# Act for Affordable Data Care

Saikat Guha
Microsoft Research, India
saikat@microsoft.com

Srikanth Kandula
Microsoft Research, Redmond
srikanth@microsoft.com

**Abstract** — Data breaches, *e.g.* malware, network intrusions, or physical theft, that lead to the compromise of users' personal data, happen often. The impacted companies lose reputation and have to spend millions of dollars providing affected users with identity and credit monitoring services. Users can suffer from fraudulent transactions and identity theft. At present, there are no mechanisms that both cover the risk from accidental data breaches and incentivise best practices that would prevent such breaches. This paper proposes a data breach insurance mechanism and the associated risk assessment technology to meet these goals. In so doing, we break from (failed) past approaches that seek to solve the problem solely through technology.

## Categories and Subject Descriptors

D.4.6 [**Security and Protection**]: Information flow controls—*Insurance*

## General Terms

Security, Economics

## Keywords

Privacy, Data Protection, Insurance

## 1. INTRODUCTION

Car insurance has a simple value proposition: accidents are inevitable and costly; insured entities are protected from damages for a small fee. An individual's fee depends on the frequency and severity of accidents across other similar individuals. High-risk groups (*e.g.*, teenage drivers) pay more than lower-risk groups (*e.g.*, clean history). Individuals are incentivized to follow good practices. For instance, teenagers can save up to 30% by installing a monitoring device that tracks their driving behavior [19]. And, strong financial disincentives further discourage high-risk behavior, *e.g.*, denying coverage if the device detects unlawful behavior leading up to an accident. These devices have cut crashes by 20% in the 17–25 age group [13]. Financial incentives can therefore have a significant effect on changing user behavior.

Data breaches are inevitable and costly. In July 2012, Yahoo! suffered a breach that leaked passwords and usernames of 400K users [4]. The passwords were in plain-text and were posted on public forums. In June 2012, LinkedIn suffered a breach that leaked 6.5M passwords [3]. While these passwords were hashed, they were not salted prior to hashing; using dictionary-attacks and rainbow-tables, hackers were able to decode and post online over half of the leaked passwords [16]. It triggered a $5M lawsuit [6]. Between April and May 2011, Sony suffered a string of breaches where personal information (potentially including credit card numbers) of over 100M subscribers of the Play Station Network were compromised [2]. Sony's cleanup cost is estimated at $171M [9]. According to Privacy Rights Clearinghouse, which maintains a chronology of publicly reported data breaches [18], 562 million financial records and social security numbers (SSNs) have been compromised in over 3200 incidents since 2005 (774 million records if including user passwords). In 2010 alone, an estimated 8.1 million users were victims of identity-theft, with each incident costing individuals on average $630 [15].

The recent focus on building privacy-preserving systems, to ultimately protect against the misuse of data, misses this larger picture. Privacy-preserving systems, be they for advertising and personalization [8, 11, 22], social-networks [1], location services [12, 14], or analytics [5] are well-architected but seem to be out-of-reach for an industry where even the most tech-savvy users and companies fumble with the basics of data security. We find businesses routinely store passwords, credit card numbers, and even SSNs in the clear, and do not secure portable devices that contain such data (§2.1). Users, valuing convenience over data security, routinely share passwords across sites and allow websites to store their credit card numbers indefinitely (§2.2).

We take a perhaps more pragmatic position. We ask if a general framework can provide some level of relief against breaches today while, over time, driving changes in user behavior and business practices.

Consider what might be called "data breach insurance": 1) if an insured user's data is breached, the insurance provider will, for a small fee, underwrite damages due to fraudulent transactions and credit accounts


*Hotnets '12,* October 29–30, 2012, Seattle, WA, USA.

opened in the user's name. 2) If an insured business suffers a breach, the insurance provider will, for a fee, underwrite the costs of lawsuits brought against the compromised business, costs of providing credit-monitoring services to affected users, and other costs to clean up after the breach. Such a framework would bring immediate relief to affected users and businesses.

The harder problem is to incent good user behavior and business practices. In particular, if there is no way to quantify behavior that is directly correlated with data breaches, there would be no *fair* fee and hence no incentive for users and businesses to change their practices. Indeed, the key to using insurance to change user behavior lies in the ability to assess risk, *i.e.*, to distinguish good actors and good practices from the bad.

