DISSERTATION

# QUANTITATIVE ECONOMICS OF SECURITY: SOFTWARE VULNERABILITIES AND DATA BREACHES

Submitted by

Abdullah Mahdi Algarni

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2016

Doctoral Committee:

Advisor: Yashwant K. Malaiya

Indrakshi Ray
Indrajit Ray
Robert Kling

# ABSTRACT

## QUANTITATIVE ECONOMICS OF SECURITY: SOFTWARE VULNERABILITIES AND DATA BREACHES

Security vulnerabilities can represent enormous risks to society and business organizations. A large percentage of vulnerabilities in software are discovered by individuals external to the developing organization. These vulnerabilities are often exchanged for monetary rewards or a negotiated selling price, giving rise to vulnerability markets. Some of these markets are regulated, while some are unregulated. Many buyers in the unregulated markets include individuals, groups, or government organizations who intend to use the vulnerabilities for potential attacks. Vulnerabilities traded through such markets can cause great economic, organizational, and national security risks. Vulnerability markets can reduce risks if the vulnerabilities are acquitted and remedied by the software developers. Studying vulnerability markets and their related issues will provide an insight into their underlying mechanisms, which can be used to assess the risks and develop approaches for reducing and mitigating the potential risks to enhance the security against the data breaches.

Some of the aspects of vulnerability—discovery, dissemination, and disclosure—have received some recent attention. However, the role of interaction among the vulnerability discoverers and vulnerability acquirers has not yet been adequately addressed. This dissertation suggests that a major fraction of discoverers, a majority in some cases, are unaffiliated with the software developers and thus are free to disseminate the vulnerabilities they discover in any way they like. As a result, multiple vulnerability markets have emerged. In recent vulnerability

discovery literature, the vulnerability discoverers have remained anonymous. Although there has been an attempt to model the level of their efforts, information regarding their identities, modes of operation, and what they are doing with the discovered vulnerabilities has not been explored.

Reports of buying and selling the vulnerabilities are now appearing in the press; however, the nature of the actual vulnerability markets needs to be analyzed. We have attempted to collect detailed information. We have identified the most prolific vulnerability discoverers throughout the past decade and examined their motivation and methods. A large percentage of these discoverers are located outside of the US. We have contacted several of the most prolific discoverers in order to collect firsthand information regarding their techniques, motivations, and involvement in the vulnerability markets. We examine why many of the discoverers appear to retire after a highly successful vulnerability-finding career. We found that the discoverers had enough experience and good reputation to work officially with a good salary in some well-known software development companies.

Many security breaches have been reported in the past few years, impacting both large and small organizations. Such breaches may occur through the exploitation of system vulnerabilities. There has been considerable disagreement about the overall cost and probability of such breaches. No significant formal studies have yet addressed this issue of risk assessment, though some proprietary approaches for evaluating partial data breach costs and probabilities have been implemented. These approaches have not been formally evaluated or compared and have not been systematically optimized. This study proposes a consolidated approach for identifying key factors contributing to the breach cost by minimizing redundancy among the factors. Existing approaches have been evaluated using the data from some of the well-documented breaches. It is noted that the existing models yield widely different estimates. The reasons for this variation are

examined and the need for better models is identified. A complete computational model for estimating the costs and probabilities of data breaches for a given organization has been developed. We consider both the fixed and variable costs and the economy of scale. Assessing the impact of data breaches will allow organizations to assess the risks due to potential breaches and to determine the optimal level of resources and effort needed for achieving target levels of security.

# ACKNOWLEDGEMENTS

First of all, I would like to thank God (Allah) for granting me the health, patience, and ability to successfully obtain my PhD degree. Therefore, praise be to Allah the Almighty.

I am grateful to the academics who assisted me during my work in this dissertation. I give my special thanks and great appreciation to my advisor, Dr. Yashwant Malaiya, for his guidance, encouragement, and supervision throughout my PhD journey. I was lucky to work with him, and I received many benefits from his vast experience. I also give special thanks to my doctoral committee members, Dr. Indrakshi Ray, Dr. Indrajit Ray, and Dr. Robert Kling, for reviewing my dissertation and providing valuable advices and suggestions to improve my work.

I would like to dedicate this dissertation to the people nearest to my heart: my father, mother, brothers, and sisters. Their encouragement, prayers, support, and love reduce the distance between us and make me feel that I'm near them. This has helped me a lot during my studies.

I would like to thank my lovely family: my wife, Haila; my daughters, Sulaf and Rtal; and my son, Eyas. I dedicate not only my dissertation but all of my life's achievements to them. They have shared my success, been patient throughout the long years, and have given me love and encouragement. Without their support, I would not have reached my goal and earned my PhD.

Finally, I acknowledge my beloved country, Saudi Arabia, and my university, King AbdulAziz University, for their support and for giving me an opportunity to complete my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# Chapter 1

# Introduction

The potential exploitation of software security vulnerabilities has emerged as a major security threat to organizations, some economic sectors, and national defense. A software vulnerability can be defined as a software defect or a weakness in a security system that could be exploited by a malicious user, causing loss or harm [1]. A vulnerability exploit is a code that exploits a vulnerability and serves as proof that the vulnerability is indeed exploitable. A vulnerability that has not been disclosed and its associated exploit are often termed a "zero-day". A patch, when applied, can remedy a vulnerability. The number of unremedied vulnerabilities in a system represents the degree of security risk. During the software lifecycle development process, it is important to evaluate and manage this risk in order to assess how it will impact users, organizations, and society.

The vulnerabilities are traded or exchanged in vulnerability markets. These markets may be classified as regulated, in which the transactions are properly recorded and disclosed; black, in which the transactions are not disclosed and there is no attempt to safeguard society; or gray, in which the transactions can be either termed legitimate or improper depending on one's point of view. In the regulated markets, the buyers are the original software developers, third-party security service organizations follow proper practices when disclosing the vulnerabilities, and the transactions are well-documented [2, 3]. Recent reports suggest that government agencies in several countries have become major players in the gray markets [4], and thus, there has been a remarkable change in the vulnerability markets.

However, the vulnerability markets have a huge impact on economics and security of organizations and governments. This impact comes from security data breaches. The risk of compromising the confidentiality of records has increased with the increasing trend of businesses collecting, storing, and transmitting financial and personal information in electronic form. A data breach incident refers to malicious data disclosure to unauthorized parties [5]. Such breaches occur for different reasons. The hackers' intent is often to steal data for financial gain, but they may also intend to harm individuals or organizations by disclosing their confidential information or intellectual property. In the past few years there have been several well-publicized data breaches. In spite of significant effort by government agencies, those involved in such cybercrimes are often caught only after a long period of activity, and some are never caught. The potential victims of such crime thus have the burden of assessing their risk and investing in countermeasures.

The study of vulnerability markets, their main players, and the risk presented by those markets are interesting and essential, including the issue of lost or stolen data via data breaches by the exploitation of vulnerabilities. The subject is especially compelling, considering few academic researchers have tried to explore this area and no academic group, for example, has constructed a systematic model to estimate data breach risks, including the cost and probability of data breaches. Both vulnerability markets and data breach risks are main issues in software/website industries in terms of the data security concept and its impact on economies and societies.

### 1.1 The Dissertation Objective

The current situation of the vulnerability markets encourages discoverers to sell their vulnerabilities in illegitimate markets instead of legitimate markets because the large monetary

rewards are attractive to them. Furthermore, some people or groups who have special agendas buy vulnerabilities to harm others. This situation generates many security risks, such as data breaches, and these can cost a great deal of money. Therefore, studying the types of discoverers and consumers and their motivations, vulnerability markets and the money flow, and making a consolidate model to estimate the data breach risk will support the goal of reducing the security risks by improving security awareness for all people who has related to software vulnerability. As the result, the legitimate markets should be more attractive for vulnerability discoverers, who might otherwise resort to selling their findings on the black market, and software developers should be ready to any data breach especially they can estimate the data breach cost and probability and this is important since the security mistake will cost their organizations a lot of effort and loss business.

However, the topic of vulnerability markets is huge, and it is related to other overlapping areas, such as software quality, risk management, software engineering, security measurement, and economics. Therefore, the motivation of the dissertation will be diverse. The dissertation focus on the three major directions, which are as the follows:

- **Analysis and organize the vulnerability markets:**

As we mentioned before, the vulnerability markets could lead to some security risks in the form of lost information and economic risks in the form of data breach costs or damage costs if a vulnerability is sold to disreputable people, organizations, or governments who attempt to harm others to further their own agendas. Therefore, analyzing the types that vulnerabilities discoverers can sell is very important. In addition, knowing how the money flows through those markets and how those markets interact with the vulnerability lifecycle is crucial to understanding these markets. Studying the markets from the inside will assist in organizing them

and making some markets more attractive than others. This will also reduce security risks, which is one of our main goals of this dissertation.

- **Analysis the main players in the vulnerability markets and their motivations:**

Studying the types of vulnerability discoverers (sellers) and consumers (buyers) will help us to identify the kinds of people who are interested in vulnerabilities. Moreover, every player in the markets has different motivations that led him or her to choose a specific vulnerability market in which to buy or sell vulnerabilities. This means that responsible buyers, such as software vendors, can create rewards to attract certain discoverers, depending on the discoverers' motivations. This will create new interest in certain less-attractive markets. Therefore, the transactions in illegitimate markets will be reduced. Then, the security and economic risks will be reduced as well.

- **Investigating the consequences of the misuse of the vulnerability markets:**

Usually, if the owners of a software program or computer system know the risks of vulnerability exploitation in terms of their reputation, information, security, and economic worth, they will focus on all possible causes of such risks. Therefore, studying academically the factors that affect the direct or indirect of data breach costs and making a consolidate and systematic model to estimate and predict the cost and probability of data breach will encourage software owners to find solutions for the key issues in the vulnerability markets.

### 1.2. The dissertation's Contributions

In this dissertation, we obtain some data for analysis and then identify the research objectives by using various approaches, such as direct surveys, using the current dataset, and collecting additional data manually. These approaches assist us in studying the software vulnerability markets and the main players in those markets, and estimating the data breach risks.

We investigate the types of vulnerability markets and their classifications, types of vulnerability discoverers and consumers and their motivations for selling or buying the vulnerability, and monetary flow between discoverers and consumers during vulnerability markets. We found that there are four different classifications for the current software vulnerability markets: producers (vulnerability discoverers types), vulnerability markets types (Regulated markets, online forum, gray markets, black markets), consumers (buyers of vulnerability), and the buying reasons such as attacking other organization or patching the vulnerability. In addition, we found that major vulnerability discoverers are individuals from external the software development organizations and their main motivation for discovery is financial rewards and this result comes from direct questions that when we ask them about their motivation in addition to other questions.

For the security data breach risks, we study the existing approaches for calculating the cost of data security, we identified and examined the existing data breach cost calculators that are available online. We collect the factors that impact the cost and probability for each calculator. Then, we choose the significant factors and remove the insignificant factors depending on scientific approach.

After we collect sufficient actual data regarding significant factors that affect the cost and probability of data breaches, and we know how every calculator computes the cost or probability of a data breach. Subsequently, we make a consolidated, systematic model to estimate the cost and probability of a data breach.

### 1.3. Dissertation Structure

This dissertation is organized as follows: Chapter 2 provides some background concepts and related work to aid the reader in understanding the main topic and all its major branches.

Chapter 3 presents the results for the analysis of vulnerability markets, as well as important information related directly to these markets, such as their money flow. Chapter 4 shows the methods and motivations of the main players of the vulnerability markets, especially the vulnerabilities' discoverers, who are considered the main reason for launching those markets. The consolidated factors approach for estimating data breach costs is demonstrated in Chapter 5. In addition, a quantitative risk assessment model of a data security breach is presented in Chapter 6, and it includes the data breach cost and probability. Finally, Chapter 7 presents research challenges, possibilities for future work, and a conclusion.

## 1.4 Publication History

Most of the results that have already been obtained through experiments in this dissertation have been published in peer-reviewed conferences [6, 7] and a journal [8]. Some of these publications have a good number of citations by other research groups.

# Chapter 2

# Background and Literature Review

In this section, some of the background concepts and terminology are discussed in order to understand the problem at hand and the approaches that appear later in this dissertation. In addition, related work is mentioned to know what other researchers did and to illustrate what this research contributes to this topic.

## 2.1 Fundamental vulnerability markets and security terms

Many concepts in vulnerability markets are essential to explain in this dissertation. Following are basic security and economics terms that are related to vulnerability markets.

- **0-day vulnerability:** it likes a normal vulnerability (hole in a software), but it is unknown to vendor. Thus, the attacker can exploit the vulnerability before patching it [9]. Zero-vulnerability or normal vulnerability are both significant in the vulnerability markets, but any zero-vulnerability should be more expensive in the markets because it is unknown and this leads to unexpected attack (0-day attack or exploit).
- **Vulnerability Markets:** "where security-related information can be traded" [10].
- **Vulnerability Discoverer:** This is an individual who or an organization such as a vendor, independent researcher, cyber-criminal, or government agency that discovers a new vulnerability [11]. The discoverer could sell the discovered vulnerability to other.

- **Hacker:** This is an entity that releases exploits for the vulnerabilities in the products of software [12]. There are three different types of hacker (exploiter): White, gray, and black hat.
- **Attack:** An attacker needs to exploit at least one vulnerability in the system to reach to his aim [13].
- **0-Day Attack (Exploit):** This is a cyber-attack exploiting a vulnerability that has not been disclosed publically [14].
- **Threat:** The likelihood or frequency of a harmful event occurring [15].
- **Bug Broker:** "an authorized negotiator that acts on behalf of society in order to increase transparency of the bug bounty process" [16]. This is in vulnerability rewards programs, but the broker usually found in the gray/black markets to sell and buy vulnerability to people who are interested in vulnerabilities trade.
- **Severity:** This is calculated by a risk score that assigned by CVSS and indicates the size of the risk associated with a vulnerability [15].
- **Reward for Vulnerability Discoverer:** "It provide an incentive for security researchers not to sell their research results to malicious actors in the underground economy or the gray world of vulnerability markets" [17]. Usually the reward provides in legitimate markets and depends on the severity of the vulnerability, but in other markets, it calls vulnerability price.
- **Full Disclosure:** This is a security philosophy that holds that the details of security vulnerabilities should be available to everyone in a timely fashion [11].
- **Patch Release Time:** The number of days elapsed between the vendor notification date and the patch release date [18]

- **0-Day Patch:** This is a patch where the vulnerability is disclosed the same day the patch is released by the vendor [19].

