Collection of Quantitative Data on Security Incidents

Thomas Nowey
Department Management of Information Security
University of Regensburg
Germany
thomas.nowey@wiwi.uni-regensburg.de

Hannes Federrath
Department Management of Information Security
University of Regensburg
Germany
hannes.federrath@wiwi.uni-regensburg.de

Abstract

Quantitative data about security threats is a precondition for a precise assessment of security risks and consequently for an efficient management of information security. Currently such data is hardly available, especially for small and medium-sized organizations. In this paper we discuss different ways of gathering quantitative data and present a new approach for the collection of historical data on security incidents. We propose a platform that collects, aggregates and evaluates data on security incidents from multiple organizations. We identify basic requirements for such a platform and show approaches for satisfying them. We especially emphasize the aspects of security and fairness. Finally we introduce a prototype that shows how an implementation could look like.

1 Introduction

Organizations of all sizes face a growing need to determine security risks and to evaluate costs and benefits of possible security investments. Various factors intensify this trend. First, there is the growing importance of security in general because of the growing number of threats and attacks over the past years (see [7]). Second, due to the law of decreasing marginal utility, we know that there must be a utility maximizing bundle of security measures. So security managers are looking for that optimal bundle to invest in or in other words they seek to answer the question “How much is enough” ([12]). Finally there are compliance requirements. A variety of external regulations like Sarbanes-Oxley Act and Basel II stipulate that organizations are able to assess their risks, including security risks.

Analyzing costs and benefits of security investments and assessing security risks are corresponding tasks. The biggest part of benefits of security measures cannot be measured directly since their main utility lies in the reduction of risks. This explains why security management should be based on a “business risk approach” ([9]). Risk can be defined as the combination of the probability of an event and its consequence ([9]). The assessment of these two factors is the basis for the risk treatment step and consequently for every economic oriented evaluation of security. On higher levels this information can be aggregated in business ratios like ROSI, transferred into security scorecards and the like or be used as input for simulations.

Most existing approaches for security management and security risk management are based on best practice or at most on qualitative data and expert judgments. The disadvantage of best practice approaches is their inability to decide between different alternatives. Qualitative judgments at least enable rudimental comparisons between different alternatives, although they cannot be used to compare security investments to other investment alternatives. However, for precise risk analysis and sound decisions quantitative data is a prerequisite (see [12]).

Many approaches assume that there is quantitative data readily available without telling from what sources to get that data. Hence we will identify and evaluate different possible sources for quantitative data in section 2. In section 3 we develop and present the idea of a platform to share and collect historical data on security incidents. Afterwards we identify special requirements for such a platform in section 4. Finally we present a prototype implementation (section 5) of our concept and conclude by lining out challenges for the future.

2 Sources for quantitative data

2.1 The challenge

Different authors state that the quantitative data required for efficient security management is not available ([12],[13]). The problem can be outlined as follows: As
shown above the main utility of security measures lies in the reduction of risks. Consequently the data that is required in the field of security management is data on risks i.e. information on possible future security events with a negative outcome. Soo Hoo [12] identifies five key variables for which data is required:

- frequency of bad events,
- consequences of bad events,
- measures of safeguard efficacy,
- costs of implementing safeguards,
- and additional profits.

In our opinion the first two of them are the most challenging to determine. Needless to say that the future cannot be exactly predicted. As in other fields we have to use probabilities, probability distributions and expected values. Since those values cannot be directly measured the challenge is to find good mechanisms to make estimates of those parameters for the future.

These problems are well-known in other areas. In the field of security the collection of such data is especially challenging, because of the fact that it is an extremely complex and fast evolving field. To date there is to our knowledge no generally accepted way of collecting quantitative data and there are no such databases available.

Data that can be used for security management must fulfill various requirements. We propose the following criteria for data sources that are relevant for security management:

- data quality: precise values, good prediction quality, objective and un-biased, originally quantitative
- up-to-dateness: up-to-date data, periodic updates possible
- organization specific: consider parameters that are unique for the organization, like risk associated with custom made software
- completeness: consider all possible threats and risks, not restricted to the technical level
- practicability: easy to assess, working solutions

In the following we will analyse different data sources and discuss how good they adapt to these challenges.

