Prioritizing Security Spending: A Quantitative Analysis of Risk Distributions for Different Business Profiles

Armin Sarabi, Parinaz Naghizadeh, Yang Liu, Mingyan Liu
University of Michigan, Ann Arbor
{*arsarabi, naghizad, youngliu, mingyan*}@*umich.edu*

Abstract

This paper aims to understand if, and to what extent, business details about an organization can help provide guidelines for better resource allocation across different preventive measures, in order to effectively protect, detect, and recover from, different forms of security incidents. Existing work on analyzing the distribution of risk across different incident categories, most notably Verizon's latest Data Breach Investigations Report, provide recommendations based solely on business sector information. In this paper, we leverage a broader set of publicly available business details to provide a more fine-grained analysis. Specifically, we use incident reports collected in the VERIS Community Database (VCDB), as well as data from Alexa Web Information Service (AWIS), to train and test a sequence of classifiers/predictors. We show that compared to using business sector information alone, our method can achieve the same accuracy by allowing organizations to focus on a sparser set of incident types, thus achieving the same level of protection by spending less resources on security through more judicious prioritization.

1 Introduction

Data is an important asset in every business; the valuable data of an organization may include private information such as medial records, customer credit card numbers, or even trade secrets, as well as public information such as data hosted on a cloud service provider, or the website of an online commerce company. Any incident involving such data, whether intentional (targeted attacks) or unintentional (internal errors), can disrupt a business and inflict damage on its assets and reputation. Therefore, a portion of an organization's resources should be dedicated to protecting itself from security incidents; preventive measures include maintaining regular backups, keeping software up to date, and employee education in order to reduce miscellaneous errors.

However, determining how to allocate resources in protecting one's assets, as well as choosing an optimal level of investment in each preventive measure, is not a trivial task, as there is a wide variety of ever-changing attack methods. To help identify common forms of data incidents, a number of projects have been created to collect information about incidents that involve some sort of data loss. Some of these projects, such as [26] and [14], focus exclusively on hacking attacks, while some (e.g. [22]) cover a broader range of incidents, including human errors, and physical loss of data due to theft. Using these reports, organizations are able to identify prevalent incident vectors, and invest in self-protection in a more optimal way. However, a point that should not be overlooked is that not all businesses should be treated the same, as each business is prone to different forms of incidents. For instance, a cloud hosting company might be more likely to suffer from hacking or denial of service attacks, while a medical institution with a large number of personnel runs a relatively higher risk of data loss through human error.

In this paper, we aim to better understand how a collection of publicly available information about a business is correlated with its risk of falling victim to different forms of data incidents. This further allows us to narrow down the recommendation on the most effective preventive measures, depending on the types of incidents the organization is most likely to face.

To this end, we use an incident dataset collected by the VERIS community [22] reporting a broad class of data incidents; these reports consist of detailed information about the incident itself (e.g. type of attack, assets involved), as well as the victim organization (e.g. business sector, number of employees). We combine these with statistics obtained from Alexa Web Information Service [3] about the website of the victim organization. The business information obtained from VCDB and the website statistics from AWIS together constitute the *business details* of the victim organization. We then consider three different categorizations for the incidents: (1) by type of data incident (e.g. error, hacking, etc), (2) based on the source of the incident (external, internal, or partner) and the motive behind it, and (3) by considering the assets that were involved in the incident (e.g. media, server, etc). Our results show that there is a clear correlation between each incident category and the victim's business details; this information can be used to provide guidelines on how an organization with limited budget for security should prioritize its security investment in allocating resources to different forms of self-protection.

We note that while this type of correlational study has been done before, most notably using business sector information, see e.g., Verizon's annual Data Breach Investigations Report [25], our goal is to use additional business information to enable a more fine-grained study, whereby the incident type distribution is quantified not just for an entire business sector, but a more refined definition based on other features such as employee size, region of operation, etc. This allows us to generate sharper (more highly concentrated) incident type distributions; that is, with more fine-grained definition of subsets within a sector, we are able to see incidents concentrated over a smaller number of types. An immediate consequence of this is that security investment and resource allocation decisions informed by such analysis are much more targeted and effective. We show that on average an organization can protect against 90% of all incidents by focusing on 70% of incident types; in some cases the latter can be significantly lower.

Our results can also be viewed as a type of prediction of the conditional distribution of incident types given that an incident occurs; this complements existing work on estimating the probability of an incident happening in the first place. In practice, the absolute risk of experiencing an incident provides the organization with insight on the total amount of resources that should be allocated to self-protection, while the conditional risk can be used to decide the allotment of these resources to different forms of preventive measures. In addition, the current study can guide better breach detection efforts. From this perspective, our study is aligned with the growing "assume breach" mentality in the security community [12, 8, 24]: everyone is a target hence all organizations should take measures to prevent, detect, and respond to incidents, in the most effective way. Last but not least, these findings can be used as guidelines in the emerging cyber-insurance market. A study of the distribution of risk among different forms of data incidents can help insurance providers better assess the potential amount of loss which in turn helps determine the contract terms, including premiums and coverage levels.

The rest of the paper is organized as follows. In Section 2 we describe the datasets used in this paper. In Section 3 we summarize existing work relevant to this study. In Section 4 we explain in detail how we build our risk assessment model, and we discuss and analyze the results in Section 5. Section 6 concludes.

2 Dataset

In this section, we illustrate the two datasets used in our study, namely the VERIS Community Database (VCDB) [22] and the Alexa Web Information Service (AWIS) [3].