## 2.2 Vulnerability lifecycle phases

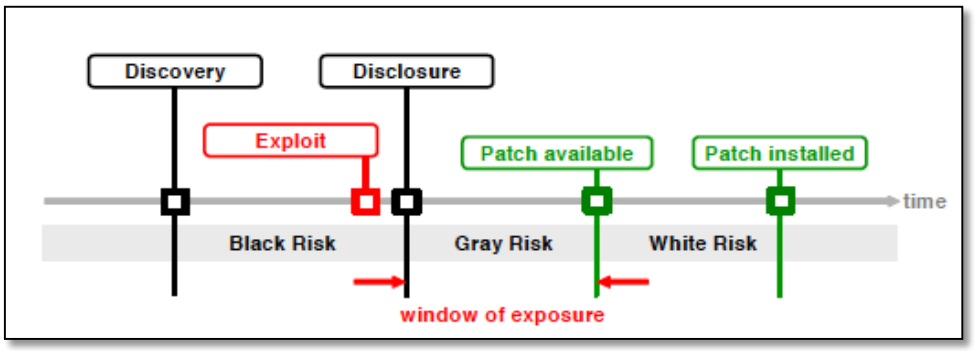


Figure 2.1: Lifecycle of Vulnerability [19]

Arbaugh et al. [20], Marconato et al. [21], Frei et al. [22], and Okamura et al. [23], demonstrated the lifecycle model that describes the states of a vulnerability over its lifetime (see Figure 2.1). The lifecycle is separated by distinctive points in time that divide the lifecycle of vulnerability into phases, where each of them reflects a state and an associated risk [20]. Thus, the term *vulnerability lifecycle* implies the fixed and linear progression from one phase to another phase, to understand the behavior of vulnerabilities. All possible states of vulnerability were addressed as follows:

- **Birth:** This denotes the creation of a software flaw or defect.
- **Discovery:** Discover of the vulnerability in the software product. The vulnerability discoverers could be black hat or white hat.
- **Disclosure:** The discoverers can expose the vulnerability details to the developer, the public or to a wide audience.
- **Correction (Patching):** Correct the vulnerability by releasing a software modification through the software vendors or developers.

- **Publicity:** A vulnerability becomes public in any of several ways.
- **Scripting (Exploitation):** A new vulnerability is successfully exploited by someone with moderate skills.
- **Death:** When the vulnerability is patched or the attackers are no longer interested in it.

Studying vulnerability lifecycles can support the development, deployment, and maintenance of software systems in designing future security policies and conducting audits of past incidents. This also leads to assessment of the security risks in software products of different vendors [12]. The sequence of exploit, disclosure, and patch is not always fixed [22] because sometimes the exploit and the patch can be before, at, or after the discovery state.

### 2.3 Vulnerability databases

There are several vulnerability databases organized by government-affiliated or private organizations. These include the National Vulnerability Database (NVD) [24], Open Source Vulnerability Database (OSVDB) [25], the vulnerability data collected by Frei et al. [12] (FVDB), Exploit Database, and IBM X-Force Vulnerability Database. NVD provides information about exploitation requirements and consequences, while OSVDB provide solution types available for a particular vulnerability [26]. CVE stands for Common Vulnerabilities and Exposures [27], which is a list of standardized names for vulnerabilities and information security exposures. A CVE-number indicates a standardized identifier for known vulnerabilities.

### 2.4 Common vulnerability scoring system (CVSS)

CVSS is a specification for measuring the relative severity of software vulnerability [15]. It uses three groups of metrics to calculate the vulnerability scores (Figure 2.2).

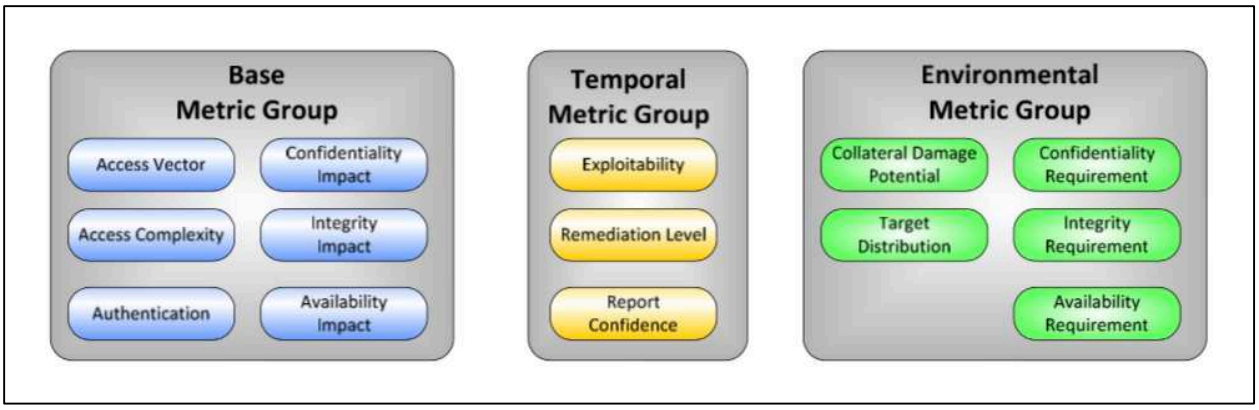


Figure 2.2: The CVSS Metric Groups [15]

- *Base metrics* are vulnerability attributes that are constant over time and across all implementations and environments.
- *Temporal metrics* are vulnerability attributes that change over time but apply to all instances of a vulnerability in all environments such as the public availability of exploit code or remediation techniques. The temporal vulnerability score is calculated with an equation that uses the score of both base and temporal metrics as parameters.
- *Environmental metrics* are also vulnerability attributes that are organization- and implementation-specific, such as how prevalent a target is within an organization. The environment score is calculated with an equation that uses the score of both temporal and environmental metrics as parameters.

The vulnerability *severity* is measured by CVSS metrics that target vulnerability exploitation and prioritize the mitigation of new vulnerabilities. The severity rankings include Low (0.0 to 3.9), Medium (4.0 to 6.9) and High (7.0 to 10.0).

### 2.5 Vulnerability security impact

The chances of loss of information and records have increased with increases in personal information for most people being collected, stored and transmitted by businesses in electronic

form. Data breach incidents caused by malicious hacking can exploit some software vulnerabilities and harm people or organizations by breaching some data, and this can lead to harassment, embarrassment, impersonation or theft.

### 2.6 Economic costs (impact) of data breaches

The cybercrimes that occur following data breaches and the impact of these crimes on the economy in terms of damage and cost can be enormous. In many cases, the damages to individual organizations have been estimated in the hundreds of millions of dollars. Some costs are not easily measurable, such as impact on segments of the economy or national security.

### 2.7 Related work

There are several academic references that discuss vulnerability markets. Most are preliminary works that attempt to study those markets, describe the highlights, and identify the main players that affect them. These references did not provide enough information or sufficient explanations as to how vulnerability markets are comprehensively classified or how the main players interact with them. In addition, another question remains: how does money flow between players and through the markets and what is the data breach impact for the organization that face a breach. The main difference between our research and others' research is that ours is specific and comprehensive: it concerns all vulnerability market classifications, the main types of players and their motivations, and the flow of money, which is the main fuel that drives these markets. Furthermore, we study the actual vulnerability markets rather than hypothetical markets, which have often been used in recent literature. In addition, how the cost and probability of security data breach are calculating for the breached organizations. However, the related work here will be divide into four parts: all vulnerability markets in general, a specific vulnerability market

(black market), another specific vulnerability market (vulnerability rewards programs), and data breach risk (include data breach cost).

### 2.7.1 Vulnerability markets

Fully-fledged vulnerability markets appeared in 2002, whereas the initial full disclosure movement began in the 1990s [3]. Schechter proposes a vulnerability market in which the software vendors can provide some rewards for the vulnerability finders (testers) in their products [28]. These markets can be used to improve the product's security. A first step in creating a mechanism for vulnerability markets to increase the system's security was introduced by Camp and Wolfram [29]. They confirm that governments can make the system more secure when they issue some incentive for vulnerabilities.

Examination of the vulnerability disclosure mechanisms is essential to know which mechanism has a better impact on the software security in the end. For instance, Kannan and Telang [30] made an initial and simple structure model of vulnerability disclosure mechanisms that include the vulnerability markets mechanism and the competition between benign users and hackers during the vulnerability discovery phase to analyze the impact of monetary incentives, which are provided by some organizations (see Figure 2.3). That model has four main participants including the infomediary (such as iDefense or CERT), the discoverers, who are either benign or malicious, and the software users. The vulnerability markets are unregulated if the infomediary leaks vulnerability information. Otherwise, the markets are regulated. The authors of this paper present an early mathematical study of vulnerability markets and compute the user loss probability of each case of the security attacks on the software users and the welfare probability of obtaining of incentives for vulnerability reporters (identifiers) and the subscription fees between software users and the infomediary. For example, when the infomediary receives the information from identifiers, it will use that information to protect only the users who

subscribe to the infomediary's service. The main difference between the CERT mechanism and a market-based mechanism is the second mechanism provides monetary rewards to vulnerability reporters when they submit a vulnerability. They come to the conclusion that a federally controlled mechanism such as CERT would benefit society more than an unregulated market-based mechanism as implemented by iDefense because the disclosure rules of the private vulnerability markets are socially suboptimal. In contrast, a regulated market-based mechanism is better than the CERT mechanism under some conditions.

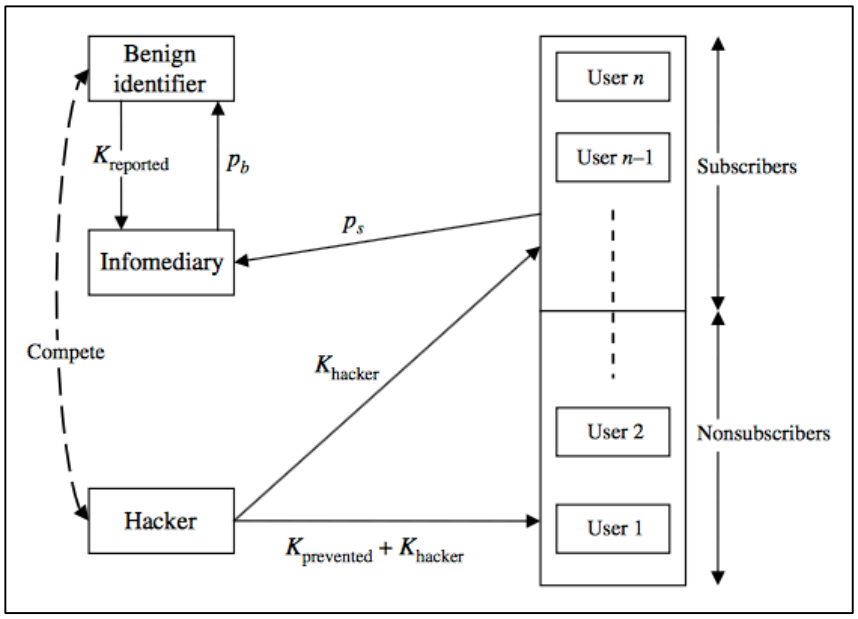


Figure 2.3: General Structure Model of Vulnerability Disclosure [30]

In addition, Ransbotham et al. [31] examine market-based vulnerability disclosure mechanisms' effectiveness through an empirical examination of two years of security alert data. Figure 2.4 shows the primary pathways to public disclosure of software vulnerability including immediate disclosure by security professionals such as security mailing lists (BugTraq), nonpublic disclosure like CERT, which immediately notifies the vendor and then discloses the vulnerability to the public (usually 45 days after notifying vendors) when the patch is available, market disclosure such as IDefense and ZDI which can sell the vulnerability through these

markets by security professionals. However, attackers always use one pathway to sell their discovered vulnerability into black markets, which leads to direct exploitation. As a result of studying those mechanisms, the authors determined that market-based disclosure controls vulnerability exploitation diffusion; this reduces the exploitation risk and the number of exploitation attempts.

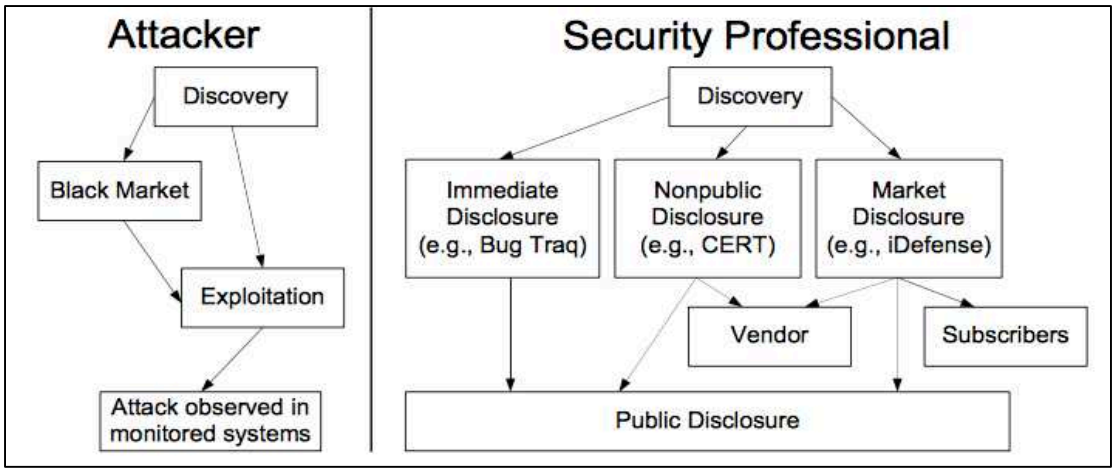


Figure 2.4: Vulnerability Disclosure Pathways [31]

One of the oldest economic concepts is selling the vulnerability in auction. Therefore, McKinney [3] briefly discusses vulnerability commoditization in software industries and other elements that are related to the vulnerability bazaar (i.e. an auction) such as the main players who deal with markets, gray markets, and vulnerability auctions. Furthermore, Ozment [32] describes auctions as having the best design for understanding vulnerability markets to improve defense plan against security attacks. He uses economic concepts of auctions, such as auctions beginning at low prices and then rising in price until bids are accepted by sellers. The vulnerability auctions' concept like WabiSabiLabi has disappeared since 2007 [33], because the buyers of the vulnerabilities are frequently suspects who can harm others and this is against the law in most countries.