2.2 Potential sources

We have identified four potential sources for quantitative data on security risks:

- **Questionnaires and other expert judgments.** Based on their knowledge and experiences internal or external experts can make predictions of a firm’s future risk exposure. Those judgments can be assessed systematically using a questionnaire or more freely in an expert interview.

- **Simulations.** Simulations are a technique frequently used in risk management to test the effects of different scenarios on future developments. They require some input data, like for example information on probability distributions. Historical simulations (see [10]) as well as Monte-Carlo Simulations (see [2]) have been proposed for the field of information security investments.

- **Historical data on security events/incidents.** In other domains, like for example the insurance sector, the use of historical data to predict the future has a long tradition. Insurance companies have huge collections of data on past events that is used to estimate risks for the future. Usually the data is extrapolated to adapt to assumed future developments.

- **Market mechanisms.** The lack of quantitative data on computer security (see [11]) was one of the reasons for the development of market mechanisms for information security. One idea is for example to issue so called exploit derivatives that can be traded like other stocks. The current quote of these securities could be used as an estimate for the probability of an exploit being available, which correlates with the probability of a security incident based on that exploit. For a profound discussion of market mechanisms see [1].

2.3 Evaluation

To our knowledge different forms of expert judgments are the most frequent way of generating data for risk assessment at present. Besides problems like bias and incompleteness the main argument against an exclusive use of expert judgments is the fact, that they are not originally quantitative. Expert estimations are usually qualitative and have to be transformed to quantitative data using some heuristic.

Simulations are a valuable concept for testing different scenarios of the future. But even the best simulations cannot be done with some input data that has to be determined from other sources. Consequently they should be used on top of a primary data source. For example by doing historical simulations on historical data one could overcome a part of the constraints of the latter.

There remain only two originally quantitative data sources: historical data and data from the capital market. Historical data is well approved in other areas. It cannot
only be used as a direct input for risk assessment but also as an input for simulations. Critics remark that in the field of security management, historical data is of limited value, because risks change so quickly that it is almost impossible to predict the future from past data. Up till now historical data is not broadly applied for security risk management purpose due to a lack of mechanisms for collecting the relevant information.

Market mechanisms are a relatively new approach and there are no practical implementations so far. On a theoretical level one can criticize that it is not possible to deduct probability and impact of an event directly from the market mechanisms that are known so far. Instead they only give an estimate for the value of an exploit. Besides that one has to be aware of the other imperfections of markets.

For market mechanisms as well as for historical data one has to trade off a broad data base against organization specific data. If data is collected only for a specific organization, it will also contain organization specific factors, but on the other hand the amount of data might be too small to be significant.

This inspection shows, that all of the possible sources have their individual shortcomings. Therefore we recommend a combined approach using data from different sources and to improve the existing ways of getting data from these sources.

The general problem with the evaluation of the above mentioned sources remains that they have to remain on a theoretical level since there are hardly any practical experiences. Especially there are no comparisons of the prediction quality of different approaches. We are convinced that the collection of historical data on security incidents can be a valuable tool in security risk management. But a large base with that kind of data is necessary to test this hypothesis and to compare it to other approaches. For that reasons we focus on developing a working solution for the collection of quantitative historical data on security incidents in the remainder of this paper.

3 The basic idea

As we have seen above, the use of historical data is well proven in other areas. It could also be useful in security management as a direct estimate and as an input for simulations. To our knowledge there is no database on security incidents that gives information about impact and frequency of security incidents. Of course many organizations record and sometimes also evaluate their security incidents and could use this data. But using data only from one single organization is not sufficient, especially because of so called HI/LF risks i.e. risks events that have a low probability of occurrence but a high, sometimes catastrophic, impact. However, especially that type of risk is extremely important, since they can threaten the continuity of the organization. Therefore an inter-organizational approach is necessary.