2.1 VERIS Community Database

The VCDB is comprised of 4786 reports on publicly disclosed data breaches. The dataset includes incidents that occurred up to and including 2014, with 4526 entries corresponding to incidents after 2010. For our current study, we focus only on the 2013 and 2014 incidents, consisting of 1729 and 592 entries, respectively. The reports cover a wide variety of events, some examples of which are given in Table 1.

Each entry in the VCDB is reported using the Vocabulary for Event Recording and Incident Sharing (VERIS) [23]. The VERIS framework, as well as the VCDB, are initiatives by the Verizon RISK Team facilitating a unified approach to documenting and collecting security incidents. The VERIS fields for an incident are populated to answer "who did what to what (or whom) with what result?" [24]; details include the type of incident and the means by which it took place, the actor and motive, the victim organization, the assets which were compromised, timeline of the incident, and links to news reports or blogs documenting the incident. However, each entry might be only partially populated, since victim organizations tend to not disclose all the details regarding the incident.

We now explain the fields extracted from VCDB which are of interest in training and testing our classifier. The first set is information regarding the type of attack, based on which each incident can be put in one of seven general categories: `environmental`, `error`, `hacking`, `malware`, `misuse`, `physical`, or `social`. Each type may include additional fields that can help further differentiate incidents of the type. For instance, a `physical` incident might be further categorized as theft or loss, while a `hacking` incident might be identified as a SQL injection or a brute force attack. The second set identifies the actor responsible for the incident, falling in one of three types: `external`, `internal`, or `partner`. The dataset may further include fields identifying the motive for each of these actor categories. The third set identifies the assets that were compromised during the incident. There are six possible asset types: `kiosk/terminal`, `media`, `network`, `people`, `server`, and `user device`.

Time	Report summary
Apr 13	Hackers breach website of Hong Kong police force and publish non-public data, deface webpage
Aug 13	A Lima, Ohio clinical psychologist is in the process of notifying clients that their office was robbed
Sep 13	Pharmacy accidentally dumped hundreds of private medical records at a recycling depot
Sep 13	Janitor is blackmailed into gathering documents from a court
Sep 13	Parents of children at Hopkins Road Elementary Schools say their kids came home with sensitive data belonging to other students
Dec 13	Multiple Brazilian government sites defaced by Anonymous in protest to upcoming FIFA World Cup
Jan 14	Hacking group DERP launches DDoS against Xbox Live networks
May 14	Someone hacked into an electronic traffic sign on Van Ness Avenue in San Francisco, posting alerts that said “Godzilla Attack” and “Turn Back”.
Jul 14	Anonymous takes down 1,000 Israeli government and business websites for #OpSaveGaza

Table 1: Incident examples from the VERIS Community Database.

We also extract three features about the victim organization from the existing VCDB fields as input for our classifier: industry code, number of employees, and the region of operation of the victim organization. The industry code provided is the North American Industry Classification System (NAICS) code [13] for the victim, which specifies the organization’s primary economic activity. Although NAICS codes can extend to up to 6 digits, each further detailing the sector, we only extract the first two digits of the code for our incidents; this classifies the company as one of 25 different sectors. The employee count captures information about the size of the organization; this entry may be a numeric range (1-10, 11-100, 101-1000, 1001-10000, 10001-25000, 25001-50000, 50001-100000, and over 100000), or simply *small* or *large* (for approximately below or over 1000 employees, respectively) when an exact number is not available. Finally, we use the region of the organization as a feature by extracting the continent of operation for the victim. Note that any said features can be missing for a VCDB entry. In such cases, we generally add an additional *unknown* category.

2.2 Alexa Web Information Service

AWIS is a service offered by Amazon Web Services (AWS) [6] that provides information and statistics about websites; these include traffic volume, number of visitors, speed, number of pages linking to the website, and information about the organization that maintains the website, such as address, contact information, and stock ticker symbol.

We gather the following data from AWIS about the victim organization. We include the global and regional rank, as well as the number of pages linking in to the target website, as indicators of the popularity or familiarity of an organization. We also include the 30 day average and standard deviation of the website’s global rank for a one month period before the incident, to identify recent trends in popularity. Other selected features include speed of the website (as a percentile compared to other websites), the age and locale of the website, the top three categories associated with it, and whether the underlying company is publicly traded in the stock market. The aforementioned attributes of an organization can provide further insight into its sector, region, familiarity, size and network size. By combining these with features obtained from the VCDB, we are able to build a detailed description of a business, which can in turn help identify its risk distribution over different incident types.

2.3 Pre-processing

To be able to combine these datasets for our study, we first have to match each incident report with the website of the victim organization. To obtain this information, we find the name of the victim organization through the *victim_id* field in VCDB, and extract the first Google search result for the organization name. We then manually verify the results to ensure that the websites match the victim organizations. For ambiguous victim IDs (e.g. Indian Government Website), we further read the incident report provided by a news report or blog entry to find the website of the entity that suffered the data breach. For the 2322 incidents that occurred in 2013 and 2014, we were able to extract the website for 2065 of them, giving us 1688 unique domains. The mapping between a victim organization and its respective website will allow us to combine entries in the VCDB with data collected from AWIS. Note that for a given year, we omit duplicate incidents for each organization. As an example, there are over 200 entries in the VCDB corresponding to error incidents in the United States Department of Veterans Affairs. We count all of these incident only two times, once in 2013 and once in 2014. If there are additional entries corresponding to other forms of data incidents (e.g. hacking), we include them as well.