Another economic concept used to trade vulnerabilities is the vulnerability market. Some researchers have focused on that type of market. Miller discusses the zero-day legitimate markets in terms of the problems of 2005—such as the lack of transparency in pricing of vulnerability and the difficulty of finding sellers and buyers in legitimate markets, since these markets are not attractive for some reasons—which are almost identical to our current issues [2]. He also suggests possible solutions, like trying to find direct auctions or trusted third parties to attempt the sale of zero-day exploits as legitimate incomes for both sellers and buyers of vulnerability, as described in two case studies. There was no actual implementation.

Finally, Böhme discusses the advantages and disadvantages of five different types of vulnerability markets to find the best market type by comparing criteria such as information, incentives, risk balancing, and efficiency [10]. These criteria were important during his studies on economic values concerning zero-day exploits in computer security. Vulnerability finders can sell their vulnerabilities into thriving markets such as some big security companies like iDefense and ZDI.

### 2.7.2. Black markets

Other literature has focused on specific markets or used markets to obtain specific results. For instance, Radianti and Gonzalez [34] explain the features of vulnerability and the issues of black markets by proposing a preliminary model that demonstrates the factors that affect specific vulnerability markets such as black markets and how software vulnerability has been mitigated. Figure 2.5 demonstrates the vulnerability discovery flow into legal markets or black markets. After that, the authors have built many sub-models for the black markets and the main players such as sellers and buyers.

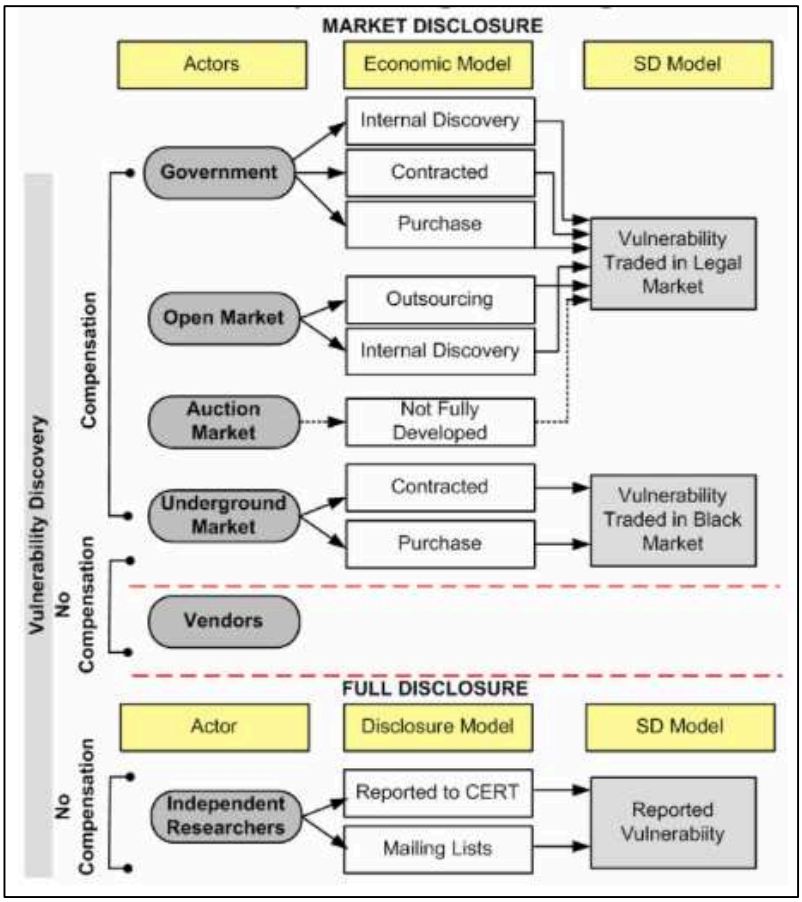


Figure 2.5: Vulnerability Discovery Flows and Simplification in System Dynamic Model [34]

Moreover, Radianti et al. [35] have continuously focused on the vulnerability of the black market and its development as a marketplace for selling or buying exploits by malicious hackers or others. These authors proposed a dynamic model to capture the structure of vulnerability, as discovered through black and white markets, and to examine the successful and unsuccessful factors of those markets in order to prevent further developments in black markets. Radianti et al. proposed five main sectors including the following (see Figure 2.6): chain of vulnerability sector, exploits sector, black market sector, security researchers sector, and payment sector. The chain of vulnerability and exploits sectors affect each other since the exploits lifetime to some extent depends on the secrecy of the vulnerabilities. Otherwise, vulnerabilities are repeatedly detected from the circulated exploit. Exploit identifications sometimes spark exploit trading in the black

markets. The security researchers sector includes people who are able to discover vulnerabilities or establish exploits, viruses, and other malware, and consists of different groups: black hat hackers in the black market, white hat researcher volunteers, and the black hat and white hat researchers in legitimate markets. The legitimate and black markets are mostly very attractive for the security researchers due to monetary incentives. This last is captured in the payment sector.

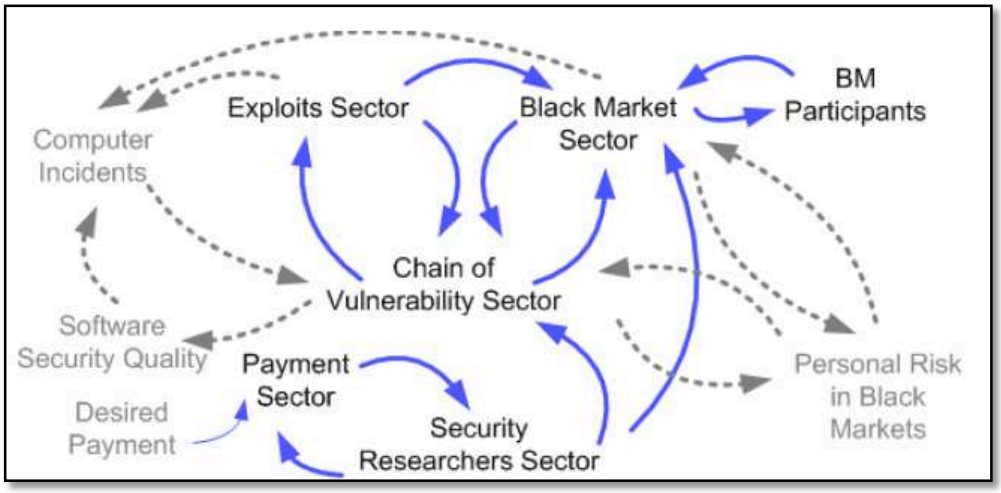


Figure 2.6: Main Sectors in a Vulnerability Black Market Model [35]

However, Allodi et al. [36] tried to answer several questions: Are black markets relevant for final user security, and does it make sense to use vulnerability information from the black markets to design patching policies? To accomplish this, they utilized two steps: first, they checked for the relevance of exploit kits' vulnerabilities in the general attack scenario, and second, they developed a model to estimate the reduction in risk using a typical CVSS-based strategy and a black market-based strategy. The authors mainly used data on vulnerabilities traded in the black markets that they collected themselves, and real attack data as reported by Symantec's WINE data-sharing program. After this was complete, they classified the information into several categories to allow assessment of confounding influences on the probability of vulnerabilities exploitation. Their preliminary observations showed that the

vulnerability risk score (CVSS) versus attacks performs well, but it leaves 40% of the attacks uncovered. Conversely, the vulnerabilities in exploit kits drive between 10% and 40% of attacks received by the final users. Therefore, cyber-crime black markets are an important source of risk for final users. Active and efficient monitoring of the markets may lead to more efficient patching strategies. In addition, the efficacy of patching strategies seems to vary with the category of the vulnerable software, so there may be a need for ad hoc policies for different software products.

### 2.7.3 Vulnerability rewards programs

Remediation strategies are very useful to reduce the impact of loss of information. One of these strategies is vulnerability rewards programs. Therefore, providing rewards to motivate people to find software defects or weaknesses before they are exploited by black hat exploiters is critical to improving computer security. There are only a few current vulnerability reward programs, and most of them were created a few years ago. The idea of reward programs is still quite new, and needs more development and improvement. Currently, many companies have rewards programs that provide a reward for software discoverers who discover a defect or flaw in software. A few vendors, like Adobe and Oracle, do not have their own VRPs because they argue that these rewards programs do not represent the best return on investment on a per-bug basis. Finifter et al. [17] examine the features and characteristics of two well-known vulnerability rewards programs (VRPs) from 2009 to 2012. Those two programs appear to be two to one hundred times more cost-efficient than hiring expert security researchers to find vulnerabilities. The authors recommended that software vendors focus on three important points to make their VRPs successful. First, the rewards for participants should follow a tiered structure, such as is used in Chrome, where participants who submit critical vulnerabilities receive a high amount and those who submit low-impact vulnerabilities receive a low amount. In

addition, the high variance in patch release time for critical vulnerabilities in Firefox could negatively affect the security community, since participation would reduce responsibility through the vulnerability rewards programs. The third factor, providing high rewards for interesting vulnerabilities, such as Chrome's rewards, will encourage participants to apply to the VRPs.

### 2.7.4 The risk of security data breaches

The potential risks that appear after the exploitation of a vulnerability should be discussed with and published to raise awareness for the public and specialists. For example, cybercrime is estimated to have lost the global economy around US $445 billion in 2014 [37]. The costs of those breaches could be the main way to call organizations' attention to security risks to vulnerability markets. Few academic groups focus on analysis and study data breach risks in terms of cost or probability. Layton and Waters [38] estimated the tangible costs of data breaches using two case studies (Telstra and LinkedIn), focusing on a salary guide and rough estimation of the work hours of the people who were involved to manage the data breach. They estimate labor costs, regarding them as the only tangible cost. In terms of intangible costs, Layton and Waters only consider the loss of reputation. It is interesting to note that they argue that the stock price was not negatively impacted after announcement of data breach in the two cases considered. The authors do not calculate the probability of data breach which is a part of the risk.

# Chapter 3

# Discoverers and Buyers in the Software Vulnerability Markets

## 3.1 Introduction

The vulnerability discoverers represent a critical source of security risk, should they choose to sell the vulnerability to malicious organizations or individuals. A vulnerability sold to the developing organization [39] results in a patch that minimizes the risk. However, a vulnerability could also be sold to an organization interested in using it for exploitation. Reports suggest that some exploitable vulnerabilities can command market prices exceeding $100,000[40, 41].

This study examines real vulnerability markets as they exist. In a market, a commodity (here an undisclosed vulnerability) is made by the producers, and is bought by the consumers or resellers of the commodity. The price is determined by supply and demand. A market may be regulated to some extent or it may be largely unregulated. The presence of a market itself enhances the level of production and consumption. Kannan and Telang [30] present an early mathematical study of the vulnerabilities market, where the discoverers are either benign or malicious, and come to the conclusion that a federally controlled mechanism would be better for society. As we present here, the real markets that have emerged are more complex and are international in nature.

As we discuss below, a large percentage of vulnerabilities are found by experts external to the actual software development organizations. They are free to disclose the vulnerabilities that they discover in any way they like. The hackers who are vulnerability exploiters are often classified as white hat, black hat, and gray hat [42, 43]. These classifications do not apply to the vulnerability discoverers who are security researchers engaged in finding vulnerabilities since they may choose different markets for different vulnerabilities. Also note that, in general, the

exploiters and the discoverers are distinct groups. Discovering vulnerabilities is not an illegal or anti-social activity, it is a respectable profession; whereas exploiting vulnerabilities is generally the opposite (unless it is part of a service).

The vulnerability markets, as shown in (Figure 3.1), may be classified as regulated, where the transactions are properly recorded and disclosed; black, where the transactions are not disclosed and there is no attempt to safeguard the society; or gray, where the transactions can be termed legitimate or improper, depending on the point of view. In the regulated markets the buyers are original software developers, third-party security service organizations follow proper practices for disclosing the vulnerabilities, and the transactions are well-documented [2, 3].

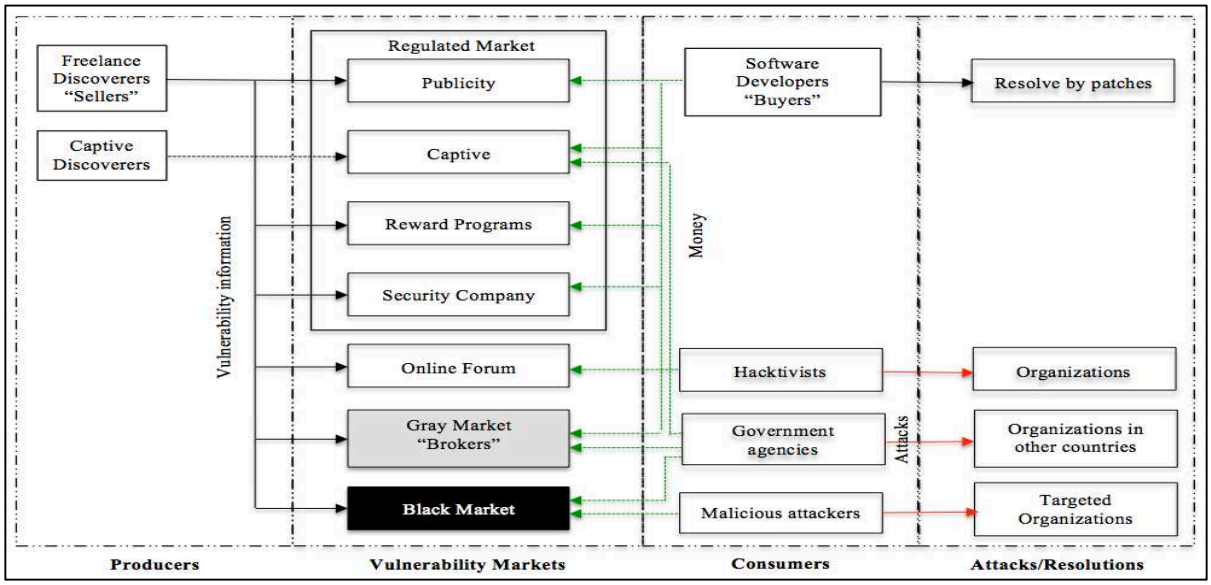


Figure 3.1: The current software vulnerability markets

The current software vulnerability reward programs are a major part of the legitimate markets that attempt to attract the vulnerability discoverers who might otherwise resort to selling their findings on the black market. These programs are relatively new and sometimes limited. They attempt to bring a discovery to the legitimate market, which significantly reduces the risk

to society. Some government agencies in some countries are becoming main buyers of software vulnerability.