We ask how technology can enable risk assessment of users and businesses. Consider a software agent that a user may opt-in to installing on his devices, something akin to the car monitoring device. The software monitors (in a privacy-preserving manner) the user's online behavior, for instance whether the user re-uses bank website passwords on other websites, whether the user opens email attachments from unknown senders, and so on. In addition, the software can (gently) nudge the user towards better alternatives. This software would compute a per-user "data-safety rating" (much like a credit rating) which is used to set the user's insurance premium. From a user study, we identify several easy-to-measure aspects of user's online behavior that are correlated with identity-theft (§4). By offering deep discounts to users with a high safety-rating, the insurance provider can incent or reinforce good behavior.

Risk-assessing businesses is, in general, harder since internal systems and processes of a business are typically not open to third-party audits. Ideally, one would simply require the business to publicly disclose its practices (*e.g.*, whether passwords are stored in the clear, whether full-disk-encryption is enabled on portable devices to protect against theft, etc.). But, businesses may lie. Conveniently, regulatory agencies (*e.g.*, the Federal Trade Commission (FTC)) are already tasked with oversight against misleading or incorrect disclosures today. However, businesses today conform by not disclosing anything meaningful.[1]

Suppose that the insurance provider risk assesses businesses based only on what is disclosed in its privacy policy. It can offer deep discounts to businesses with a high safety-rating. This is a new financial incentive for businesses to publicly disclose meaningful policies and to improve their practices over time.

Overall, this paper has three contributions:

**1) Characterization.** To set the stage, we present a longitudinal study of the nature and severity of over 3200 data breaches (§2). We also characterize risky online behavior among 653 surveyed users. The data shows that the point solutions offered by past research has limited impact on the larger data breach problem.

**2) Framework.** We propose an insurance mechanism for data breaches (§3). Our framework provides immediate financial relief to affected parties without (initially) requiring any changes. It has financial incentives that, over time, steer users and businesses towards practices that prevent data breaches.

**3) Feasibility.** We present a proof-of-concept study of our framework. In particular, from our user study data, we present a decision-tree classifier that risk-assesses online behavior (§4). We reveal a few easy-to-measure behaviors that correlate with data breaches. We also quantify user's willingness to opt-in to such an insurance and how much they are willing to pay (§6). We also analyze privacy policies of 89 Alexa-ranked sites towards risk-assessing them (§5).

## 2. CHARACTERIZING BREACHES

In this section we characterize the data breach problem. We use two datasets. The first is a public dataset of breaches [18]. The second is a survey of online behavior of 653 users that we conducted. We appear to be the first users of the former dataset, and are not aware of any dataset equivalent to the latter. Overall we find that data breaches are rife, and that users seem to prefer convenience over protecting themselves.

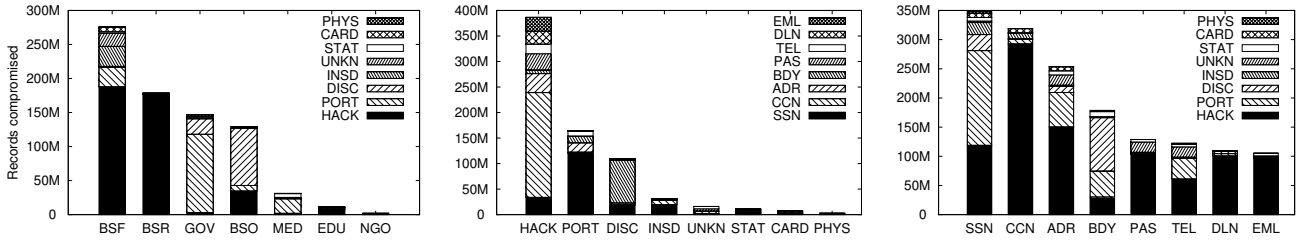
### 2.1 Data Breaches

**Dataset:** The Privacy Rights Clearinghouse maintains a publicly accessible chronology of reported data breaches since 2005 [18]. The list is not complete since it only contains breaches that are disclosed. Breaches are labeled with the type of business breached (see Fig. 1), the type of the breach, some commentary and the number of records breached (if known). A single record may include multiple types of data (*e.g.*, SSNs, credit card numbers, etc.) for a user. We determine the type of data by looking for keywords in the commentary (*e.g.*, credit card), excluding instances with a nearby no or not; we found from manual inspection that the heuristic, while rudimentary, was quite accurate. As of July 2012, the dataset covers 3226 incidents.

