction. *International Statistical Review / Revue Internationale de Statistique*, 58(2):153–171, Aug. 1990.

[3]When we did the same experiment with ZAGA claim distributions, bonus-malus did not significantly impact capital, as ZAGA distributions already model the probability of filing claims.

[3] P.-C. Cheng, P. Rohatgi, C. Keser, P. A. Karger, G. M. Wagner, and A. S. Reninger. Fuzzy Multi-Level Security: An Experiment on Quantified Risk-Adaptive Access Control. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 222–230, May 2007.

[4] N. Dimmock, A. Belokosztolszki, D. Eyers, J. Bacon, and K. Moody. Using trust and risk in role-based access control policies. In *Proceedings of the Symposium on Access Control Models and Technologies*, pages 156–162, 2004.

[5] I. Dinov. Expectation maximization and mixture modeling tutorial. *Statistics Online Computational Resource*, 2008.

[6] H. U. Gerber, E. S. W. Shiu, and N. Smith. Methods for estimating the optimal dividend barrier and the probability of ruin. *Insurance: Mathematics and Economics*, 42(1):243–254, Feb 2008.

[7] A. Ghosh and A. Roth. Selling Privacy at Auction. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, pages 199–208, 2011.

[8] M. Gilli and E. Këllezi. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, 27(2-3):207–228, May 2006.

[9] K. Hoo. How much is enough? A risk-management approach to computer security. Working paper, Center for International Security and Cooperation, June 2000.

[10] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory Using R*. Springer, second edition, 2008.

[11] G. Lebanon, M. Scannapieco, M. Fouad, and E. Bertino. Beyond k-Anonymity: A Decision Theoretic Framework for Assessing Privacy Risk. In *Privacy in Statistical Databases*, volume 4302 of *Lecture Notes in Computer Science*, pages 217–232. Springer Berlin / Heidelberg, 2006.

[12] I. Molloy, P.-C. Cheng, and P. Rohatgi. Trading in risk: using markets to improve access control. In *Proceedings of the 15th New Security Paradigms Workshop*, pages 107–125, 2008.

[13] N. Nissanke. Risk based security analysis of permissions in RBAC. In *Proceedings of the 2nd International Workshop on Security in Information Systems*, pages 332–341, 2004.

[14] C. Riederer, V. Erramilli, and A. Chaintreau. For Sale : Your Data By : You. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, pages 1–6, 2011.

[15] E. Tong, C. Mues, and L. Thomas. A zero-adjusted gamma model for estimating loss given default on residential mortgage loans. In *Proceedings of the Credit Scoring and Credit Control XII Conference*, pages 24–26, Aug. 2011.

[16] A. Tsanakas and E. Desli. Measurement and Pricing of Risk in Insurance Markets. *Risk Analysis*, 25(6):1653–1668, 2005.

[17] L. Zhang, A. Brodsky, and S. Jajodia. Toward information sharing: benefit and risk access control (BARAC). In *Proceedings of the 7th IEEE International Workshop on Policies for Distributed Systems and Networks*, pages 9–53, 2006.