### 3.2 Vulnerability Markets and the Main Players

Vulnerability discoverers can be internal or external to a software development organization. They seek appropriate rewards for their capabilities. The external vulnerability finders (freelancers) are often free to offer their discovery in exchange for a suitable reward. They may attempt to maximize their reward by selling vulnerabilities in the appropriate vulnerabilities markets [30, 44].

In an ideal situation, the discoverers seek no reward and submit the vulnerabilities found to a responsible disclosure mechanism. Receiving credit for a vulnerability discovered is sufficient compensation for some. However, it is not enough for many discoverers. They know that vulnerabilities can have significant economic value [10] because they can lead to zero-day exploits that might harm organizations, the economy, and ultimately, society [45]. Some exploits have sold for as much as $250,000 [46]. In addition to money, some discoverers find the fame generated by the disclosure attractive, as it can translate into further economic opportunities.

Some organizations, such as Google, have acknowledged the importance of freelance discoverers, and offer a significant monetary award in addition to the possible inclusion in their 'discoverers hall of fame'. A good example of a vulnerability discoverer who has taken advantage of such a reward program is Sergey Glazunov, a Russian student and security researcher who earned $60,000 by discovering a new exploit in Google's Chrome browser [47]. Generally, finding vulnerability exploits is legal, and some legitimate businesses sell them. The price for an exploit sold to business and government agencies in the United States ranges from $20,000 to more than $250,000 [48]. Each market has some attributes that are more attractive for

some producers (discoverers) and consumers (buyers) based on their long and short term objectives. A market is defined by its governing rules and conventions. The transactions between discoverers and buyers (software venders, those with malicious intentions, or resellers) involve an exchange of vulnerability information for a suitable price, generally money. The buyers of vulnerabilities derive the value by making their software product safer, or by the rewards a zero-day attack may bring. Below we discuss each of the markets.

### 3.2.1 Regulated Vulnerability Market

This is a regulated market that is controlled by conventions and laws that attempt to prevent any improper actions towards the society as a whole. It includes the four markets discussed below. In all of these markets the vulnerability information is transferred to the software venders, who then patch the vulnerability in their products before it is disclosed.

#### 3.2.1.1 Publicity

In this case, the discoverer submits the finding to an authority, where it undergoes a well-defined responsible disclosure procedure. The discoverer gets the recognition and the software developer gets the chance to develop a patch before the vulnerability is disclosed. The publicity generated may enhance a discoverer's reputation as a capable researcher. CERT and other similar organizations provide such a market. This market would not be attractive for discoverers who have already established themselves or who need money more than publicity. For some, the recognition received may eventually translate into economic opportunities.

#### 3.2.1.2 Captive Market

In this market the discoverers are captive to an organization and are thus not permitted to reveal the vulnerabilities externally. This includes vulnerability finders working within a software development organization or those working for them under contracts. Researchers within security service organizations are also, in effect, captive. In some countries the

government may be the only permitted buyer, although in that case the government is free to use the vulnerability based on its national priorities.

#### 3.2.1.3 Vulnerability Rewards by Vendors

The reward programs offered by software venders are a good option for vulnerability discoverers who can sell their finding to the venders directly in an easy, legitimate way. The reward offered for a vulnerability can be significant, although modest in most cases. In addition, the discoverers receive appropriate credit [32].

Rewarding security researchers and others who make software products more secure is important. Providing rewards to motivate people to find software defects or weaknesses before they are exploited by black hat exploiters is critical to improve computer security.

There are only a few current vulnerability reward programs, and most of them were created a few years ago. The idea of reward programs is still quite new, and needs more development and improvement.

The current reward programs include these listed below. The key information about them is provided in (Table 3.1).

Table 3.1: Some current vulnerability rewards programs

| Program | # Vulns. type | Max reward | Min reward | # of beneficiaries | Trend |
|---|---|---|---|---|---|
| *Vulnerability Reward Program for Google web properties* | 5 | $20,000 | $100 | 2010: 51<br>2011: 122<br>2012: 189<br>2013: 226 | Increase |
| *Chrome Vulnerability Reward Program* | Any security bug | >= 10,000 | $500 | 543 | N/A |
| *The Mozilla Security Bug Bounty Program* | Certain bugs depending on some criteria | $3000 (US) cash reward and a Mozilla T-shirt | $500 | N/A | N/A |
| *Facebook* | Certain qualifying security bugs | No maximum | $500 | Prior to 2011: 43<br>2011: 46<br>2012: 111<br>2013: 235 | Increase |
| *WordPress Security Bug Bounty Program* | 11 | $1000 | $25 | N/A | N/A |
| *CCBill Vulnerability Reward Program* | 7 | $ 500 | $300 | 42 | Hold |
| *Secunia Vulnerability Coordination Reward Program (SVCRP)* | Most bugs depending on some criteria | Most Valued Contributor& Most Interesting Coordination Report | N/A | N/A | N/A |
| *ZDI Rewards Program (TippingPoint)* | Particular bugs depending on some criteria | $25,000 | $1000 | N/A | N/A |
| *iDefense (Verisign)* | N/A | N/A | N/A | Significant number | N/A |

• Vulnerability Reward Program for Google web properties [49]: This program was created in November 2010. People who discover one of five types of vulnerabilities, such as remote code

execution, SQL injection, and other common web flaws, are rewarded $100 to $20,000. The number of discoverers who have received approval from the reward panel has ranged between 53 winners in the fourth quarter of 2010 to 39 winners in the fourth quarter of 2012.

• Chrome Vulnerability Reward Program (Chromium Security Reward) [50]: All vulnerabilities are considered in this program, provided the vulnerability is identified as being of sufficiently high severity. The rewards range from $500 to $10,000 and up.

• The Mozilla Security Bug Bounty Program [51]: The rewards range from $500 to $3,000 depending on the severity rating of the vulnerability, and the reward includes a Mozilla t-shirt.

• Facebook [52]: This program is similar to most other reward programs. It offers a bounty for certain qualifying security bugs. The reward has a minimum of $500 with no specified maximum, and is based on severity and creativity.

• WordPress Security Bug Bounty Program [53]: This program has two different bounties: one for WordPress and another for WordPress Plugins. The minimum reward is $25, and the maximum reward is $1,000.

• CCBill Vulnerability Reward Program [54]: CCBill is an Internet billing service. The rewards range from $300 to $500, depending on the types of vulnerabilities found, such as SQL Injection, DoS, and persistent XSS.This program has been temporarily placed on hold due to corrections needed in the reported bugs.

Eventually, Microsoft has announced last June 2013 a month-long vulnerability rewards programs for the Internet Explorer (IE11) developer preview [55]. Microsoft was not interested in paying rewards programs since it does use outside consultant organizations to test their software on a contract basis, however [56].

#### 3.2.1.4 Rewards by Security Service Companies

Some companies that provide security services also acquire vulnerabilities. The vulnerabilities acquired are used to provide a higher degree of safety to their security customers, and may be provided to the software developer using a suitable compensation mechanism. These organizations do not sell the vulnerabilities to others. For example, Microsoft patched at least 17 vulnerabilities reported by the two programs in 2006 [40]. There are some third-party security companies that buy the vulnerabilities and sell them to software makers or vendors such as ZDI and iDefense [57]. Such reward programs include the following:

- Secunia Vulnerability Coordination Reward Program (SVCRP) [58]: There are two special awards: most valued contributor and most interesting coordination report.
- ZDI Rewards Program [59]: The Zero Day Initiative (ZDI) provides reward points each time a vulnerability submission is purchased. These points determine the ZDI status, which are bronze, silver, gold, platinum, and diamond. The rewards range from $1,000 to $25,000.
- iDefense Vulnerability Contributor Program: This is one of the oldest reward programs, and a few top discoverers mention working with iDefense. Detailed reward information is not available.

The security service companies may have their own internal vulnerability researchers. Their discoveries primarily serve to promote the organization.

### 3.2.2 Vulnerability Brokers

As opposed to the security services companies, vulnerability brokers buy as well as sell vulnerabilities. They come closest to an open market, since buyers and sellers can negotiate their prices [60, 61]. It is considered a legitimate, but only partially regulated, market that has some

general rules. A vulnerability broker is an organization or person who provides a link between a vulnerability discoverer and the highest bidder. It has been reported that the commission might reach up to 15% of the selling price [57]. Therefore, the broker may sell that information to the software venders or to some government organization, depending on who can pay more. Several international government organizations are said to have become significant buyers [46] in recent years, but their policies are not generally disclosed.

Vupen, located in France, is an example. They can sell vulnerabilities to a government, provided that the government belongs to NATO, ANZUS, or ASEAN alliances [57]. If the software venders buy the vulnerability information, they will use it to patch their products, but if it is purchased by a government agency [62], they might use it for military purposes, to inflict damage to an opponent as a cyber-weapon, or to collect sensitive information (espionage) from opposing governments or organizations. Security experts have claimed that the Stuxnet malware was specifically created by government agencies in the United States and Israel to attack Iran's nuclear facilities in 2010 [63]. We can regard the vulnerability brokers to be a gray market, which can be legitimate from the point of view of national priorities. However, considering the amount of funding that governments can bring to the table, such markets will reduce the number of public disclosures [64]. “The Grugq” a Bangkok-based security researcher, is regarded as an influential global vulnerability broker. He arranges deals between vulnerability discoverers and western government agencies for a 15% commission [41]. A vulnerability auction site WabiSabiLabi was active a few years ago [33].

### 3.2.3 Online Forums

Online forums exist where information about vulnerabilities and exploits can be exchanged. In some cases, the exchange may not involve money—rather, the members (called hacktivists)

have a special or private agenda to attack specific organizations. LulzSec was a famous hacktivist group that attacked several user accounts and websites in different countries in 2011. Anonymous is a loosely connected network of hacktivists located in different places that choose the same targets to attack. It is likely that such groups do not have access to zero-day vulnerabilities since they would be too valuable to reveal without any financial gain. It is likely that they rely on installations that have not yet applied the patches needed.

### 3.2.4 Vulnerability Black Market

The vulnerability black market is not a regulated market and it is not controlled by any laws. This market allows any groups or organizations such as cyber criminals, terrorists, or government agencies to buy vulnerabilities. The price paid to the vulnerability discoverers is said to be five to ten times the amount of the other vulnerability markets, depending on the attributes of the vulnerabilities [31, 41]. The estimated price range given by some in-the-field experts for a zero-day exploit is given in (Table 3.2) [46, 48]. Many governmental and commercial organizations, such as the International Monetary Fund, Intel, the Indian Defense Ministry, and the Pacific Northwest National Laboratory, have suffered from the malicious attacks [61]. Government agencies in several countries have programs to develop new cyber weapons, and they may be significant players in the black market for zero-day vulnerabilities [65].

Table 3.2: Price list for zero-day vulnerability exploits

| Products | Minimum price for zero-day exploits "2011" | Minimum price for zero-day exploits "2013" |
|---|---|---|
| *Adobe Reader* | $5,000 - $30,000 | N/A |
| *Mac OSX* | $20,000 - $50,000 | N/A |
| *Android* | $30,000 - $60,000 | $100,000 |
| *Flash or Java Browser Plug-ins* | $40,000 - $100,000 | N/A |
| *Microsoft Word* | $50,000 - $100,000 | N/A |
| *Windows* | $60,000 - $120,000 | $40,000 - $250,000 |
| *Firefox or Safari* | $60,000 - $150,000 | N/A |
| *Chrome or Internet Explorer* | $80,000 - $200,000 | $200,000 - $500,000 |
| *IOS* | $100,000 - $250,000 | $50,000 - $200,000 |

### 3.3 The Consumers (Buyers) of Zero-Day Vulnerabilities

Ultimately, the buyers of vulnerability information are either software developers who intend to eliminate the vulnerability by developing a patch, or an organization that intends to use it for purposes that would seem malicious to its opponents. When a government agency is a buyer, it can bring a substantial amount to money to the market that other buyers may be unable to match. This may raise the prices of the vulnerabilities and in turn encourage more experts to enter the profession of vulnerability discovery. The motivations of consumers differ depending on the type of consumer, but we can divide those motivations into two categories: patching the vulnerability or attacking others.

### 3.4 Potential Impact of Money Flow

In the past few years, several government agencies associated with different countries have started investing in offensive and defensive capabilities for engaging in cyber warfare and espionage [46]. In comparison with conventional military hardware, the cyber capabilities are potentially much more cost-effective. Reports suggest that some vulnerabilities, along with their exploits, can bring a significant amount of money. This could cause a significant shift in the markets. As we discuss, it might lead software developers to be more aggressive in their reward programs.

### 3.5 Vulnerability Markets and the Risks to Society

A few researchers have recently evaluated security risk based on the vulnerability life-cycle [66]. However, they have not considered the impact of the vulnerability markets. Figure 3.2 shows the vulnerability flow involving the markets.

As Figure 3.2 shows, even disclosed vulnerabilities can be a source of risk. Some vulnerabilities can be disclosed without a patch either because of logistics reasons or because

they are judged to be inconsequential (state $e_{wp}$ Disclosure without Patch). When the patch is available, some users may not apply it, immediately leaving it in an exposed state (state $e_{pn}$ Patch not Applied). Some of the conventional vulnerability scanning products offer protection against such states.