**Observations:** While the median sized data breach compromised 2K records, 2.4% of the breaches (55 incidents) compromised over a million records each.

Records were mostly breached at businesses in the financial services (35%) and retail (23%) sectors and at

[1] *e.g.*, Dropbox updated its privacy policy to replace "All files stored on Dropbox servers are encrypted (AES256) *and are inaccessible without your account password*" (emphasis added) with "All files stored on Dropbox servers are encrypted (AES 256)". It also changed "Dropbox employees aren't able to access user files" to the more vague "Dropbox employees are prohibited from accessing user files". The changes were due to an FTC complaint [21].




**(a)** Entity breached

**(b)** Breach Type

**(c)** Data compromised

| | | | | | |
|---|---|---|---|---|---|
| | | PHYS | Non-electronic records lost, or stolen | SSN | Social-security # |
| BSF | Financial Services Business | CARD | Credit card leaked (non-hacking) *e.g.*, skimmers | CCN | Credit/debit card # |
| BSR | Retail/Merchant Business | STAT | Stationary device lost, or stolen | ADR | Mailing address |
| GOV | Government and Military | UNKN | Unknown | BDY | Birthdate |
| BSO | Other Business | INSD | Insider *e.g.*, disgruntled employee | PAS | Password |
| MED | Healthcare Providers | DISC | Unintended disclosure *e.g.*, posted on website | TEL | Phone no. |
| EDU | Educational Institutions | PORT | Portable device lost, or stolen | DLN | Driver's license |
| NGO | Nonprofit Organizations | HACK | Hacking or malware | EML | Email address |

**Figure 1:** Data breaches collated by Privacy Rights Clearinghouse

government and military organizations (19%). Medical, educational and nonprofit institutions account for less than 6% of breached records.

Hacking or malware was the primary cause of breached records (48%). Loss of portable devices and unintended disclosure (e.g., data posted on a public web site) accounted for 21% and 14% respectively.

The types of data breached in order of frequency are SSNs (45%), credit card numbers (40%), addresses (32%) and birth-dates (22%). Other data types (passwords, phone numbers, driving license numbers and email addresses) were present in 16% or fewer breached records. Note that a breach may compromise multiple types of data. Hence these numbers add up to more than 100%.

Correlating across these axes leads to some interesting observations. We see from Fig. 1a that breaches at the government, military and medical institutions largely due to loss or theft of portable devices and unintended disclosure. On the flip side, retail businesses attribute almost all their breaches to hacking or malware. We see from Fig. 1c that four data types (credit cards, passwords, driver license numbers and email addresses) are much more likely to be revealed due to hacking and malware than expected based on the frequency of hacking. On the flip side, the other four data types have a more than expected contribution from loss of portable devices and unintended disclosure.

## 2.2 User Behavior

**Dataset:** We performed a survey where we asked participants about their online behavior. Fig. 2 summarizes the questions in the survey.[2] The answer choices were none, 1–3, 4–10, 11+, and no-answer. We recruited volunteers over email, social-networks, and Amazon Mechanical Turk (AMT) workers in the United States. To control quality, we ignored responses that were submitted two standard deviations faster than the mean time to complete the survey, and those that were self-inconsistent *e.g.*, where the user reports being unable to recover more devices than he reports to have lost or misplaced. In all, 653 people successfully completed the survey. We were surprised at the response quality, while we expected to discard upwards of 25% of respondents, we ended up discarding less than 10%.

We included informational queries about gender, age and profession to identify sample bias. The 25 to 34 age group, and the 35 to 49 age group were the most common, accounting for 39% and 28% of the participants respectively. In large part due to the diversity of AMT's user-base, we got an unbiased sample along gender (49% male:51% female) and along professions.

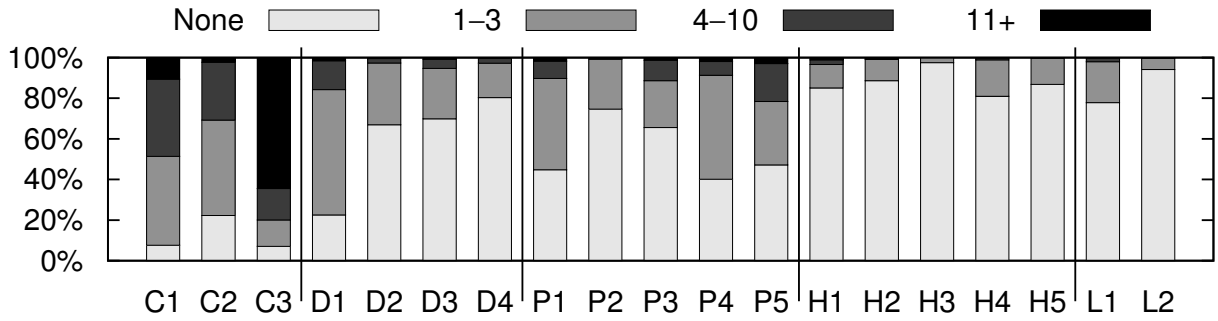
**Observations:** Overall, we found that users seem to have a strong preference for convenience over protecting themselves, and lacking solutions that offer both, are prone to risky behavior.