Especially in the US there have been efforts to promote the sharing of security information between organizations for several years ([6]). Unfortunately most of the resulting initiatives do not provide a technical platform and do not consider incentive schemes. Today inter-organizational data collection is frequently done by CERTS. Since their focus is more on early warning and countermeasures they are virtually not collecting information on damage and frequency. So currently there is no technical basis for the collection and sharing of such information.

Our goal is to establish a platform for the exchange of information on security incidents with a focus on the collection of data for risk management. Participating organizations will collect information on every security incident that occurs and submit them to a central platform. Within the platform the data from all participants will be harmonized and stored. The data will then be aggregated, analyzed and interpreted. The results are afterwards distributed to the participants either on special request or in regular reports. The organizations get data probabilities, distributions and expectation values and can thus improve their security management. The participants should also have the possibility to use this platform for the exchange of information and experiences. The platform provider can publish regular reports with aggregated data.

We consider especially small and medium enterprises as participants, since they have very limited capabilities to collect data on their own. The only alternative for them is to regularly buy expert opinions which is very costly. Expert interviews have shown that it is reasonable to assume, that firms that don’t see it as their key-competence are willing to offer some data in exchange for improving the efficiency of their information security measures if some requirements are fulfilled (see next section).

3.1 Architecture

The actors in the basic architecture are shown in Fig. 1:

- **Participants** are organizations that participate actively in the sharing of security information as a closed user group. They provide information on security incidents
and receive aggregated data for their security management.

- **The platform provider** is a party that is not actively participating in the information sharing. Its main task is to collect, harmonize, aggregate and interpret the data. It also provides and administers the technical infrastructure and sets the organizational framework.

- **External data providers** add additional information to improve the quality of the data in the platform.

**Figure 2. Main functions**

The main functions (as illustrated in Fig. 2) of the system are:

- **Submission**: Every time a security incident occurs within the organization the participants identify and assess the damages and submit the relevant meta-data electronically to the platform.

- **Collection and storage**: The platform provides an interface for the submission of incident data and stores the submissions. Possibly data from other sources is added in this step.

- **Harmonization**: The submitted data has to be standardized in order to make it comparable between organizations.

- **Aggregation**: The harmonized data will then be clustered and aggregated along different dimensions like for example industry of company, type of incident, etc.

- **Interpretation**: In this step the data is evaluated. The goal is to determine likelihood and severity of different types of incidents and their influencing factors.

- **Distribution**: After aggregating and analyzing the data, the platform manager makes the results available to the participants. This can happen in form of regular reports or by providing an interface for data analysis. Another possible way of information distribution would be to provide detailed information on single incidents, thus giving participants the possibility to find other organizations who have previously encountered the same problem.

### 3.2 Microeconomic foundation

Interviews with different companies have shown that there is a strong interest and effort to participate in such a platform. Besides that we can also find some theoretical findings in microeconomics.

Gordon, Loeb and Lucyshyn [5] have identified various economic questions that arise in any “organizational arrangement focused on the sharing information related to security breaches”. They draw parallels to the fields of trade associations and research joint ventures and conclude that the development of a microeconomic model for this case is necessary. In [6] they develop such a model for the analysis of information sharing on information security. Their key finding is, that with information sharing a firm can attain its optimal level of information security at lower cost and social welfare is improved. This shows the basic utility of the proposed platform. However, the authors point out, that incentive mechanisms are absolutely necessary to prevent misbehavior like free-riding.

A review of microeconomic literature points out the principal utility of such forms of information exchange for participants as well as for the overall welfare. However, there are different points that have to be considered to make a solution work. We will analyze various requirements in the next section.

### 4 Special requirements

When designing such a system, naturally various objections arise from potential participants. The main concern is of course about the confidentiality of the submitted data. But there are also other things that have to be kept in mind for a working solution. In the following sections we identify the main requirements for such a platform and sketch how they could be satisfied.

#### 4.1 Security

The data processed in the platform is very sensitive. Therefore it is obvious that security has to play an important part in the design of the platform.