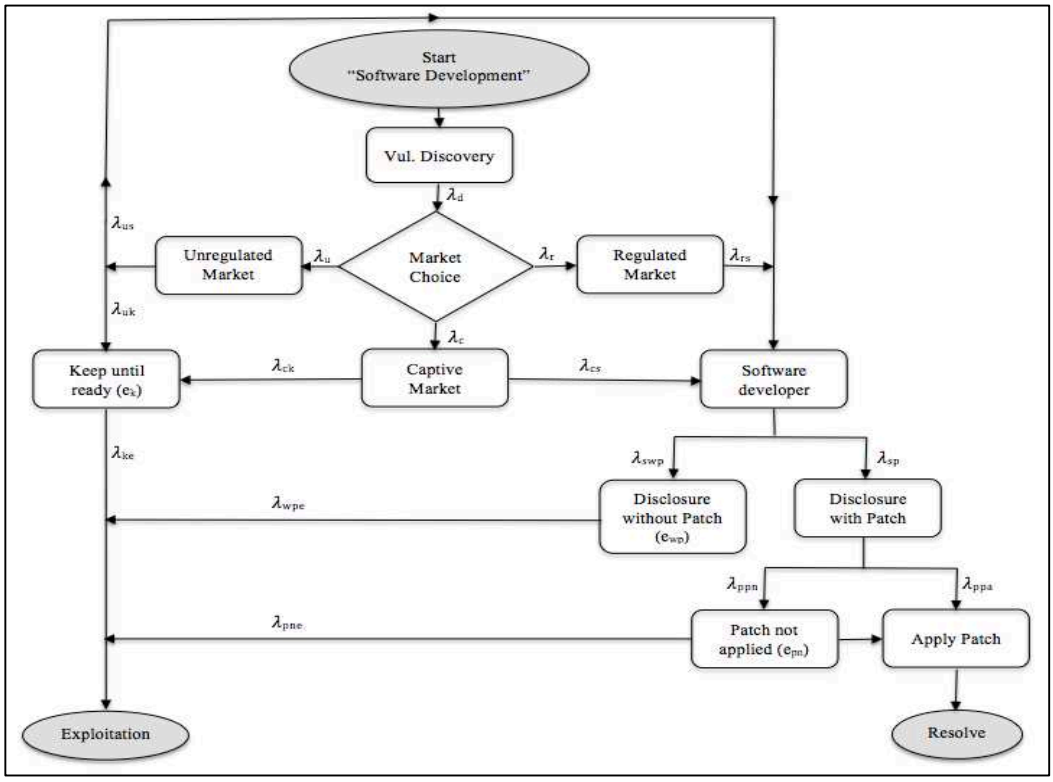


Figure 3.2: Vulnerability flow through markets to zero-day exploitation or patching

There is no protection against zero-day vulnerabilities, however, which have not been publically disclosed (state $e_k$ Keep until Ready). We can note that the captive market has two options: sell the vulnerability to software developers or keep it until it can be used for an attack, for example. Even highly secured systems can be potentially exploited using the zero-day. They can be expensive to acquire, but can be used for cyber warfare, cyber terrorism, espionage, or an attack on vital institutions of an opposing nation.

Figure 3.2 shows the three states $e_k$, $e_{wp}$, $e_{pn}$, in which the vulnerability is exposed. The risk due to an exploitation of a vulnerability during a time window (t1, t2) is given by [66] as in (3.1):

$$Risk(t_1,t_2) = \int_{t_1}^{t_2} \sum_i P(e_i)\lambda_i \; dt.Exploitation\ impact \tag{3.1}$$

Where i is one of the one of the exposed states ($e_k$, $e_{wp}$, $e_{pn}$) and $\lambda_i$ is the transition rate from that state to the exploitation state. $P(e_i)$ is the probability of being in the state $e_i$.

Note that a zero-day attack is only possible for a vulnerability passing through the unregulated markets, with the exception of a captive market (such as a defense lab) where the objective is to discover vulnerabilities for exploitation.

A significant fraction of the cyber-attacks on systems belonging to individuals or organizations occur though $e_{wp}$, $e_{pn.}$ An attack through $e_{wp}$ is through a known risk, and one through $e_{pn}$ could be considered a consequence of negligence. However, an attack through $e_k$ would be completely unexpected, and, depending on the target and the type of breach, can have devastating consequences on an organization or a society.

As we observe, a large fraction of successful vulnerability discoverers is from regions that are not as industrially developed. Some of these regions are also known for their sophisticated vulnerability exploiters [67]. This suggests that economics might play a significant role in potential approaches for keeping the society safe.

An attractive reward program based on vulnerability criticality can provide a significant alternative to the gray and the black markets. A few software developers and security organizations now run a small number of such programs. These programs ensure time for patch development before a disclosure. Some of the top discoverers that we contacted suggest that sometimes the reward programs do not pay enough, and a better reward may be obtained on the black market (although none of them admitted to selling any vulnerabilities in such a market.)

We note that after a few years of very successful vulnerability discovery, many of the top discoverers apparently disappear from the scene as credited discoverers. Some of them suggest

that they find it more profitable to contract out their security auditing services to software developers. This can also significantly reduce the risk to the society.

Companies and organizations need to design attractive vulnerability reward programs for their products. This will allow the legitimate markets to compete with the black market.

Some reward programs, such as the one for Google Chrome, appear to have been successful. While the amount of money committed to the reward programs is only a tiny part of the company's revenues, Google is giving out some of the best monetary rewards.

A significant part of the global vulnerabilities market is quite opaque. Even the emerging legitimate markets have not been studied in detail, although some mathematical studies based on the classical market theories have appeared. There is a need to examine actual data and practices in order to understand the vulnerability discovery and disclosure.

The zero-day vulnerabilities with exploits are a serious issue. The number of high-profile attacks that use the zero-day has increased sharply during the first three months of 2013 [68], demonstrating the amount of risk associated with the unregulated markets [36]. Mechanisms need to be developed to make it more profitable for researchers to sell their discovered vulnerabilities in the legitimate markets, therefore reducing trading in the unregulated markets.

# Chapter 4

# Motivation and Methods of the Most Successful Vulnerability Discoverers

## 4.1 Introduction

Vulnerability discovery models that attempt to model the vulnerability discovery process have been recently proposed [69, 70]. However, there has not been a study of actual vulnerability discoverers and what motivates them. The individuals who discover the vulnerabilities (termed *discoverers* here) and those who exploit them (*exploiters*) are two separate groups. Discovering a vulnerability takes a much higher degree of technical skill and insight. The exploiters do not require a comparable skill—in fact, in the presence of an exploit (code that exploits one or more vulnerabilities), a patient hacker may achieve a security breach largely mechanically. The vulnerability discoverers represent a critical source of risk, should they choose to sell the vulnerability to malicious organizations or individuals. For example, Google has twice paid a $60,000 reward for details on a single vulnerability [39], suggesting that the potential damage caused by these vulnerabilities could have been enormous.

Many vulnerability discoverers seek to preserve the right to their claim of having discovered a vulnerability, since it serves to acknowledge the discoverer's expertise. For example, the well-known University of Cambridge researcher Ross Anderson mentions a vulnerability he and his student discovered in 2003 [71]. A mid-year peak in vulnerability discovery, specifically in Microsoft products, can be explained by the coinciding date of a major conference, wherein security experts often present their vulnerability findings [72].

As presented in this study, a large percentage of vulnerabilities are found by experts external to the actual software development organizations. They are free to disclose the vulnerabilities they discover in any way they like. The hackers who are vulnerability exploiters are often

classified as *white hat*, *black hat*, and *gray hat* [42, 43]. These classifications do not apply to the security researchers engaged in finding vulnerabilities. However, the vulnerability markets may be classified as *legitimate*, where the transactions are properly recorded and disclosed; *black*, where the transactions are not disclosed; or *gray*, where the transactions are at the borderline. The current software vulnerability reward programs are a major part of the legitimate markets that attempt to attract the vulnerability discoverers who might otherwise resort to selling their findings on the black market. Those programs are relatively new and sometimes limited. They attempt to bring a discovery to the legitimate market, which significantly reduces the risk to the society. It is possible that some groups, such as government defense agencies, may be willing to pay a much higher price in the black or gray market [4].

Any unpatched vulnerabilities in a software program can allow hackers to attack the system, harming an organization or compromising sensitive information. Therefore, remedying any newly discovered vulnerabilities before they are exploited is critical. While many discoverers are likely to be responsible professionals, they need to be provided the opportunity to use their skills in a positive, productive way in order to avoid passing the information to those who might exploit the vulnerabilities. If there is a lack of incentives from organizations in the field, they might be tempted to sell the information in the vulnerability black market, resulting in possible exploitation of systems.

The motivation for vulnerability discoverers has been considered briefly by researchers in the past [73], but has never been studied using actual data. The discovery and disclosure of vulnerabilities are processes that are significantly impacted by the economics involved [74]. A few researchers have considered theoretical modeling of the vulnerabilities market. This paper asks these questions: who are the actual vulnerability finders, and what motivates them?

Vulnerability discovery is done by either researchers affiliated with a major organization (and who generally follow proper disclosure policies) or by freelance researchers, who may sell their findings, either in the legitimate or in the gray or black markets.

We study the current types of vulnerability discoverers. We find initially in Figure 4.1 that vulnerability discovery is done by either researchers affiliated with a major organization (and who generally follow proper disclosure policies) or by freelance researchers, who may sell their findings, either in the legitimate or in the gray or black markets. Some vulnerabilities are sold in the legitimate market via vulnerability reward programs, or by contacting venders directly. The developers get a chance to develop software patches for the vulnerabilities before the vulnerability is disclosed. On the other hand, when the vulnerabilities are sold on the black market, they are likely to be exploited before public disclosure. The key strategy would be to encourage the researchers to sell their discovered vulnerabilities in the legitimate market instead of the black market (dotted circle on the figure). This will reduce trading in the black market and more vulnerabilities will enter the legitimate market.

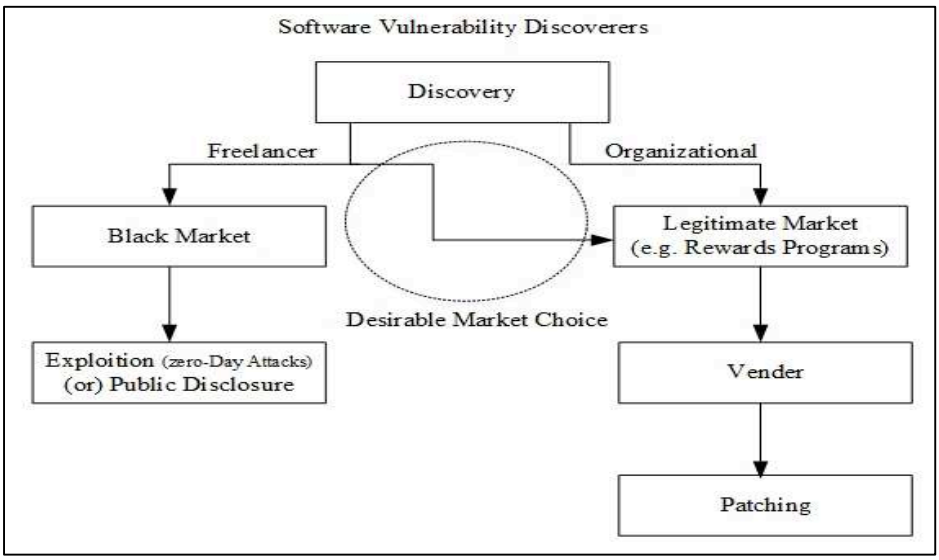


Figure 4.1: The events in the vulnerability lifecycle

However, we also investigate who the actual vulnerability finders are and what motivates them. To answer these questions, we use several steps. The following subsections describe those steps in detail.

### 4.2 Top Discoverers

To understand the vulnerability discovery process, we examine the records of the top vulnerability discoverers. Since each of them has successfully discovered a significant number of vulnerabilities, we can presume that they did not just get lucky—rather, they have a system that has been demonstrated to work. To find the top vulnerability discoverers, we obtained data from the OSVDB database (until May 2013). We identified the top fifteen vulnerability discoverers in the database who found the most vulnerabilities, as given in (Table 4.1). The actual names are not identifiable in some cases; they are generally known by the login identifier that they use in their blogs. Table 4.1 includes some of information about them obtained from blogs and discussion boards in addition to OSVDB. (Table 4.1) illustrates the global distribution of vulnerability finders. A significant number of them are in eastern and western Europe, with a few in the Far East, in addition to some in the United States. This shows why the legal framework within a single country cannot regulate the vulnerability markets. Thus, while ideal vulnerability markets can be proposed, they cannot be implemented. Ultimately, economics will govern the markets.

Table 4.1: The top vulnerabilities discoverers on OSVDB

| Discoverer | Country | Period | # Vuln | # Vuln types | Why they're interested | Stopped/Continued |
|---|---|---|---|---|---|---|
| *r0t* | Latvia | 2005-08-09 to2010-09-16 | 810 | 10 | N/A | N/A |
| *Janek Vind "waraxe"* | Estonia | 2003-08-08 to2013-03-21 | 319 | 8 | Vulnerability website | N/A |
| *Lostmon Lords* | Spain | 2004-06-20 to2009-08-15 | 279 | 8 | Security Researcher | Worked until July 2012 |
| *rgod* | Italy | 2005-06-06 to2012-08-29 | 277 | 12 | Hacker | Worked until Aug. 2012 |
| *Luigi Auriemma* | Italy | 2000-07-08 to2013-03-16 | 267 | 9 | Hobby | N/A |
| *Russ McRee* | USA | 2008-01-14 to 2012-03-02 | 237 | 4 | Specialist in security | N/A |
| *Aliaksandr Hartsuyeu* | Lithuania | 2005-12-28 to2011-02-03 | 229 | 6 | Security Company | Still working 2012 |
| *James Bercegay* | USA | 2003-06-03 to2008-09-04 | 200 | 12 | Web developer | Worked until 2011 |
| *Kacper* | Poland | 2006-05-12 to2007-08-10 | 199 | 3 | N/A | N/A |
| *luny* | N/A | 2006-05-18 to 2006-07-13 | 142 | 6 | N/A | N/A |
| *Diabolic Crab* | N/A | 2004-09-25 to 2005-07-12 | 140 | 6 | N/A | N/A |
| *JeiAr* | USA | 2003-05-29to 2004-05-04 | 120 | 7 | Web developer | Worked until 2011 |
| *Tan Chew Keong* | Singapore | 2004-07-29 to 2009-09-28 | 102 | 9 | Information Security Specialist | N/A |
| *Stefan Esser* | Germany | 2000-11-09 to 2012-06-03 | 86 | 10 | Security Consultant | Still do jailbreak until 2012 |
| *M.Hasran Addahroni* | Indonesia | 2006-02-09 to 2009-02-07 | 80 | 2 | Security Gossiper&Bugs Hunter | N/A |

However, Figure 4.2 gives a plot of the yearly discoveries of all of the top discoverers (line marked with a diamond). Year 1 corresponds to the first year of the discovery. Most of the top discoverers here are credited with discovering the vulnerabilities during the first three years (line marked with a rectangle). However, a few discoverers have continued to discover vulnerabilities for several years. This raises an intriguing question. Why do some very successful discoverers disappear from the scene after two to three years? A possible explanation is that, during those two to three years, they acquire the notoriety of being accomplished vulnerability discoverers. After that, they start offering their services to software developers or security service companies on a contract basis or as employees. Some of them may be able to start their own small organizations. In both cases, they are able to obtain steady and significant remuneration rather than a few rewards from the reward programs. This is supported by the information that some of them have provided us, as discussed below.