The best behavior was in the physical security category (H). Loss of laptops, wallets and phones was infrequent (H2–H5). Further, only a few users left devices unlocked and unattended (H1).

In the data protection category (D), several users reported emailing documents containing personal identifiable information and sensitive numbers in clear text (D3, D4). Many more reported using public devices, e.g., at kiosks and libraries (D2). Even more were prone to stay persistently logged into many sites (D1). In context of their likelihood to leave devices unlocked and unattended (H1), this could be an easy breach.

A sizable fraction of users reuse passwords across multiple sites (P2, P3, P5). Email passwords were reused more frequently than bank passwords. In the context of emailing sensitive information (D3, D4), and using email for recovering forgotten passwords, a breached email account could cascade into a much broader breach. On

[2]Survey questionnaire posted at `http://bit.ly/ScrMBF`




| | | | |
|---|---|---|---|
| | **Credit and Debit Cards** | | **Password Security** |
| C1 | Online transactions (monthly) | P1 | Bank passwords changed (last year) |
| C2 | Sites credit card saved at | P2 | Banks with same password |
| C3 | Monthly statements checked (yearly) | P3 | Non-bank sites with bank password |
| | **Data Protection** | P4 | Email passwords changed (last year) |
| D1 | Sites persistently logged into | P5 | Non-email sites with email password |
| D2 | Public devices used (monthly) | | **Physical Security** |
| D3 | Identity document emailed (last year) | H1 | Device left unlocked and unattended (last year) |
| D4 | Sensitive number emailed (last year) | H2 | Phones lost or misplaced (last year) |
| | **Loss** | H3 | Laptop/tablet lost or misplaced (last year) |
| L1 | Fraudulent transactions (last two years) | H4 | Wallet/keys lost or misplaced (last year) |
| L2 | Transactions above resulting in money loss | H5 | Items above that weren't recovered |

**Figure 2:** Survey responses of 653 participants

the positive side, several users report frequent changes to both bank and email passwords (P1, P4).

We find that 80% of the users trust websites to store their credit card numbers. Over 30% do so at four to ten different sites (C2). Online transactions using cards were quite frequent (C1). Combined with financial and retail businesses being most susceptible to hacking and malware attacks (Fig. 1a), it is no surprise that fraudulent transactions are common (L1), several of which lead to monetary loss (L2).

In §4 we will show a strong correlation between fraudulent transactions (L1) and various user behavior. We also surveyed users about their willingness to pay for insurance, and to opt-in for a monitoring software, which we discuss in §6.

## 3. A DATA BREACH INSURANCE?

To check that our proposal for a data breach insurance makes sense, we verify whether the seven necessary criteria for an insurance [17] hold in this context.

**1. Scale.** *There should be a large pool of prospective clients.* The number of users susceptible to identity-theft is large. Similarly, the number of businesses susceptible to data breaches is large.

**2. Non-Catastrophic.** *Losses must not happen all at once that the insurance provider is bankrupt.* The median breach affected 2K records. The largest involved less than 10% of the Internet population of which significantly fewer may suffer an actual identity-theft.

**3. Loss.** *Losses must be large enough to justify paying the insurance premium.* As discussed, this is the case today for businesses where the clean-up cost of breaches can run into millions. For users, as mentioned, an identity theft incident costs on average $630 [15].[3]

**4. Premium.** *Premiums must be large enough to cover losses and admin costs, yet low enough that clients will lose much more if not insured.* Our feasibility analysis in §6 finds that user premiums could be as low as $20, which 77% people report willingness to pay.

**5. Incident Reporting.** *Loss must result from an incident (*e.g.*, unauthorized access, or theft) with a report of the time, place, and cause of the incident.* Many jurisdictions (*e.g.*, 46 out of the 50 US states) already have mandatory breach notification laws [10].