According to the principles of multilateral security the interest of all parties involved should be considered. The input data, i.e. information on security events and incidents, discloses a lot of information about the participants. Therefore confidentiality is of great importance. The possession of that data gives insights in security measures, potential vulnerabilities and incident handling strategies. A disclosure of that information could give attackers valuable infor-
mation. However, the predominant barrier to sharing information on security incidents with others is the fear of negative publicity (see [4]). Garg et al. [3] have shown that the disclosure of information on security incidents frequently leads to negative stock market reactions.

The data that is provided by the platform is used by the participants as a basis for their security management decisions. Hence the integrity of that data is critical since wrong data might lead to wrong decisions. Furthermore the functioning of the system is based on a closed user group. The platform provider must be sure that only authorized users can submit and access data.

Attacks on the system could be performed by insiders as well as by outsiders. Potential attackers include the platform provider, participants, external data providers as well as other external parties. In the following we have identified the main challenges, divided up along the three protection goals.

Confidentiality

1. Confidentiality of communication between participants and platform. Nobody should be able to wiretap the data submitted.
2. Confidentiality of data stored in the database. Neither the participants nor external parties should have direct access to the information stored in the database.
3. Unlinkability. Even other participants should not be able to link a reported security incident to a company without the approval of that firm.
4. Privacy. It should be possible to participate in the system without revealing one’s identity to others.

Integrity

1. Integrity of communication between participants and platform. It should be impossible to modify data exchanged between platform and participants without being noticed.
2. Integrity of data within the platform.
3. Authentication and authorization. Since the basic concept requires a closed user group, it must be guaranteed that only authorized users can submit and receive data.

Availability

The proposed platform is not a real-time system. Nevertheless availability issues should be addressed. This is true for the submission system as well as for the data evaluation components. Especially when it comes to the integration of third party data, huge amounts of data have to be processed.

To reach these goals, we propose the following security mechanisms:

1. Establishment of a trusted third party. The platform manager has a central role. It needs to have access to all data since it has to conduct the data evaluation and interpretation. Because of this critical role we propose the platform manager to be an independent trusted third party that is accepted by all of the participants.
2. Use of cryptographic mechanisms to enhance integrity and confidentiality. The data between the platform and the participants should be transmitted encrypted. Authentication and authorization mechanisms can ensure that no outsider can access data or provide information under a wrong identity.
3. Use of privacy enhancing technologies. To hide from outsiders who is participating in the system, the connections between participants and platform should be performed using traffic anonymization techniques. To enable information sharing within the platform without revealing the identity of the organization that has reported as certain incident pseudonyms should be employed. As we will see in the next section privacy and fairness can be conflicting goals.

4.2 Fairness

An issue closely related to the previous section is the aspect of fairness. By this term we mean mechanisms that mainly address the problems of free-riding and the submission of wrong data.

Free-riding is a well-known problem in economics. Gordon, Loeb and Lucyshyn [6] give an overview over that phenomenon in the field of information security. In our case the term describes the phenomenon when one party is always taking, but never giving. That means it participates in the system without ever adding information itself. It just uses the data provided by others to improve its own security management. Microeconomic models show that free-riding can endanger the success of the whole platform. Without mechanisms to prevent free-riding, participants even tend to this behavior.

Another issue that has to be considered is the submission of wrong data. A party might submit wrong data with the goal to force other participants into suboptimal security investment decisions. Such a party might be a competitor who wants others to make economically wrong investment decisions or an attacker trying to influence security measures of the organizations for example by downplaying the potential damage caused by some kind of attack.

The third challenge in this area is about the misuse of data. The data collected and published in the system is intended to be used for the improvement of an organization’s risk management. In exchange for submitting their data the
participants get information that outsiders don’t get. So it must be ensured that no party discloses information to outsiders.

Since all of the above problems are closely related the mechanisms against them are similar. We see four categories of measures to foster fairness in the platform:

1. Incentive systems. An incentive system could reward the submission of attempted or successful security breaches to reduce the incentives for free-riding. Another approach might be to charge a certain amount of money if the number of submitted security events remains under a certain threshold. Gordon and Loeb [5] show, that participants in similar systems tend to submit too low values for damages. They discuss a submission charge for security events with a pricing shrinking with the rise of the reported event.