Finally, we also use the size of the network hosting a victim organization’s website as an additional feature in our classifier. To identify this network, we use whois information gathered by the top Regional Internet Registries (RIR) including AFRINIC [2], APNIC [4], ARIN [5], LACNIC [10], and RIPE [16]. By looking up the IP address of a website in these databases, we can find the IP address block/prefix that they belong to. These prefixes identify the smallest unit that includes the website IP and that has been allocated by the RIR to an organization. We then look up the owner ID for each of these prefixes, and find the total number of IP addresses registered to the same ID as the organization's network size. We note that, however, these prefixes cannot always identify the network of the business in question. For instance, when a business is utilizing a 3rd party hosting company to maintain its website, the identified network may potentially include addresses belonging to other organizations as well. In such cases, the network size is not necessarily representative of the victim organization's network size; nevertheless, we use it as a best estimate given the available data.

3 Related Work

The main contribution of this study compared to existing literature is an in-depth and quantitative analysis of the risk distribution over security incident types for a given organization, which can help the latter more strategically allocate resources for prediction, prevention and detection.

Data analysis: The most relevant study to this paper is Verizon's annual Data Breach Investigations Report. The most recent report for 2014 [25] contains detailed analysis on more that 63,000 security incidents from multiple sources including VCDB. The report contains a detailed analysis on statistics of the data including action types and vectors, actor types and motives, and assets involved, as well as victim demographics, industry, and organization size. Moreover, the authors identify nine patterns describing 92% of the incidents in their report. By categorizing the incidents into separate patterns, it is possible to analyze the distribution of incident varieties within each pattern and provide entities with a more specific recommendations on how to invest in their security. The report also provides the spread of attack patterns within each industry, to narrow down the risk even more. For instance, it is pointed out that the main threat to organizations providing accommodation services is through POS intrusion, which describes 75 percent of the incident reports within this industry. Furthermore, Thonnard et al. perform a similar analysis on spear phishing targeted attacks in [21]. The authors identify risk factors at the organization level (industry sector and number of employees), and individual level (job level and type, location, and number of LinkedIn connections), that are positively or negatively correlated with the risk of experiencing targeted attacks.

As mentioned earlier, compared to [25], we aim to provide a more fine-grained framework to give more specific guidance to organizations not only based on their industry, but utilizing a host of other features available to us. For incidents studied herein, this includes demographic information, details about the size of the business and its popularity, and business sector information. Moreover, compared to [21], our goal is to study a broader range of data incidents, including targeted and untargeted physical and cyber attacks from both internal and external sources, and incidents due to error.

Prediction of cyber incidents: The notion of predicting cyber incidents (rather than detection) has also enjoyed popularity recently. In [20], Soska el al. apply machine learning tools to predict the chance of a website turning malicious in the future, and show that their method can achieve 67% true positive and 17% false positive. In our recent study [11], we examine to what degree security incidents may be predicted by using a range of security posture data. In contrast to the above two studies, the present work is much more specific to incident types, with an emphasis on the relative risk each incident type poses to a particular organization.

Other works closely related to this paper include studies on the correlation between data breach incidents and market value [1, 9, 7]. Moreover, in [17] Romasky et. al. provide an empirical analysis of data breach litigation, and in [18] discuss the impact of breach disclosure laws on identity theft.

4 Methodology

In this section, we discuss in detail how to build a risk assessment model using the business features and incident reports described in Section 2.

4.1 Construction of the classifiers

Our ultimate goal is to provide risk assessment for an arbitrary organization given its features, i.e. a conditional distribution of risk over all incident types. In this context, the raw output of a decision tree classifier can be used as a means to density estimation. Toward this end, we use Random Forest classifiers, an ensemble learning method that constructs multiple decision trees over the training data, and outputs the average of all individual trees' predictions [15]. Random forest classifiers improve upon single decision trees by reducing over-fitting over the training set.

A naive way to build a risk assessment model is to take the incident signature (i.e. action, actor, and asset) of an entry as a class label, and the victim business features as input data for the classifier. However, given the large number of possible incident signatures, there are only a small number of samples per signature vector. Furthermore, as we have mentioned before, a significant number of incident entries provide only partial information about

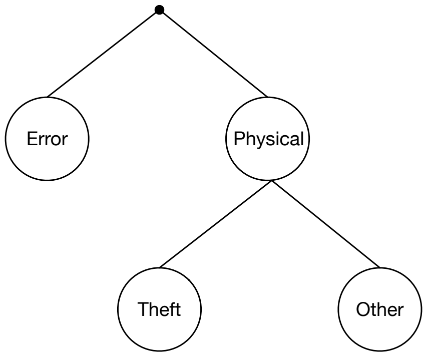


Figure 1: A sample risk assessment tree.

their corresponding incident. Ignoring such entries will leave us with even fewer samples.

Our solution to the above problem is to build multiple classifiers, each of them estimating a portion of the incident signature. An interpretation of this technique is the chain rule in probability. Assume that we want to estimate the risk factor for an organization of type t for experiencing a physical theft incident. We can break this risk into two parts as follows:

$$\Pr(\text{Theft} \mid t) = \Pr(\text{Physical} \mid t)\Pr(\text{Theft} \mid \text{Physical}, t)$$

As a result, entries that cite a physical incident without specifying additional details will still be included for building and testing the first classifier (first term in the RHS of the equation), but will be ignored when building the second classifier (i.e. theft). This method can be visualized as a tree as shown in Figure 1, where each node represents a data breach type. The risk score at a node is the result of multiplying the risk at its parent node by the output of the classifier corresponding to said (child) node.