**6. Accident.** *Loss must be outside client's control.* Businesses and users do not control when and how data breaches happen. We discuss insurance fraud in §6.

**7. Risk-assessment.** *It must be possible to estimate the probability of loss, and the magnitude of the loss. To incentivise adoption of best practices, the probability of loss must be estimated for each individual.* The risk-assessment system we propose in this paper (§4) addresses this requirement. The magnitude of the loss can be estimated from historical data, and policy caps can bound risk.

With this analysis, we believe that a data breach insurance can work.

## 4. RISK ASSESSMENT: USERS

**Intuition:** To assess risk, we need to identify aspects of user behavior that are correlated with them suffering a data breach. Based on data from survey responses, Table 1 identifies the features that discriminate between

[3]Sadly, [15] does not further itemize these costs.

| Response | all | victims (Δ) |
|---|---|---|
| C1: Online transactions (monthly) $\geq 4$ | 48% | 67% (+18%) |
| H4: Wallet/keys lost or misplaced (last year) $> 0$ | 18% | 31% (+12%) |
| C2: Sites credit card saved at $\geq 4$ | 30% | 42% (+12%) |
| D1: Sites persistently logged into $> 0$ | 75% | 87% (+11%) |
| D3: Identity document emailed (last year) $> 0$ | 29% | 40% (+11%) |
| H5: Items above that weren't recovered $> 0$ | 12% | 20% (+8%) |
| P5: Non-email sites with email password $\geq 4$ | 21% | 28% (+7%) |
| C3: Monthly statements checked (yearly) $\geq 4$ | 78% | 85% (+6%) |
| D4: Sensitive number emailed (last year) is 1–3 | 16% | 23% (+6%) |
| D2: Public devices used (monthly) $> 0$ | 32% | 37% (+5%) |
| P4: Email passwords changed (last year) is 1–3 | 50% | 55% (+5%) |
| P1: Bank passwords changed (last year) is 0 | 43% | 37% (-5%) |
| P2: Banks with same password $> 0$ | 24% | 30% (+5%) |

**Table 1:** Behavior correlated with loss

all users and victims, *i.e.*, those who reported non-zero fraudulent transactions (L1). We list all features that occur frequently and have a probability difference of at least ±5%. We find that victims of fraudulent transactions are likely to engage in more online transactions, are more likely to have lost or misplaced important objects, save their credit card information on more sites, email sensitive documents, etc. as compared to the average.

**A Potential Risk Assessment Technique:** Combining the features above, we want a technique that assigns a risk score to users. Naively combining individual features results in a combinatorial explosion. Instead, we use the C5.0 decision tree classifier [20] which takes as input a set of labeled data items and recursively partitions the dataset into two by choosing at each juncture the feature that most reduces the entropy of each partition. Users who responded with a non-zero value for L1 are labeled At Risk, and the rest labeled Low Risk. Given a bound on the acceptable classification error, C5.0 prunes the tree to avoid over-fitting.

To evaluate the predictive power of this technique we used 75% of the data for training, and reserved 25% for testing. Fig. 3 shows the decision tree learned by C5.0 from our data. This tree manages to correctly classify 84% of victims in the test data, which is quite good. Note that a user labeled as low risk in the data may, in fact, be at risk (based on his behavior) but simply lucky enough to *not yet* have seen unauthorized transactions. Hence, we chose C5.0 parameters to lower false negatives (known victims classified as low risk).

A couple of points are worth noting. First, on the decision tree, each path to the leaf encodes a conjunction of boolean expressions that result in the classification listed on the leaf. For example, the left-most branch (H4 $> 0 \Rightarrow$ At Risk) concludes that users that have recently lost important objects (like wallets, or keys) are likely at risk. Similarly, the right-most branch concludes that of the remaining users (H4 $= 0$), those that don't check their monthly statements more than 3 times a year and have emailed identity documents in the past (C3 $\leq 3 \wedge$ D3 $> 0$), but don't ever use public devices (D2 $= 0$), pose low risk. Second, note that the nodes at lower levels on the tree are based on small subsets of the original dataset and may hence be unreliable. That we could get meaningful divisions with limited samples indicates that risk assessment is viable. More data and more sophisticated learners could improve the results.