2. Reputation systems. An approach that is used in many electronic platforms are reputation systems. Participants get rated by other participants based on the perceived quality of their submissions. In our systems this rating could be based on the submitted incident reports. Rating just samples of the submitted data could reduce the workload significantly.

3. Legal framework. Some legal framework could enforce fairness. This might be some kind of self-commitment signed by the participants or even a contract between the participants or between the participants and the platform provider treating contract penalties. Of course misbehavior cannot be identified without additional mechanisms. So this can only be an amendment.

4. Technical mechanisms. Technical mechanisms along with manual and statistical plausibility checks could support the above mechanisms. The data submitted by the participants should be cross-checked against the data collected from external sources. Using statistical methods outliers could be identified. Those could then be manually checked for plausibility. This could either be done by a member of the third party or by posting the relevant data set to a message board and asking randomly selected participants for their plausibility rating.

Another important factor is that participants are not allowed to join and leave the platform randomly. The system is based on a closed user group. We expect the utility for the participants to be the higher, the longer they participate in the system. New members can join the platform only if they commit themselves to the established rules.

4.3 A common language and common metrics

The biggest merit of the platform is the fact, that participants do not have to rely solely on their own experiences and the quantitative data recorded in their own organization. They now have the possibility to use a much broader data basis which shall help them to optimize their risk assessment. Besides that they can benefit from other organizations’ experiences.

However, such a sharing of information is only valuable for the parties involved, if the information is comparable at all. This means that an organization must have the possibility to transfer the information about an incident and especially the impact to its own business context. The goal of the platform provider to aggregate the submitted data is only achievable if the data from different organizations is compatible.

To reach these goals we see two requirements:

1. A common language for the description of security events. Today most organizations record quite different information on their security incidents in quite different ways. For a sharing system we need a common taxonomy that allows an unambiguous categorisation and description. The same security incident has to be described the same way by two different organizations. Besides that, dimensions to describe the organization specific context of an incident are required.

2. Common metrics for the characterisation of damages. The main goal of the platform is the generation of quantitative data on damages. Since all organizations are different from each other it would be inexact to simply aggregate data from different sources. Common denominators are required that are easy to use and make data comparable. Besides recording the financial impact of an event there should also be other measures that provide a better comparability like hours to recovery etc.

4.4 Enabling direct information exchange

Interviews with potential participants have revealed another requirement. Besides precise statistical data, users also would like to have a possibility to exchange experiences on certain types of security events and to share best practices. This would also require to establish contacts between the participants and to show information on unaggregated level. Of course this requirement is somewhat conflicting with privacy requirements. Therefore the participants should have a possibility to determine themselves how much privacy they need and how much information they are willing to disclose to others.
4.5 Guidance

Practitioners and Researchers state that quantitative data is necessary for an economic security management. However, since such data has not been available there is to our knowledge no precise guidance on which data to use and how to integrated in the organization’s information security management system. Such a guideline should aid in assessing risks and selecting the appropriate security measures. It should also demonstrate how standards and legal requirements can be fulfilled by using quantitative data from the platform.

4.6 Usability

The more security events are reported by the participants, the better gets the quality of the output data. Hence there must be incentives for the participants to report completely. Thoroughly describing all security events can be a lot of work. Thus besides the mechanisms described in the section on fairness it must also be made as easy as possible to submit the events.

We see the following possibilities to facilitate the submission of events:

1. Easy to use web based user interface.

2. Provision of interfaces for different applications that handle security event data. For example the integration of existing incident information generated by an IDS or by a security management system could help security managers to save time.

5 Prototype implementation

We have implemented a prototype application of our basic architecture. Its main purpose is to test different mechanisms to fulfill the presented requirements and to illustrate the basic concepts to potential users. With this application we want to identify where further research has to be done, evaluate user behavior and especially the data submission component. In the following section we will give some insights into the technical realization and report on the first test results.