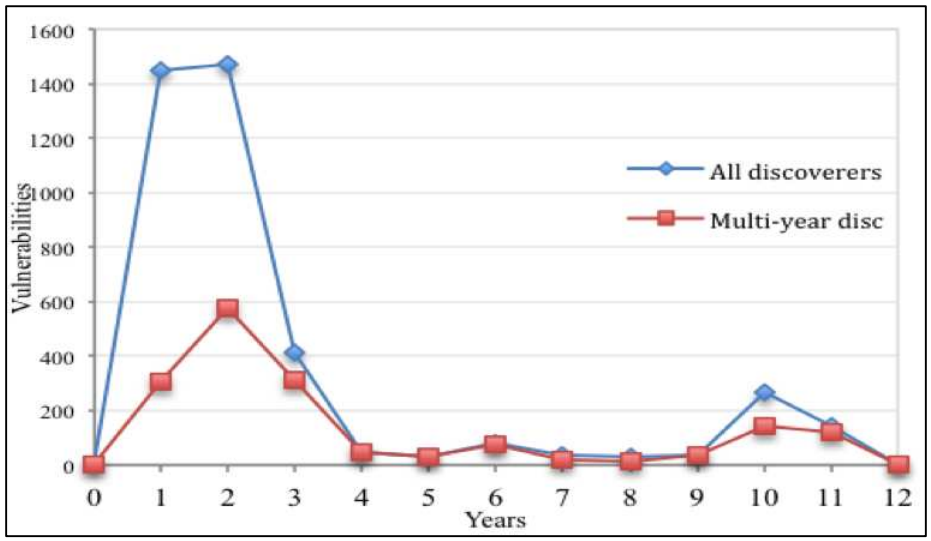


Figure 4.2: Vulnerabilities discovered yearly

### 4.3 Outsider and Insider Discoverers

One key question in understanding the vulnerability discovery process is whether a discoverer of vulnerabilities is a part of the software product team or an outsider. This will help us to understand what motivates discoverers to find and report software vulnerabilities. To address this question we examined two well-known open-source software products (as example): Safari and Google Chromium (Table 4.2, Figures 4.3, 4.4). The period we investigated was from July 1, 2012 to December 31, 2012, and we used the Open Source Vulnerability Database OSVDB as the data source.

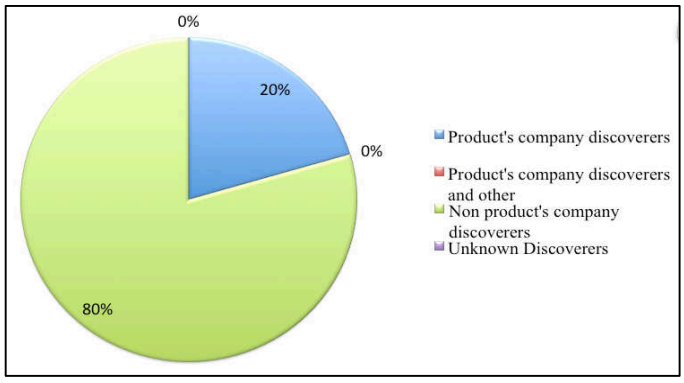


Figure 4.3: Vulnerability discoverers in Safari

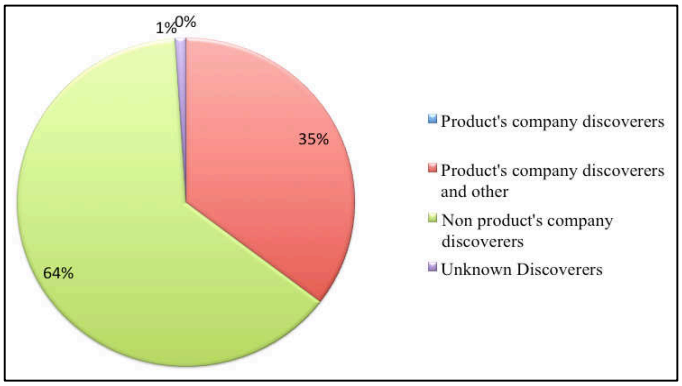


Figure 4.4: Vulnerability discoverers in Chromium

As shown in Table 4.2, for these two products, the majority of the vulnerabilities discovered were found by outsiders. Finifter et al. [17] have also found this to be true for particularly for Google Chrome although not for Firefox. This demonstrates the importance of outsider discoverers and the potential significance of providing discoverers with more enticing vulnerability reward programs, or other forms of a legitimate market. It is definitely worth knowing what would motivate the discoverers to participate in such reward programs.

Table 4.2: Vulnerability discoverers from July 1, 2012 to December 31, 2012: insiders or outsiders

| Discoverers | Safari's Vulnerabilities | Percentage | Chromium's Vulnerabilities | Percentage |
|---|---|---|---|---|
| *Product's company discoverers* | 17 | 20% | 0 | 0% |
| *Product's company discoverers and others* | 0 | 0% | 35 | 35% |
| *Outside discoverers* | 66 | 80% | 63 | 64% |
| *Unknown Discoverers* | 0 | 0% | 1 | 1% |

## 4.4 Direct Information

Some of the vulnerability markets are secretive, specifically the gray market, where the brokers serve as intermediaries, and the black market. They are, however, believed to be of great significance, and government agencies are emerging as vulnerability buyers. To understand the motivations and mechanics of different markets, we decided to directly contact the top discoverers of OSVDB to seek information. We were able to locate contact information for many of them. We then contacted them and asked some key questions, including the following:

1) What motivates you to discover software vulnerability?
2) How and when did you start?
3) What specific tools do you use for discovering vulnerability?
4) Did you stop working as a vulnerability discoverer? If so, when and why did that happen? If not, why not?
5) Do you think that vulnerability reward programs will help reduce black market transactions

and encourage the use of legitimate markets? Please explain.

6) Did you apply to one of the current vulnerability reward programs, and if so, why?

7) If you have discovered a vulnerability, when would you consider selling your vulnerability to a broker? Please explain.

8) In your view, are there any specific steps that software developers or government agencies can do to reduce the security risk to society? Please explain.

9) Do you have any other comments?

Considering that freelance vulnerability discoverers can sometimes be secretive, we were pleasantly surprised when several of them actually responded; although most of them did not reply to us. The following section includes some of the answers to the above questions. To ensure their privacy, we have replaced the discoverers' names with aliases. Table 4.3 summarizes the responses.

• Discoverer 1: He uses his own tools, "specifically [his] hands and mind, in preference to automated tools". He has not sold a vulnerability in the past ten years. He does not find the reward program to be attractive. He never sold his own discovered vulnerability to brokers or any buyers, but he has sent vulnerabilities directly to the software venders.

• Discoverer 2: The main reason he became a vulnerability discoverer was that he wanted to promote his own website and his source code review service. He only uses his own tools, which are offered on his organization's website. He states that that vulnerability reward programs are of limited use, as the black markets offer more money. Like Discoverer 1, he does not apply for any reward programs.

• Discoverer 3: He started in 2002 while following Bugtraq and other mailing lists. He uses both public and proprietary tools to discover vulnerabilities. Although he now runs his own

company, he still finds the time for discovery work. He states that reward programs pay very little for exclusive information and bug patches, which can be sold for much more on the black market. Nevertheless, he has submitted some vulnerabilities to the ZDI and iDefense reward programs in the past.

• Discoverer 4: He started in 2008 and focuses entirely on web application security flaws, largely specific to free and open source applications. To discover vulnerabilities, he uses a combination of tools such as Burp Suite, OWASP ZAP, and a number of Firefox plugins (Tamper Data), as well as simple manual testing. He thinks that, for the most part, vulnerability reward programs will help to reduce black markets and encourage legitimate markets. He acknowledges that money is always a motivator and if vulnerability discoverers are paid well via the legitimate market, hopefully they will be less likely to sell the bug on the black market. He claims that he does not sell vulnerabilities. He always coordinates his findings with Secunia but does not take any further action regarding the vulnerability.

• Discoverer 5: He believes that the most profitable option for a vulnerability discoverer is to offer software security auditing services. His first discoveries were done between 1992 and 1993. The tools that he uses for discovering vulnerabilities are Notepad++ for PHP and other scripting languages, which allow him to search specific text strings through multiple files and color-coding. He also uses Apache/PHP/MySQL on his home PC, and all of his web application research is done using @localhost. Discoverer 5 usually works manually, without automatic vulnerability scanners. He believes that vulnerability reward programs will surely lessen damage, and is aware of hundreds of zero-day findings sold to ZDI and other vulnerability reward programs. He has worked with ZDI and iDefense because they pay for findings, arrange all communications with developers, and give him credit in the public advisory.

• Discoverer 6: He began to delve further into discovering after he did an interesting exercise during a computer security lab when he was a university student. He uses popular open-source tools, such as the ones distributed with Kali Linux, as well as his own scripts. He is still working in vulnerability discovering as a web application security pentester. He thinks that vulnerability rewards programs should be lucrative since pentesting requires a great deal of time, which results in a personal output of money and effort. Therefore, if one cannot earn money through legitimate channels, he will sell his discovered vulnerabilities on the black market. He is an active participant in all major bug bounty programs for two reasons: money and renown in the community. However, he claims that he himself has never considered selling the bugs he discovers to a broker.

• Discoverer 7: He began doing so in his early teens by finding security issues in the online games that he played. He used various intercepting proxy/tools for replaying requests, such as LiveHTTPHeaders or Burp Proxy, for vulnerability discovering. He is still involved in vulnerability discovering as a full-time employee on the Product Security team at Facebook. He has submitted issues to many of the bug bounty programs that currently exist; he believes that these programs are a great way to apply his skills and have his efforts rewarded and recognized. As a responsible vulnerability discoverer, he always tries to disclose an issue to the vendor before selling the vulnerability to a broker.

• Discoverer 8: He used HttpWatch and Burp Suite to capture the http traffic and did the rest of the work manually. He has taken a break from discovering vulnerabilities since he did not get enough time. He thinks that legal rewards programs offer white-hats good money, so there are fewer chances that white-hats will become black-hats. He has been participating in rewards programs for the past year and a half. He only reports vulnerabilities to vendors, and claims that

he never thinks about selling vulnerabilities to brokers or anyone else. If a vendor does not respond properly, he discloses the vulnerability in a blog post, but he does not sell it. He believes that government agencies should support and encourage new rewards programs, such as HackerOne (a bug bounty program for the internet).

Table 4.3: Top vulnerability discoverers' answers to specific questions about their vulnerability discovering and reward programs

| Discoverer | Motivating Factors | Stop Discovering | Impact of Rewards Programs | Applying to Rewards Programs |
|---|---|---|---|---|
| *DISCOVERER 1* | Hobby and lifestyle choice | No | N/A | No |
| *DISCOVERER 2* | Make his website more popular | No | Limited impact | No |
| *DISCOVERER 3* | Curiosity | No. He has a company | Not much impact | ZDI and iDefense |
| *DISCOVERER 4* | Enjoyment | Yes. Not enough time | Mostly, yes | No |
| *DISCOVERER 5* | Fun, profit, auditing | No | Yes | ZDI and iDefense |
| *DISCOVERER 6* | Lean new discovering skills | No | Yes | All major programs |
| *DISCOVERER 7* | Enjoyment | No. Part of his job | Mostly, yes | Many programs |
| *DISCOVERER 8* | Passion, profit, learn new technologies | Yes. Not enough time | Yes | Many programs |

We note that many of the discoverers acknowledge the significance of the gray market and the black market in vulnerabilities. Many of them have found it profitable to engage in contract work after having established credentials as expert vulnerability finders.

Another notable observation from this study is the fact that the freelance discoverers appear to rely on their expertise more than on specific tools. Some of them have developed their own tools based on their experience. This suggests that the discovery of previously unknown vulnerabilities is a research activity requiring considerable technical skill, rather than something that can be completely automated using algorithmic methods. This should be contrasted with vulnerability scanning tools that look for known, i.e. already disclosed, vulnerabilities.

## 4.5 Discussion

Upon investigating the factors that influence vulnerability discovery and disclosure, we summarize our findings as follows.

We note that freelance discoverers play a significant role in vulnerability discovery. In some cases, they have even formed their own companies or groups. An unusually high number of

successful vulnerability discoverers are from Eastern Europe, a region also known for its sophisticated vulnerability exploiters [67]. That may be attributable to a high degree of technical skill combined with weaker local economies. The rewards for finding vulnerabilities often come from international software organizations based in the United States.

Discovering vulnerabilities requires considerable technical and research skills. Some of the discoverers have an extensive background in software security. The responses by the top discoverers to our questions suggest that they tend to rely on their expertise and intuition rather than just the tools. The discoverers, like other well paid software professionals or researchers, expect to be fairly compensated for their services. With a few critical, high-severity vulnerabilities in hand, they may be in a position to bargain.

An attractive reward program based on vulnerability criticality can provide a significant alternative to the black market. A few software developers and security organizations now run a small number of such programs. These programs ensure time for patch development before a disclosure. Some of the top discoverers that we contacted suggest that sometimes the reward programs do not pay enough, and a better reward may be obtained on the black market, but none of them admitted to selling any vulnerabilities on the black market.

We note that after a few years of very successful vulnerability discovery, many of the top discoverers apparently disappear from the scene as credited discoverers. Some of them suggest that they find it more profitable to contract out their security auditing services to software developers.

The black market may often provide better rewards for some individual vulnerabilities than current legitimate programs. The black market might sometimes be more attractive because applying to reward programs may be tedious or slow. Limited awareness of reward programs

may also contribute to the attractiveness of the black markets. The reports suggest that many of the buyers in the black market may be affiliated with various governments [46], bringing a significant amount of money to the black market.