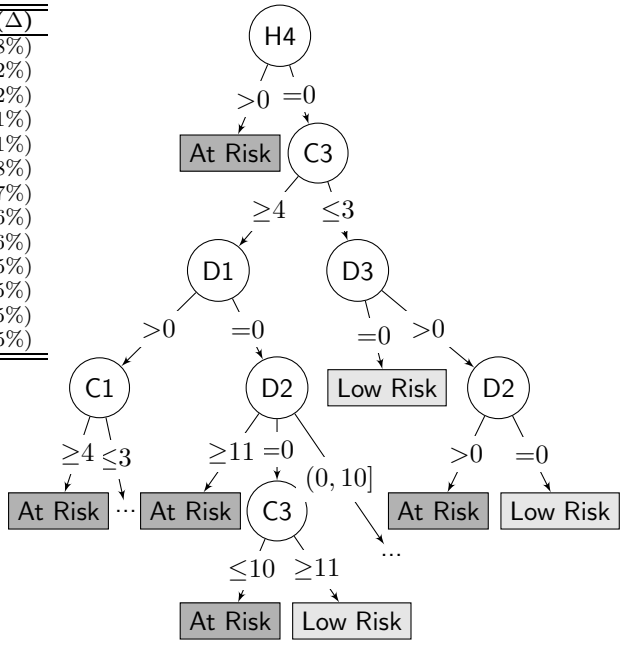



**Figure 3:** Decision tree classifier

| Privacy-policies mentioning | # |
|---|---|
| Personally identifiable information | 65 |
| Credit or debit cards | 30 |
| Use of encryption (other than SSL/TLS) | 12 |
| Employee confidentiality policy | 8 |
| Password encryption policy | 1 |
| Unintended disclosure policy | 1 |
| Portable device policy | 0 |

**Table 2:** Analysis of Privacy-Policies

## 5. RISK ASSESSMENT: BUSINESSES

For businesses, we analyze the privacy policies of top-ranked websites. Our automated crawler was able to find privacy policies of 89 sites out of the top 500. We manually identify keywords based on Table 2, to shortlist policy clauses of interest that we then review manually.

While most sites mention what data they collect (*e.g.*, 65 mention PII, *i.e.*, personally identifiable information), few mention how data is stored. The few that mention encryption (other than in the SSL/TLS context), do so only in vague terms (*e.g.*, "we take reasonable security measures [which] include firewalls and encryption"). This is troubling given the LinkedIn incident and Dropbox policy-change where improper application of cryptography rendered it impotent. Only one site (`mozilla.org`) concretely states "your password will be stored on our servers in an encrypted form called a hash".

While some sites explicitly mention their employees may access user data (but are contractually bound to

confidentiality), none mention security measures on employees' portable devices. Recall, lost portable devices account for 21% of breached records.

We believe businesses can be risk-assessed by correlating features extracted from privacy policy clauses with breach notifications. The current state however is that, except for one site, none of the sites disclose anything meaningful. One way to solve this impasse is for the insurance provider to publicly list a set of "potent" privacy policy clauses (*e.g.*, Mozilla's password encryption clause) that businesses can choose to (or choose not to) include in their privacy policy.

## 6. DISCUSSION

**Premium.** We asked our survey participants how much premium (if any) they would pay for a policy cap of 10 times that amount. 77% users reported they would pay. The median premium was between $20 and $40 per year. Note that this is less than identity-theft monitoring services (*e.g.*, Zendough, $15/month) that users subscribe to today.

For a rough estimate, assume the premium is $20, and the insurance company pays the full policy cap ($200) to the victims. The insurance company earns $10K in premiums, from the 500 willing users in our dataset, and pays out $6.4K to the 32 victims, for a $3.6K profit per year. The break-even point is when at least two-thirds of the willing users actually sign up.

**Behavior change.** We asked our survey participants which (if any) risky behaviors they would like a software monitor to warn them about for a 30% discount on their premium. An overwhelming fraction (94%) indicated they want to be warned, with 28% selecting all six risky behaviors listed (*e.g.*, banking password reuse, saving credit card on file).

**Fraud, and Moral hazard.** Insurance companies today deal with fraud through a variety of mechanisms including investigating suspicious claims. We expect the same will apply to data breach insurance. Moral hazard is a situation where the insured party has a tendency to take more risks than if they were not insured. Insurance companies address this today by either factoring this risk into the premium, or by limiting payout (e.g., policy cap) and transferring the extra risk back to the party. We expect the same will apply to data breach insurance.