Given the training and test samples (incidents belonging to 2013 and 2014, respectively), we first train a binary classifier for each node, using a Random Forest model consisting of 20 trees. To prevent over-fitting, we set the minimum number of samples at each leaf of the decision trees to 25. However, we may still experience some over-fitting due to the large number of features available to our classifier. To help alleviate this problem, we limit the number of features used for each Random Forest as follows: we always use the three features extracted from the VCDB, namely industry, employee count, and region. Out of the remaining 10 features, we select the most significant through cross validation, i.e. training multiple classifiers using different combinations of features, and selecting the one with the best performance. The list of features used for each classifier, as well as their importance in the resulting Random Forest classifiers, are included in Table 6 in the Appendix.

4.2 Incident categorization

Using the classification method described above, we apply our risk assessment scheme separately to three parts of the incident signatures: action, actor, and asset. Each of these classifiers focuses on a separate aspect of an incident.

Action type The action type falls into one of the seven general categories discussed in Section 2.1. We omit `environmental` incidents, of which there are only 4 samples between 2013 and 2014. We further categorize `hacking` events into two sub-categories: (1) hacking incidents that involve data breach through compromised credentials, including stolen credential, brute force, and backdoor attacks, and (2) all other forms of hacking, 75% of which are SQL injection and Denial of Service attacks. We also divide `physical` incidents into two subcategories of (1) theft and (2) everything else, 88% of which are due to tampering.

Knowing the action type can provide significant information on the types of preventive measures that can be used to reduce loss. For instance, the first group of `hacking` incidents can be prevented by setting strong passwords and changing them on a regular basis, as well as not storing unencrypted credentials at insecure locations. `Error` and `misuse` can be reduced by employee education, setting and enforcing internal regulations, and avoiding unnecessary access privileges for employees and/or business partners.

Actor type and motive In addition to action types, we train our classifier based on the actor responsible for the incident. `Internal` actors are separated based on their motive into two sub-categories of (1) financial motives, and (2) other motives, including convenience, espionage, grudge, ideology, and fun. `External` actors are similarly sub-categorized into (1) financial, (2) espionage, (3) ideology, and (4) fear, fun and grudge. Incidents due to `partners` are not further sub-categorized due to insufficient samples.

Assessing risk associated with actor types can prompt organizations to determine policies for employee education and access to data (for `internal` types), guard their network periphery from `external` attackers, and perform due diligence when selecting `partners`.

Asset type Finally, we look at the types of assets that were compromised during the incident. Asset types include `kiosk/terminal`, `media`, `people`, `server`, and `user device`. We have omitted `network` related assets due to insufficient number of samples. Knowing what asset types are more likely to be affected can significantly improve our ability to estimate the amount of potential

Incident type	Crimeware	Cyber Esp.	DDos	Stolen Cred.	Error	Skimmers	PoS	Misuse	Web app	Else
# of samples	67	16	106	326	333	66	19	272	399	356

Table 2: VCDB data categorized using DBIR 2014 patterns. Only 82% of the data can be described by the 9 patterns.

loss following security incidents. This can guide insurance underwriters in designing more appropriate policies catered to specific client organizations. It can also be used to advice network administrators to keep regular backups when assets such as `media` and `server` are involved.

4.2.1 Comparison with DBIRs' categorizations

Our choice of categorizations is consistent with the one adopted by Verizon in the 2008-2013 DBIRs, but differs from the categorizations proposed in their latest 2014 report [25]. DBIR 2014 uses hierarchical clustering to identify 9 incident classification patterns (combinations of actions, assets, and actors) that can be used to describe 94% of all incidents. Examples of these patterns include cyber-espionage, point of sale intrusions, and insider misuse. Despite the effectiveness of this clustering method in accurately describing incidents in the dataset used by Verizon, an application to the subset available through VCDB would fail to provide a similar precision, see Table 2: due to lack of sufficient details, 18% of the VCDB data will not fit the 9 proposed patterns (as opposed to only 6% in Verizon's larger dataset). This is one of our main motivations for selecting three different categorizations based on VERIS primitives only, i.e., actions, actors, and assets.

5 Results

5.1 Risk distributions

To gain insights on how details about a business can affect their risk of experiencing various types of data breach, we start by deriving the distribution of risk over incident action types for each industry sector. The results for 9 business sectors, as well as the overall distribution are included in Table 3; these results use only sector information in training the corresponding classifiers. Note that this is equivalent to simply measuring the distribution of incidents in each sector, since the Random Forest classifier is using only a single feature. There are a few observations on the risk distribution of different sectors. For instance, information companies are more prone to both types of hacking, and less likely to sustain damage due to physical incidents. In contrast, the healthcare industry has low risk in hacking but high risk in physical attacks, especially theft. These observations are intuitively to be expected, since information companies' most valuable assets are generally stored in non-physical formats (e.g. on the cloud), while the healthcare industry may still use physical forms of archiving sensitive data such as patient information.

To highlight the additional gain we get by using more features than just industry sector information, we also show in Table 3 a number of examples. In these cases our classifiers can generate much more specific risk predictions. For instance, we can see that compared to a typical information company, Russian Radio has less risk in malware, social, and hacking through compromised credentials, but higher risk in error, misuse, and physical. Verizon and Macon-Bibb County exhibit a more uniform risk across the board. The higher risk for Verizon in error and misuse (also the lower risk of Macon-Bibb County in the same categories) can be attributed to their respective sizes. As the number of employees grows larger, so does the risk of data incidents due to human error and malevolent employees. These much more refined and targeted predictions would not be possible without using additional features. As we shall show later in Section 5.4, with proper thresholding the actual incidents in these organizations were also correctly forecasted.