Companies and organizations need to design attractive vulnerability reward programs for their products. This will allow the legitimate markets to compete with the black market. Some reward programs, such as the one for Google Chrome, appear to have been successful. Google has a good reputation in technology as well as management, and has recognized the discoverers as high-achievement professionals. While the amount of money committed to the reward programs is only a tiny part of the company's revenues, Google is giving out some of the best monetary rewards.

A significant part of the global vulnerabilities market is quite opaque. Even the emerging legitimate markets have not been studied in detail, although some mathematical studies based on the classical market theories have appeared. There is a need to examine actual data and practices in order to understand the vulnerability discovery and disclosure.

### 4.6 Study Limitations

There is not yet enough data to start development of key hypotheses regarding the mechanics of multiple vulnerability markets. Considering the nature of the field, it will not be possible in the near future to obtain representative samples. Here, we explore some potential limitations of this research due to the size of the sample and its potential bias.

Sample Size: We tried to analyze the OSVDB dataset to find the top discoverers there because other databases do not include the names of discoverers. Unfortunately, the OSVDB database is not available for direct analysis. We relied on some manual analysis in addition to their reports to identify a larger number of top discoverers. We thus left out the discoverers who

have discovered only a small number of vulnerabilities. We attempted to locate the email addresses of the top discoverers on OSVDB and sent emails to the individuals. We were happy to note that several of the discoverers were willing to share information with us. It is unlikely that a significantly larger sample of top discoverers would be willing to participate in a study. We have also contacted some of the discoverers who have been active in several vulnerability rewards programs. Only three of them have responded so far.

Sample Bias: It is likely that those who responded to the questions were much more likely to be on the "white hat" side of the business. However, some of the respondents candidly acknowledged the lure of the black market, although none of them actually directly acknowledged having been a part of it. Innovative methods need to be developed that would allow researchers to better assess the black market in vulnerabilities.

# Chapter 5

# Estimation of Data Security Breach Costs: A Consolidated Approach

## 5.1 Introduction

Estimates of the costs incurred in such data breaches are published frequently. However, those estimates vary widely. Without proper disclosure by the impacted organizations, it is hard to compare the risk of information loss and its potential costs. Thus, the collection and aggregation of such information is vital. A careful quantitative analysis of data from the disclosures can be used to allocate resources for prevention and recovery after a breach. However, there is no established approach, although a few organizations such as the Ponemon Institute and NetDiligence are involved in data collection. This makes empirical cost estimation challenging. To assist in prioritizing potential actions and obtain more information on the nature of data breaches, the assessment of immediate and long-term costs, and the development of effective countermeasures, detailed studies are needed.

Data breaches impact a large fraction of all organizations. For instance, 43% of companies in the United States experienced a data breach in 2013 [75]. In 2015, there were 781 data breaches that were reported in U.S. companies (including small businesses), and according to the Identity Theft Resource Center [76], more than 169 million records were exposed. The cost of data breaches has increased 23% since 2013 according to studies conducted by IBM and the Ponemon institute [77].

According to the Ponemon Institute's studies of U.S. data breaches, the average cost of a compromised record has not changed significantly over the last five years. However, according to the NetDiligence studies, the average cost per record steadily increased between 2012 and 2015

(Figure 5.1) because the median number of records exposed was very small. The average data breach cost has also been increasing over the last three years (2013, 2014, and 2015) for most countries included in the Global Analysis study of the Ponemon Institute (Figure 5.2). For Germany and Australia, the average cost per record has varied slightly or remained almost the same. Some data costs of Canada and the Arabian Cluster (including Saudi Arabia and United Arab Emirates) are not available for some of the years.

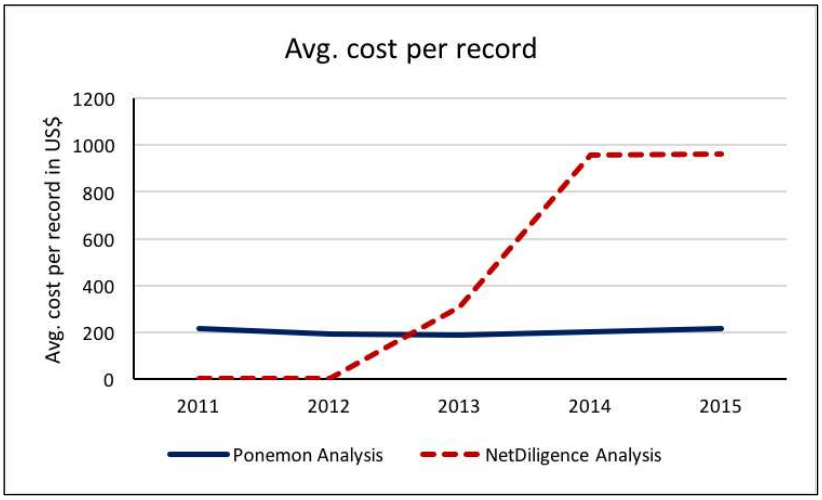


Figure 5.1: Average cost per compromised record over the last five years according to Ponemon and NetDiligence

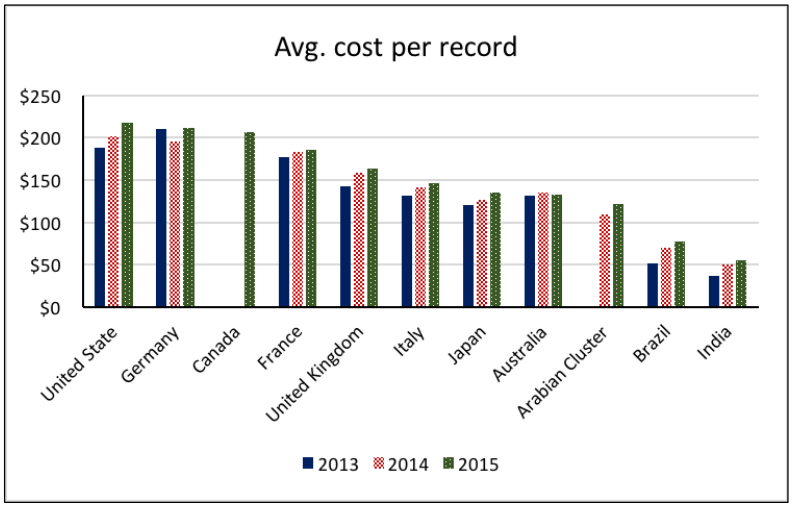


Figure 5.2: Average cost per compermised record of data breaches in 11 countries over the last three years according to Ponemon

A systematic study of data breach costs can help optimize data breach response plans, and this can potentially lead to a reduction in the costs related to breaches of up to 47% according to a Symantec/Ponemon Institute study [78].

The objective of this study is to examine, compare, and evaluate the existing approaches for assessing data breach costs and identifying the issues that need to be addressed. This is a major challenge that requires significant effort for the methods to have good predictive capability, as can be seen from the following sections. Some organizations have developed online cost calculators based on their collection of data and modeling assumptions. Unfortunately, the computations incorporated in the calculators are proprietary. Since these tools represent a significant effort by individuals with access to detailed data, we begin by examining these data breach cost calculators. The calculators have been implemented independently and vary considerably in their approaches. The results obtained by these calculators can be different by as much as two orders of magnitude. We attempt to extract the trends and logical assumptions used in the model and identify the questions that need to be resolved to obtain a consolidated approach that can be explained and calibrated using available data.

Utilizing an exhaustive range and combinations of inputs, we examine how each calculator computes the cost of data breaches for difference scenarios. We identify the factors that affect the costs of data breaches used in the various calculators and then classify those factors into logically related categories. We then attempt to identify the factors that are correlated in the same category to remove redundancy among factors. Further simplification is possible by identifying factors that have a significant impact in these calculators as well as ones that have little or no impact. We then model an approach needed to construct complete systemic models for data breach costs based on

the calculators and actual data regarding data breach costs. We validate components of the model using data from some recent well-documented breaches.

## 5.2 Background

This section presents background information necessary to discussing data breaches and their impact.

### 5.2.1 Types of Data Breaches

Data breaches have significant impacts on the economics and finances of individuals, organizations, and in some cases may even impact segments of the economies in various countries. There are several classes of data breaches, such as misuse of systems by employees or partners, intrusions with no theft of data, intrusions leading to the theft of personal data such as credit card numbers or social security numbers, and intrusions leading to the theft of intellectual property or confidential correspondence/documents.

### 5.2.2 Causes of Data Breaches

There are several possible causes of data breaches. These include exploiting weak or stolen login credentials, malware, presence of vulnerabilities or accessible back doors, and social engineering.

### 5.2.3 Economic Costs (impact) of Data Breaches

The cybercrimes that occur following data breaches and the impact of these crimes on the economy in terms of damage and cost can be enormous. In many cases, the damages to individual organizations have been estimated in the hundreds of millions of dollars. Some costs are not easily measurable, such as impact on segments of the economy or national security.

### 5.2.4 Main Cyber Insurers

Several insurance companies cover parts of the costs due to security breaches. For some of the breaches, a significant fraction of the costs was covered by insurance. The data collected by

NetDiligence, which was used for the Hub International Calculator, was based on insurance claims.

### 5.2.5 Databases of Data Breaches

There are several databases that provide information about security data breaches and help researchers analyze and contribute their results to reduce the impacts that result from data breaches. The main databases include Privacy Rights Clearinghouse (PRC) [79], DATALOSSdb [80], the Veris Community Database (VCDB) [81], and Web Hacking Incident Database (WHID) [82]. While the information in these databases was not directly applicable to this phase of the study, they can be valuable for validation of results once a computational model, based on the observations here, has been constructed.

## 5.3 Approach

To study the existing approaches for calculating the cost of data security, we identified and examined the existing data breach cost calculators that are available online. While there exist some reports on the overall observations of the data that some of the calculators are based on, details of the approaches taken by the calculators are not available. These calculators are intended for estimating the costs of data breaches. They require values of several factors as inputs. Here we examined the factors impacting the calculations and attempt to classify the factors into those impacting breach costs and those impacting the probability of breaches. Then, we developed a methodology for obtaining the breach costs (in most cases, the cost is per record) for several values of each factor. In some cases, we examined several combinations of factors that could have a mutual impact. Some calculators only generate the cost while others also generate the probability of a breach of the type examined.

After collecting the necessary data, we systematically identified the factors that appeared to be somewhat redundant. The redundant factors were merged. We also identified the factors that

emerged as significant and removed the insignificant factors. Some calculators apparently do not use any built-in data and require users to input all the parameters needed for computations. In addition, some of the calculators did not provide any breakdown of the overall costs and thus were not useful at this time.

We noted that using different calculators can result in very different estimates. To evaluate the calculators, we examined two major cases involving massive data breaches in large companies.

## 5.4 Analysis of Calculators

Here we examine the existing online data breach cost calculators and the factors they use. We use a large number of different input combinations to estimate how the calculators (in some cases: some calculators) compute both the cost and probability of a specific data breach.

### 5.4.1 Basic Information on Data Breach Cost Calculators

Various calculators for estimating data breach costs are available online for educational, illustrative, or other purposes. The calculators examined include Hub International/NetDiligence[83], Symantec/Ponemon Institute [84], MegaPath/Ponemon [85], Identity Theft (IDT911) [86], CyberTab [87], and IBM/Ponemon [88] (Table 5.1). Note that three of the calculators were implemented in conjunction with the Ponemon Institute, although they were released at different times. It is likely that they were implemented using the data collected immediately prior to the implementation of the calculator. Only two of the calculators compute the probability of the exploit of the type examined.

Table 5.1: data breach cost calculators examined

| Calculator | Date of Calculator | Data Source (powered by) | # Factors | Cost Computed | Computes Probability | Output |
|---|---|---|---|---|---|---|
| Hub Int'l & NetDiligence | 2012 | NetDiligence: 2011 Cyber Liability & Data Breach Insurance Claims | 7 | cost per breach | No | **Four partial costs:**<br>▪ Investigation<br>▪ Management<br>▪ Sanctions<br>▪ Lawsuit<br>**Overall results:**<br>▪ Total cost of breach<br>▪ Cost per record |
| Symantec & Ponemon | 2010 | 2011 Cost of Data Breach Study: United States. | 13 (9 only has impact) | cost per record | Yes | ▪ Avg. cost/record<br>▪ Avg. cost/breach<br>▪ Likelihood of data breach in next 12 months |
| MegaPath & Ponemon | 2013 | The Ponemon Institute study (2012) | 10 | - | No | ▪ Lower data breach cost estimate<br>▪ Upper data breach cost estimate |
| IDT911 | 2014 | Three different sources | 9 | cost per breach | No | ▪ Estimate 1st party response amount |
| CyberTab | 2014 | 2014 The Economist Intelligence Unit Ltd. report. | Many | cost based on user's input | No | ▪ Attack cost<br>▪ Security spending<br>▪ Prevention cost<br>▪ Return on prevention |
| IBM & Ponemon | 2014 | 2015 Cost of Data Breach Study: Global Analysis | 16 (14 have impact) | cost per record | Yes | ▪ Avg. cost/record<br>▪ Avg. cost/breach<br>▪ Chance of data breach in next 12 months |

#### 5.4.1.1 Hub International and NetDiligence Calculator

Hub International provides risk services, insurance brokers, and consultants. Since it was formed through the merger of 11 Canadian insurance brokerages in 1998, it has been a global insurance broker. Hub International launched its calculator (eRisk portal) on July 18, 2012 [89] to assist and protect information businesses against cyber loss. The calculator is powered by the data collected for a NetDiligence study that was apparently done in 2011 [90]. The calculator generates four partial costs as well as the total cost and the cost per record.

#### 5.4.1.2 Symantec and Ponemon Institute Calculator

This calculator is presented through the cooperation of Symantec (as sponsor) and the Ponemon Institute (as the organization behind the data source). The calculator was launched in 2010 based on archive.org archives, and its most recent version appears to be based on the data included in "2011 Cost of Data Breach Cost Study: United States" [91]. This study was based on the actual data breach experiences of 49 U.S. companies in 14 different industry sectors, including

data collected from 400 interviewed individuals and 268 organizations. It takes into account a wide range of direct and indirect business costs. The Ponemon Institute has been collecting similar data since 2005 and has published annual reports based on that data.

#### 5.4.1.3 MegaPath and Ponemon Institute Calculator

MegaPath Corporation is a provider of secure access solutions. It launched its data breach risk calculator on November 19, 2013 [92]. The calculator was developed by Ponemon Institute experts and is apparently based on the 2012 study, "Data Security in Small Healthcare Organizations" [92] based on the numbers used.