5.1 Preliminary theoretical work

A thorough analysis of the requirements showed, that a semantic model for the input of security events is a prerequisite for any implementation of the platform. We therefore started by identifying criteria for a good and applicable semantic model. Afterwards various existing taxonomies were evaluated against those criteria. We decided that for our purpose the taxonomy by Howard and Longstaff described in [8] was a good foundation due to its similar area of application and its numerous predefined categories. We extended it to cover all types of security incidents (like social engineering) and introduced a differentiation between internal and external attackers. The system was refined by categorizing the threats listed in the german baseline protection manual. Existing taxonomies are all limited to the description of attacks and vulnerabilities. Our focus is on estimating damages. So we had to develop a semantic model for the description of damages. The model utilizes different types of damages. To characterize the extent of damage various categories have been introduced. The main goal was to find measuring units that are comparable between organizations. To further improve comparability we also defined measures that characterize the organization itself.

5.2 Implementation details

The implementation is realised as a web-based client/server-application. The main goal of the prototype is to use it for tests and demonstration purposes. Potential participants should be able to use it without installing additional software so we decided to use a client-interface that can be used with a standard web-browser. The server-side is realised using Java Servlets and JavaServer Pages in combination with a Tomcat Server. As a database we use PostgreSQL. The application is realised as a multi-tier architecture.

To fulfill the confidentiality requirements all communication between user and platform is SSL-encrypted. Each participant gets a pseudonym that is used in combination with a password for authentication at the platform. The information stored in the platform cannot be accessed from the outside. For anonymization of the communication we recommend using some kind of anonymity service like AN.ON.

In the final system users should have alternatives to using the browser-based client for the submission of security events. We therefore decided to use XML for the submission of events to guarantee interoperability and extensibility. The semantic model is mapped in an XML-scheme. To facilitate changes in the semantic model all web-based forms are generated dynamically from the XML-scheme. The XML-solution also allows easy integration with security management applications that are already in use within the organizations.

Currently the submitted data is stored in a relational database. The evaluation is done manually by using statistical software. For the future we are planning to do automated data analysis using statistical software and data mining tools.
5.3 Conclusion

The architecture is running in a test environment. First tests showed that most users need guidance with the recording the first incidents since the systems requires many information.

Currently we are working on methods to use existing information that is for example stored in security incident reporting systems. The next step will be to implement the evaluation functionality as well as mechanisms to guarantee fairness within the platform. We are currently evaluating different kinds of incentive systems.

6 Conclusion and future work

In this paper we have shown why quantitative data is useful for economic security management and pointed out that it is not available at present. We have outlined why the collection of historical data might be a promising approach. Therefore we have proposed a complete architecture for a secure platform for the collection of quantitative historical data on security incidents. We have identified requirements which such a system has to fulfill to be practically applicable and showed ideas that satisfy those requirements. With a prototype implementation we demonstrated that such an architecture is realizable and also identified critical points.

Our next steps include the further development of the prototype and a test phase with various organizations. There are two important challenges. First, we need to implement mechanisms that guarantee fairness while still considering privacy requirements. Second, data evaluation has to be automated. In addition to that we are currently developing a process model for security management that uses experiences from the field of business engineering.

The platform developed in this paper can contribute to a more efficient security management by providing the concentrated experiences of various organizations in form of quantitative data. Participating organizations can use this as input data for their risk assessment and for investment decisions. The information generated might also be interesting for the development of cyber-insurances since it might become possible to better understand the statistical characteristics of security incidents.

We are fully aware, that historical data is not a solution to all problems. There is virtually no way to guarantee that really all relevant events are reported and of course their is a proportion of attacks and even successful attacks that is not discovered an thus cannot be reported. But as the greek philosopher Perikles is frequently quoted: “predicting the future accurately is not so important, being ready for it is”. Even if the data that participating organizations get from the platform is an imperfect estimate of the future, dealing intensively with their security management and their security incidents will help them to improve the level of security within their organization.

Acknowledgment

The authors would like to thank Johannes Bossle, Juergen Stern, Andreas Wagner and Alexander Zimmermann for their passion and dedication during their student projects helping our prototype to become reality.

References