**Advantageous (vs. Adverse) Selection.** If only the high-risk users choose to be insured, insurance would not viable as a business. In recent work, economists show that low-risk users are also much more likely to pay for insurance for peace-of-mind [7].

**Complementing Existing Mechanisms.** The tools proposed here to monitor users' online behavior, assess risk and nudge users' towards improved behavior will complement existing mechanisms. For example, if effective, they would reduce the total costs due to credit card fraud and hence credit card companies may offer their customers incentives to deploy such tools.

## 7. SUMMARY

In this paper we propose an insurance framework for addressing the larger data breach problem, providing immediate relief to victims today (without requiring any changes), and incentivising users and businesses to adopt best practices over time. We believe fine-grained risk-assessment technology can enable this. We present data that supports the feasibility of our approach.

## 8. REFERENCES


[1] R. Baden, A. Bender, D. Starin, N. Spring, and B. Bhattacharjee. Persona: An Online Social Network with User-Defined Privacy. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)*, Barcelona, Spain, Aug. 2009.

[2] BBC News. Sony warns of almost 25 million extra user detail theft. May 2011. `http://bit.ly/Nu8gSg`.

[3] BBC News. LinkedIn passwords leaked by hackers. June 2012. `http://bit.ly/Nq8XdI`.

[4] BBC News. Yahoo investigating exposure of 400,000 passwords. July 2012. `http://bit.ly/LiItsT`.

[5] R. Chen, A. Reznichenko, P. Francis, and J. Gehrke. Towards Statistical Queries over Distributed Private User Data. In *Proceedings of the 9th Symposium on Networked Systems Design and Implementation (NSDI)*, San Jose, CA, Apr. 2012.

[6] D. Clark. LinkedIn sued for $5 million over hacked passwords. *The News Tribe*, June 2012. `http://bit.ly/LiIvRv`.

[7] D. M. Cutler, A. Finkelstein, and K. McGarry. Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance. In *NBER Working Paper No. 13746*, 2008.

[8] M. Fredrikson and B. Livshits. RePriv: Re-Envisioning In-Browser Privacy. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*, 2011.

[9] D. Goodin. PlayStation Network breach will cost Sony $171m. *The Register*, May 2011. `http://bit.ly/LSfrGW`.

[10] P. Greenberg. State Security Breach Notification Laws. *National Conference of State Legislatures*, Feb. 2012. `http://bit.ly/PrhjFD`.

[11] S. Guha, B. Cheng, and P. Francis. Privad: Practical Privacy in Online Advertising. In *Proceedings of the 8th Symposium on Networked Systems Design and Implementation (NSDI)*, Boston, MA, Mar. 2011.

[12] S. Guha, M. Jain, and V. Padmanabhan. Koi: A Location-Privacy Platform for Smartphone Apps. In *Proceedings of the 9th Symposium on Networked Systems Design and Implementation (NSDI)*, San Jose, CA, Apr. 2012.

[13] J. Insley. Car insurance: satellite boxes 'make young drivers safer'. *The Guardian*, Apr. 2012. `http://bit.ly/N7ONGW`.

[14] S. Jaiswal and A. Nandi. Trust No One: A Decentralized Matching Service for Privacy in Location Based Services. In *Proceedings of the Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld)*, 2010.

[15] Javelin Strategy & Research. 2011 Identity Fraud Survey Report: Consumer Version. Feb. 2011. `http://bit.ly/Nu9kFI`.

[16] J. Kirk. How Charles Dickens Helped Crack Your LinkedIn Password. *PCWorld*, June 2012. `http://bit.ly/Nu8gS9`.

[17] R. I. Mehr, E. Cammack, and T. Rose. *Principles of Insurance*. R. D. Irwin, 8 edition, 1985. See `http://bit.ly/OKpQhX`.

[18] Privacy Rights Clearninghouse. Chronology of Data Breaches Security Breaches. `http://bit.ly/LiItt4`. Accessed Jul. 2012.

[19] Progressive Casualty Insurance Company. Snapshot Discount: Pay As You Drive. `http://pgrs.in/N35uo6`, Mar. 2011.

[20] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kauffman, 1993.

[21] R. Singel. Dropbox Lied to Users About Data Security, Complaint to FTC Alleges. *Wired*, May 2011. `http://bit.ly/Npfg4e`.

[22] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas. Adnostic: Privacy Preserving Targeted Advertising. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, Feb. 2010.