5.2 Dealing with rare events and reporting bias

Looking at Table 3, there is an imbalance in the overall frequency at which different incident types appear in our dataset. Social incidents occur rarely as compared to error and hacking incidents. It is indeed possible that social incidents are rare events, and therefore should not be a priority when determining security policies. However, an important challenge in building a risk assessment model is under-reporting of security incidents by victims. Data breach reports are largely undisclosed, as organizations tend not to expose their security posture information unless necessary. Our dataset, VCDB, is a collection of publicly disclosed breaches; these incidents have either been detected by external sources (e.g. website defacement), or are incidents which an organization is obligated to report due to the compromise of private customer information (e.g. payment information or health records). Thus, not only incidents are commonly underreported, but it is also safe to assume the existence of selection bias in the data: each incident type is represented differently as a result of both availability and variation

	Error	Hacking		Malware	Misuse	Physical		Social
		Comp. Cred.	Other			Theft	Other	
Overall	0.22	0.12	0.21	0.06	0.15	0.14	0.04	0.04
Manufacturing	0.08	0.09	0.33	0.13	0.22	0.13	0.00	0.02
Retail Trade	0.15	0.26	0.11	0.19	0.09	0.09	0.11	0.02
Information	0.09	0.28	0.41	0.07	0.04	0.03	0.01	0.07
Russian Radio	0.14	0.16	0.40	0.02	0.10	0.10	0.03	0.03
Verizon	0.28	0.17	0.22	0.08	0.19	0.06	0.05	0.05
Finance & Insurance	0.25	0.09	0.11	0.05	0.12	0.10	0.19	0.07
Pro., Sci. & Tech. Svcs	0.16	0.09	0.56	0.04	0.13	0.09	0.00	0.02
Educational Svcs	0.30	0.13	0.21	0.06	0.11	0.14	0.00	0.05
Health Care & Social Asst	0.25	0.08	0.03	0.02	0.23	0.38	0.02	0.01
Accommodation & Food Svcs	0.08	0.37	0.00	0.18	0.16	0.11	0.11	0.00
Public Administration	0.27	0.09	0.29	0.03	0.17	0.10	0.01	0.03
Internal Revenue Service	0.21	0.08	0.15	0.06	0.17	0.09	0.02	0.03
Macon-Bibb County	0.20	0.13	0.23	0.07	0.14	0.23	0.04	0.04

Table 3: Conditional risk distribution by business sector,
and for sample organizations (highlighted rows).

of detection methods, and the corresponding industries' disclosure policies. This bias could cause a tendency towards flagging and protecting from incidents that are reported more often, in turn resulting in poor protection against less commonly reported incidents.

One way to address this issue is to ignore the frequency at which incident types are reported. In other words, rather than looking at each row in Table 3, we could base our decisions on the distribution of risk within each column. For instance, we can make the observation that finance and insurance companies exhibit higher than average risk in social incidents, even though the absolute risk in this category is the second lowest in its respective row. By having different standards, or thresholds, of what signifies high risk in each category, we can alleviate the impact of potential under-reporting and report ing bias in the dataset and prevent the tendency of ignoring rare events by ensuring equal protection among all incident types. Specifically, after training our classifiers and obtaining risk outputs on the input data, we specify thresholds for each incident type separately, such that the reduction in risk is consistent among all types; this is detailed in the next section. Note that this *normalization* of risk scores is possible mainly due to the fact that we are constructing separate classifier for each incident type.

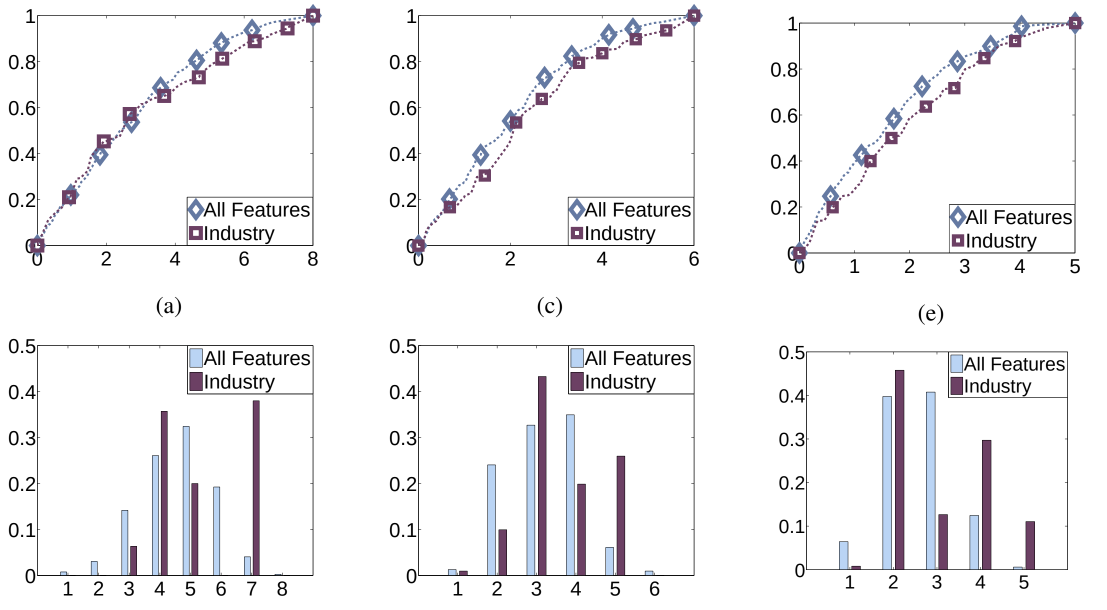
5.3 Interpreting the classifier output

After estimating an organization's risk in each category by feeding its features into our classifier, the next step is to interpret these scores by determining what range of values indicate heightened risk. Based on our discussion in the previous section, this is achieved by computing the ROC curve for each binary classifier on the training set, and choosing the point that corresponds to a predefined true positive rate. We will use the family of thresholds corresponding to these points to determine risky incident types for any arbitrary organization, hereafter referred to as the *risk profile*. Selecting a more conservative set of thresholds (i.e. higher true positive rate) will tighten the business's security by advising it to invest in a larger set of self-protection methods. This selection represents the trade-off between the amount of resources an organization allocates to self-protection, and the reduction in incidents it desires to attain. From this point on when referring to *thresholds* used for deriving the risk profile, we simply mean the family of thresholds acquired for a specific true positive rate.

5.4 Evaluation

For evaluation, we train our classifiers over 2013 incidents, and test them on the 2014 data. We first obtain the risk profiles of organizations in our test samples, for various sets of thresholds. We then calculate the accuracy of our risk assessment model, by counting the number of incidents which belong to one of the risky types forecasted by the risk profiles. An important advantage of our model is in reducing the number of risky types predicted for each organization; achieving the same accuracy by advising organizations to focus on a smaller set of incident types will help achieve the same level of protection by spending less resources on security.

Figures 2a, 2c, 2e summarize our results over action,



(a) (c) (e)

(b) Action types

(d) Actor types and motives

(f) Asset types

Figure 2: Detection rate vs. average number of risky types (top), and distribution of organizations over the number of types in their risk profiles (bottom).

Organization	Error	Hacking		Malware	Misuse	Physical		Social
		Comp. Cred.	Other			Theft	Other	
Information		▨	▨	▨				▨
Russian Radio		▨	▨ ×					
Verizon	▨	▨	▨ ×	▨	▨		▨	▨
Public Administration	▨	▨	▨	▨	▨	▨		▨
Macon-Bibb County	▨ ×	▨	▨		▨	▨	▨	▨
Internal Revenue Service	▨				▨ ×			

Table 4: Risk profiles for different sample organizations, and their corresponding industries' profiles. Gray cells signify incident types with high risk, and crosses indicate the actual incident that occurred.

actor, and asset types, respectively. Each point in the plot denotes the accuracy of risk profiles obtained from a particular set of thresholds, versus the average number of risky types forecasted by these profiles. To illustrate the improved performance of using our extended set of features, we have also included the accuracy curve of a predictor using industry information alone (see Table 3). For action, actor, and asset types we can correctly forecast 90% of the incidents in our dataset by flagging, on average, 5.6 (70% of incident types), 4.0 (67%), and 3.5 (70%) incident types, respectively. In other words, we can achieve this accuracy by eliminating at least 30% of all incident types. Using only business sector information, the numbers increase to 6.5 (81%), 4.8 (80%), and 3.6 (72%). The distinction is more visible when predicting over action and actor types.

Note that for a given point in the plot, the number of risky types in the risk profile can vary across organizations. Figures 2b, 2d, and 2f demonstrate the distribution of organizations over their predicted number of risky types, corresponding to the 80% accuracy point in the top plots. Looking at Figure 2b we can see that using all features, there are organizations whose risk profiles only consist of 1 or 2 incidents types, while others include up to 7 types.

We present a number of these samples in Table 4, whose risk scores have already been discussed in Table 3. The first two examples in the table belong to the information sector, and the last two are public administration organizations. We have included the risk profiles for these

Industry (Number of Samples)	Error	Hacking		Malware	Misuse	Physical		Social
		Comp. Cred.	Other			Theft	Other	
Overall (1426)	61.6	61.9	37.5	32.4	56.9	51.5	38.1	38.9
Manufacturing (39)	30.8	97.4	51.3	89.7	33.3	28.2	76.9	41.0
Retail Trade (63)	34.9	100.0	46.0	76.2	42.9	9.5	68.3	23.8
Information								
Small (49)	22.5	100.0	100.0	65.3	12.2	8.2	38.8	59.2
Large (41)	36.6	100.0	80.5	70.7	36.6	0.0	51.2	87.8
Finance & Insurance								
Small (53)	66.0	62.3	18.9	75.5	18.9	34.0	75.5	60.4
Large (91)	64.8	41.8	29.7	31.9	67.0	49.4	86.8	75.8
Pro., Sci. & Tech. Svcs (44)	54.6	72.7	27.3	50.0	27.3	45.5	36.4	43.2
Educational Svcs								
Small (27)	81.5	44.4	14.8	63.0	40.7	92.6	25.9	33.3
Large (46)	89.1	34.8	2.2	19.6	41.3	82.6	41.3	26.1
Health Care & Social Asst								
Small (97)	59.8	28.9	7.2	22.7	54.6	95.9	46.4	10.3
Large (97)	93.8	10.3	3.1	7.2	96.9	96.9	42.3	24.7
Accommodation & Food Svcs (33)	72.7	6.1	15.1	48.5	87.9	78.8	54.6	9.1
Public Administration								
Small (41)	95.4	85.4	24.4	22.0	63.4	51.2	9.8	19.5
Large (96)	97.9	32.3	10.4	2.1	93.8	67.7	0.0	55.2

Table 5: Average risk profiles by business sector and size.

sample organizations using our extended feature set, as well as the risk profile using only industry. For the information sector, the latter recommends focusing on both types of hacking, as well as social incidents, whereas for public administration it deems all but the second type of physical incidents risky. By contrast, using our extended feature set, we are able to eliminate malware and social incidents as likely threats for Russian Radio, and still provide an accurate risk profile. Similarly for the Internal Revenue Service we are able to narrow down the list of threats to two types without losing accuracy. Macon-Bibb County and Verizon are assessed to have a broad range of risks, more so than their respective industry average would suggest; this highlights that for these organizations they may be attacked on multiple fronts, which may call for a different type of resource allocation strategy. The point is that this type of fine-grained prediction is much more specific to an organization itself rather than using the industry average as a proxy. We also note that in all these cases our risk profile correctly captured the actual incident occurrences (as indicated by an “×”).

It is worth noting that the grey cells in Table 4 not marked with an “×” are incident types deemed likely by our classifier but unrealized in reality (not observed in our dataset). These should not be viewed as discrepancy; rather, the relationship between a predicted risk profile and actual incident occurrence is analogous to that between a dice with a certain probability of turning up each side and the outcome of tossing the dice in a particular random trial. In other words, in the example of the Internal Revenue Service, even though misuse is the only incident that actually occurred, the result suggests that an error event could just as well have happened. This is because in essence our classification constructs risk profiles by extracting details about a business and examining actual incidents that have occurred to other, *similar* companies. In this case, for organizations that share the same business model as the Internal Revenue Service, error and misuse constitute the majority of data breach reports; thus given the information available to us, both incident types are regarded risky.

To close this section, we display the average risk profile over action types of all organizations, as well as average risk profiles over action types for different industry sectors and sizes in Table 5. Each number in the table represents the percentage of organizations, for whom the respective incident type is deemed risky. For instance, 61.9% of all organizations have high risk in hacking incidents due to compromised credentials. However, for 100% of organizations in the information sector this type of hacking poses a high threat. The risk profiles are obtained for the 80% accuracy point in Figure 2a.

We highlight a number of trends in Table 5. As discussed previously, large companies tend to have higher risk in error and misuse. Sectors that are more prone to error include large healthcare, and both small and large public administration. Large healthcare and large public administration companies also run a high risk of misuse. Error incidents exhibit a substantial presence in all business types, the minimum being 21.2% for information companies. Note that overall, all of the incident types are flagged for at least 30% of our samples, even tough their occurrence rate is widely different as evidenced in the first row of Table 3. This is due to our choice of ignoring the a priori distribution of incidents, as explained in detail in Section 5.2.

Comparing Tables 4 and 5 can help provide some insight on how having additional features has helped eliminate (or introduce) possible risks for those sample organizations. For instance small information companies tend to have lower risk in social incidents, and this has helped us eliminate this category as a possible threat for Russian Radio. We can also see that small public administration and large information companies have a more uniform risk among all types, attributed to the risk profiles for Macon-Bibb County and Verizon, respectively. The Internal Revenue Service, a large public information company, is expected to have less risk in the second type of physical incidents, as well as hacking and malware. Note that one cannot completely explain the generated risk profiles by only looking at business sector and size information alone, as they are a result of analyzing the dataset's distribution over all the features in Table 6. For instance large public administration organizations tend to have higher risk in social events than small ones, even though this incident type has been flagged for Macon-Bibb County and not the IRS. In this case, other features of the IRS have contributed to it having lower risk in social incidents.

6 Conclusion

Our results demonstrate how, and to what extent, can business details about an organization help forecast its relative risk of experiencing different types of data incidents. We observe that even though there is notable correlation between organization features and the incident signatures in our dataset, it is impossible to assert with certainty the types of incident an organization is likely to face. We acknowledge the fact that there is an inherent randomness in incidents suffered by organizations: no business is prone to a single type of incident. As observed in our results, while risk in incidents such as hacking and theft may vary largely across sectors, any organization is likely to experience incidents due to miscellaneous errors. Nonetheless, feeding further information into our classifiers may help construct more accurate risk profiles. The feature set used in this paper provides only high level information about the organization itself, and not its security posture. Even though these features are the easiest to obtain, as they all are publicly available, further information indicative of an organization's security policies will undoubtedly help narrow down its risk profile. Externally observable signals (such as the ones used in [11]), as well as inside information, may be used to infer a business's security posture.

It is worth noting that incident types are often too ambiguous to act upon for a security unaware business operator, hence the need for explicit, actionable security recommendations. Note that there indeed exist frameworks providing such recommendations. For example, the SANS institute's critical security controls [19] is comprised of 20 categories of security controls, each describing a specific action or policy that can be implemented by a business in order to raise its security levels. Verizon uses this framework to provide general security recommendations in its annual Data Breach Investigations Report, and the SANS institute offers a partial mapping between these controls and the VERIS incident categorizations. Translating our risk profiles into actionable security recommendations is a direction for future work. Furthermore, our current dataset does not contain information on the monetary impact of each incident type. Obtaining such information, and combining it with the cost of protection for each incident type, will allow us to provide more economically-informed recommendations.

References

[1] Acquisti, A., Friedman, A., and Telang, R. Is there a cost to privacy breaches? an event study. *ICIS 2006 Proceedings* (2006), 94.

[2] AFRINIC Whois. `http://www.afrinic.net/services/whois-query`.

[3] Alexa Web Information Service. `http://aws.amazon.com/awis`.

[4] APNIC Whois. `http://wq.apnic.net/apnic-bin/whois.pl`.

[5] ARIN Whois. `https://www.arin.net/resources/request/bulkwhois.html`.

[6] Amazon Web Services (AWS). `http://aws.amazon.com`.

[7] Gordon, L. A., Loeb, M. P., and Zhou, L. The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security 19*, 1 (2011), 33–56.

[8] HINES, C. Why companies must adopt the "assume mentality" when it comes to breaches. https://blog.cloudsecurityalliance.org/2015/02/27/why-companies-must-adopt-the-assume-mentality-when-it-comes-to-breaches.

[9] KANNAN, K., REES, J., AND SRIDHAR, S. Market reactions to information security breach announcements: An empirical analysis. *International Journal of Electronic Commerce 12*, 1 (2007), 69–91.

[10] LACNIC Whois. http://lacnic.net/cgi-bin/lacnic/whois.

[11] LIU, Y., SARABI, A., ZHANG, J., NAGHIZADEH ARDABILI, P., KARIR, M., BAILEY, M., AND LIU, M. Cloudy with a chance of breach: Forecasting cyber security incidents. In *USENIX Security Symposium* (2015).

[12] LOS, R. "Assume breach" is not a defeatist point of view. http://blog.norsecorp.com/2015/02/02/assume-breach-is-not-a-defeatist-point-of-view.

[13] NAICS Association. http://www.naics.com.

[14] PASSERI, P. Hackmageddon. http://hackmageddon.com.

[15] Random Forest Classifier. http://scikit-learn.org/stable/modules/ensemble.html.

[16] RIPE Whois. https://apps.db.ripe.net/search/query.html.

[17] ROMANOSKY, S., HOFFMAN, D., AND ACQUISTI, A. Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies 11*, 1 (2014), 74–104.

[18] ROMANOSKY, S., TELANG, R., AND ACQUISTI, A. Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management 30*, 2 (2011), 256–286.

[19] SANS Institute Critical Security Controls. https://www.sans.org/critical-security-controls.

[20] SOSKA, K., AND CHRISTIN, N. Automatically detecting vulnerable websites before they turn malicious. In *USENIX Security Symposium* (2014).

[21] THONNARD, O., BILGE, L., KASHYAP, A., AND LEE, M. Are you at risk? profiling organizations and individuals subject to targeted attacks. In *Financial Cryptography and Data Security* (2015).

[22] VERIS Community Database (VCDB). http://vcdb.org.

[23] The VERIS Framework. http://veriscommunity.net.

[24] VERIZON. Data Breach Investigations Reports (DBIR) 2013.

[25] VERIZON. Data Breach Investigations Reports (DBIR) 2014.

[26] The web application security consortium's Web-Hacking-Incident-Database. http://projects.webappsec.org/w/page/13246995/Web-Hacking-Incident-Database.

Appendix

	Industry	Employee Count	Region	Rank	Local Rank	Rank History	Links In	Website Age	Speed	Locale	Traded	Category	Network Size
Action													
Error	21.4	25.2	18.8	x	x	x	x	9	x	x	5.7	x	19.9
Hacking	27.8	9	29.2	8.7	x	x	10.5	8.1	x	x	x	x	6.8
Comp. Cred.	0	25	8.3	x	x	x	x	16.7	25	x	x	8.3	16.7
Other	0	17.4	33.3	x	10.7	11.7	x	4.6	10.6	x	x	x	11.6
Malware	20.5	8.2	4.2	x	13	33	x	x	7.7	1.8	x	x	11.5
Misuse	17.4	9.7	6.9	24.2	x	x	19.5	9.3	x	11.4	1.6	x	x
Physical	11.3	3	7.6	x	x	33.1	6.1	x	5.6	x	0.4	33	x
Theft	26.4	0.5	2	x	x	38.7	x	6.4	6.9	x	1.9	x	17.2
Other	24.9	9.6	4.1	x	x	x	16.1	24.9	x	x	x	x	20.4
Social	14.2	21.4	18.9	x	18.8	x	x	x	26.8	x	x	x	x
Actor													
External	28.9	7.1	11.7	15.4	x	x	x	6.1	x	17.4	1.8	x	11.6
Financial	12.7	13.3	27.9	2.1	30.9	9.8	3.2	x	x	x	x	x	x
Ideology	18.7	38.5	25.8	6.8	x	x	4.1	x	6	x	x	x	x
Other	13.6	4.1	40.6	x	33.5	x	x	x	8.2	x	x	x	x
Internal	28.3	16.6	40.8	x	x	x	x	12.2	x	x	2	x	x
Financial	17.4	0	0	x	x	x	x	12.5	18	x	x	37.8	14.3
Other	8.3	0	0	x	x	37.4	18.3	x	x	x	x	20.4	15.6
Partner	19.3	11.8	12.2	x	x	x	x	16.5	x	5.3	x	22.3	12.6
Asset													
Kiosk/Terminal	13.3	11.7	5.1	x	x	9.9	1.9	x	2.9	x	0.9	54.4	x
Media	10.4	8.3	10.6	x	7.8	x	3.9	x	3.1	x	0.6	55.2	x
People	19.7	15.4	24.7	x	x	x	x	x	28.9	10.5	0.7	x	x
Server	15.7	3.1	17.6	x	13	11.9	x	3.7	x	2.2	1.6	27.2	3.9
User Device	8.3	5.5	7.2	14.3	17.3	38.9	x	6	x	2.3	0.2	x	x

Table 6: Utilized features and feature importances for all classifiers.
Crosses indicate features that have not been used in training the corresponding classifier.