#### 5.4.1.4 IDentity Theft (IDT911) Calculator

IDT911 is a Canadian insurer of major credit unions and banks. It was established in 2003 [93] and focuses on providing services related to identity protection, identity theft recovery, and data risk management. Based on archive.org archives, it appears to have launched its calculator in 2014. The calculator is based on three sources of information: proprietary data collected from IDT911's experiences handling data breaches; information available to the public through many public, private, and non-profit groups and sources; and information collected from the users of a website that tracks breach statistics.

#### 5.4.1.5 CyberTab Calculator

CyberTab is a tool intended for information security and other senior executives for assessing the potential damage to their organizations from electronic attacks. It was released in 2014 [94]. Its calculator was developed by The Economist Intelligence Unit and sponsored by the well-known consulting organization Booz Allen Hamilton.

#### 5.4.1.6 IBM and Ponemon Institute Calculator

The IBM Calculator is sponsored by IBM and based on data used in the study "2015 Cost of Data Breach Study: Global Analysis [77]". This study is the tenth annual survey conducted by the

Ponemon Institute, and it includes data from 250 organizations across 12 countries: Australia, Brazil, Canada, France, Germany, India, Italy, Japan, UK, USA, Saudi Arabia, and the UAE. The calculator appears to have been launched in 2014.

### 5.4.2 Factors used as determinants of Data Breach Costs

After analyzing the factors used to estimate the costs of data breaches in the data breach calculators, we divided the main factors into two main groups—those affecting the data breach costs and those affecting the likelihood of a breach—based on our logic and judgment. Most calculators calculate only data breach costs, but a few estimate the likelihood of breaches as well. Table 5.2 shows some of the factors that are common among calculators. The Cost group includes the following subgroups/factors: the number of affected records, the breach type/business type, business size indicators, data location, credit monitoring years, and a group of four binary (Yes/No) factors. The second group includes the main factors affecting the probability of a data breach. They include: organizational attributes (technical, commitment to security, policies), industry classification, and history of data breaches. Table 5.2 clearly demonstrates that there is considerable divergence among the calculators, which may be based on the different perspectives used by the calculator developers.

Table 5.2: Comparable factors in data breach cost calculators

| Classification | | Factors | Hub Intl & NetDiligence | Symantec & Ponemon Institute | MegaPath & Ponemon Institute | IDentity Theft 911 | CyberTab | IBM& Ponemon Institute |
|---|---|---|---|---|---|---|---|---|
| *Cost of Data Breaches* | *Number of breached records* | Total number of affected records | X | X | X | X | X | X |
| | *Type of data breached /business Type* | Types of information do employees handle | X | X | X | X | X | X |
| | *Business size* | Global headcount of organization (size) | | X | X | | X | X |
| | | Global footprint of organization (branches) | | X | | X | | X |
| | | Organization's headquarters or where does it conduct the majority of its business | | X | | | X | X |
| | *Data location* | Location of organization's data store | X | X | | X | | |
| | *Credit monitoring years* | Number of years of credit monitoring | X | | | | | |
| | *Some binary factors* | Actual fraud has occurred | X | | | | | |
| | | PCI compliance has an issue | X | | | | | |
| | | Federal class action lawsuit has been filed | X | | | | | |
| | | How does the business store information? Paper or electronic | | | | X | | |
| *Probability of Data Breaches* | *Technical* | Employees, partners, or other can access sensitive information, networks or applications from a local or remote location | | X | | X | | |
| | | Organization's practices concerning the transfer of sensitive personal information to other organizations | | | X | | | |
| | | Percentage of company's IT operations are conducted in public or semi-public cloud environments | | | X | | | |
| | *Commitments to security* | Sensitive data encrypted on all laptops or removable storage during transmission, storage, or retrieval of data | | X | | X | | X |
| | | Organization has a dedicated information security leader, such as a chief information security officer (CISO) | | X | | | | X |
| | | Strong authentication measures have been deployed to protect sensitive networks and applications from unauthorized access | | X | | | | |
| | | Organization has a data breach incident response plan in place today by its business continuity management (BCM) team | | | X | | | X |
| | | Organization's leaders and/or business owners consider information security and data protection a priority | | | X | | | |
| | | Organization's security budget adequate for preventing material data breach incidents | | | X | | | |
| | | How long does the business retain sensitive information pertaining to employees, customers, and patients? | | | | X | | |
| | *Policies* | Privacy and data protection security policy | | X | | | | X |
| | *Industry classification* | Organization's industry classification | | X | X | X | X | X |
| | *History* | Organization experienced a material data breach within the past 12 to 24 months | | | X | | | X |
| | | Did your organization ever suffer a data breach caused by a third party's negligence or criminal acts? | | | | | | X |
| | | Did your organization ever suffer a data breach caused by the loss or theft of a laptop, smart phone, tablet, USB thumb drive, or other portable data bearing devices? | | | | | | X |
| | | Did your organization ever suffer a data breach incident that was publicly disclosed to the media, advocates, and/OR regulators? | | | | | | X |
| | *External* | Likely cause of a data breach | | X | | | X | X |

### 5.4.3 Analyzing the Data Breach Cost Calculators

Careful and thorough exercising of the available data breach cost calculators is critical for modeling their functionality in terms of the factors they consider. We carefully compiled the outputs generated using a large number of strategically used input combinations. We summarize our observations as follows.

#### 5.4.3.1 Hub International and NetDiligence Calculator

The Hub International Calculator uses seven factors and computes the four partial costs for the two types of data breaches (records): (1) Personal Health information (PHI) and Social Security Number (SSN), and (2) credit cards (CC). The partial costs are influenced by the following binary factors:

***a) Whether the data is centralized store,***

***b)Whether actual fraud (actual fraud means the data has been stolen and used already before) has occurred,***

***c)Whether there has been a Payment Card Industry (PCI) compliance issue, and***

***d) Whether a federal class action lawsuit has been filed.***

The overall cost is divided into four partial costs: i) costs associated with incident investigation, ii) customer notification and crisis management, iii) regulatory and industry sanctions, and iv) class action lawsuits. Not all factors directly impact all the partial costs. Therefore, we have analyzed each partial cost individually. For the analysis, we examined all 16 binary cases (since there are 4 different binary factors). Each case is represented using 4 bits, starting from 0000 to 1111, where each bit represents each binary factor as Yes/No. We used sample sizes from 1000 to 100 million for the number of records compromised. The sizes of affected records were incremented by a multiplication factor of 10 (1000, 10000, up to 100,000,000). Then, we collected the partial cost values generated for each size for each binary case. We repeated the analysis for each of the two types of data breaches with one year of credit monitoring, which is generally offered to affected customers.

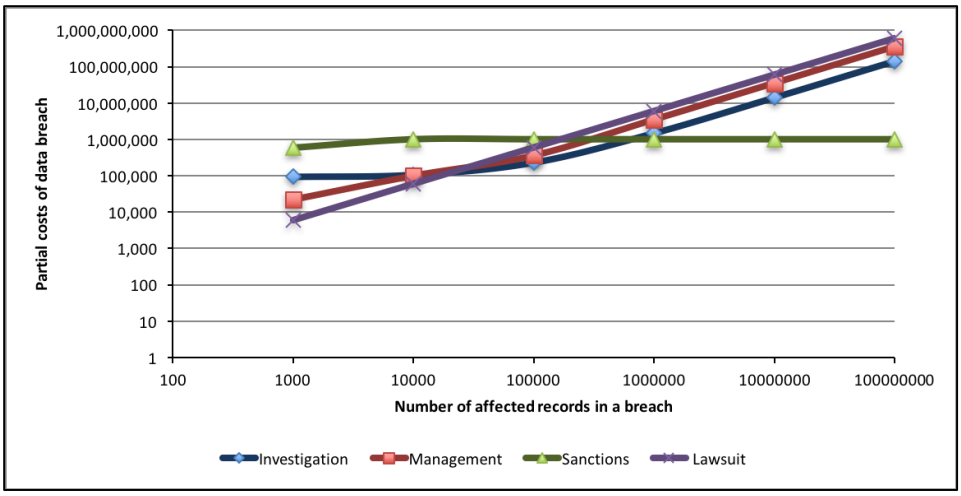


Figure 5.3: Partial costs from Hub International calculator for credit cards for the binary case: 0000 (all four factors a, b, c, d are false)

We obtained the relationship of each partial cost with the number of records from each binary case using the trendline of each chart. An example is shown in Figure 5.3. Generally, power regression provided the good fit. Next we analyzed for all power equations (y= ax^b) for each partial cost for all binary cases. We made two main observations, as discussed below.

Table 5.3: Parameters a and b for each partial costs of credit cards type in Hub International depending on the binary factor

| **Partial costs** | **Impacting factor** | **Avg. a** | | **Avg. b** | |
|---|---|---|---|---|---|
| | | ***if Yes*** | ***if No*** | ***if Yes*** | ***if No*** |
| ***Investigation*** | Centralized | 1087.33 | 474.83 | 0.53 | 0.65 |
| | Fraud | 797.69 | 764.47 | 0.59 | 0.59 |
| | Lawsuit | 846.03 | 716.12 | 0.60 | 0.58 |
| ***Crisis management*** | Credit monitoring | 41.50 | 41.50 | 0.84 | 0.84 |
| ***Sanctions*** | PCI comp. | 11308 | 610611 | 0.47 | 0.03 |
| ***Lawsuit*** | Fraud | 5.56 | 5.9 | 1.02 | 0.89 |
| | Lawsuit | 5.4 | 6 | 0.90 | 1 |

Table 5.4: Parameters a and b for each partial costs of PHI and SSN types in Hub International depending on the binary factor

| Partial costs | Impacting factors | Avg. a | | Avg. b | |
|---|---|---|---|---|---|
| | | *if Yes* | *if No* | *if Yes* | *if No* |
| ***Investigation*** | Centralized | 1482.9 | 652.19 | 0.53 | 0.65 |
| | Fraud | 1091.63 | 1043.46 | 0.59 | 0.59 |
| | Lawsuit | 1168.91 | 966.18 | 0.60 | 0.58 |
| ***Crisis management*** | Credit monitoring | 60.71 | 60.71 | 0.84 | 0.84 |
| ***Sanctions*** | PCI comp. | 19145 | 865754 | 0.43 | 0.02 |
| ***Lawsuit*** | Fraud | 0.18 | 0.04 | 0.58 | 0.58 |
| | Lawsuit | 0.22 | 0 | 1.15 | 0 |

First, as shown in Tables 5.3 and 5.4, we can observe the impacted factors for each partial cost in Hub International. For instance, the incident investigation costs are impacted by the factors of centralized data, actual fraud, and action lawsuit. The column marked *Yes* indicates the numbers when the factor applies. The cost of incident investigation will be high in cases of non-centralized data, actual fraud, or the filing of a federal class action lawsuit since the costs of forensic investigation, security remediation, legal defense, litigation, and damages are high. In addition, customer notification/crisis management costs are impacted by only one factor: the number of years of credit monitoring for each victim. Usually, the average length of credit monitoring is one year. The third cost, regulatory and industry sanctions, is affected by one factor, which is the existence of issues related to PCI compliance. This cost will be high in cases of PCI compliance issues for large numbers of affected records. The last cost, the cost of a class action lawsuit, is impacted by two factors: the occurrence of actual fraud and the filing of a federal class action lawsuit. If the two factors are applied, costs will be high since eDiscovery litigation, legal defense, and damages are also high. In addition, we collect both costs of a and b for each impacted factor in each partial cost when the factor has impacted the partial cost or not.

Table 5.5: Impact of binary factors for credit cards in Hub International calculator in power equation y=ax^b

| Partial Costs | a | b |
|---|---|---|
| ***Investigation*** | (0xx0 +0xx1)= 0xxx<br>(1xx1+1xx0)= 1xxx | (1xx0+1xx1)=1xxx<br>(01x1+00x1)=0xx1<br>0xx0<br>=====<br>1xxx<br>0xxx |
| ***Observations*** | Centralized data decreases expected investigation cost. While, lawsuit has some impact. | Centralized data increases expected investigation cost. Lawsuit factor has some impact. |
| ***Crisis Management*** | 1 (xxxx) | 1 (xxxx) |
| ***Observations*** | Constant regardless of the other factors. | Constant regardless of the other factors. |
| ***Sanctions*** | xx1x+xx0x=1(xxxx) | (xx0x+xx1x)=1(xxxx) |
| ***Observations*** | PCI compliance decreases expected sanctions cost. | PCI compliance increases expected sanctions cost. |
| ***Lawsuit*** | x1x1<br>00x1 + x011<br>xxx0<br>1001 | 1001<br>xxx0<br>00x1 + x011<br>x1x1 |
| ***Observations*** | Actual fraud has some impact, but lawsuit factor has the main impact on the lawsuit cost. | Actual fraud has some impact, but lawsuit factor has the main impact on the lawsuit cost. |

Table 5.6: Impact of binary factors for PHI and SSN in Hub International calculator in power equation y=ax^b

| Partial Costs | a | b |
|---|---|---|
| ***Investigation*** | (0xx0 + 0xx1)= 0xxx<br>(10x1+ 11x1)= 1xx1<br>1xx0<br>======<br>0xxx<br>1xxx | (1xx0 + 1xx1)= 1xxx<br>(0xx1+ 0xx0)=0xxx |
| ***Observations*** | Centralized data decreases expected investigation cost. Other factors have some impact. | Centralized data increases expected investigation cost. Lawsuit factor has some impact. |
| ***Crisis management*** | 1 (xxxx) | 1 (xxxx) |
| ***Observations*** | Constant regardless of the other factors. | Constant regardless of the other factors. |
| ***Sanctions*** | (xx1x+ xx0x)=xxxx | (xx0x+ xx1x)= xxxx |
| ***Observations*** | PCI compliance decreases expected sanctions cost. | PCI compliance increases expected sanctions cost. |
| ***Lawsuit*** | xxx0<br>(x0x1+ x1x1)= xxx1 | xxx0<br>(x0x1+ x1x1)= xxx1 |
| ***Observations*** | Lawsuit increases expected lawsuit cost, while actual fraud has some impact. | Lawsuit increases expected lawsuit cost, while actual fraud has some impact. |

Second, Tables 5.5 and 5.6 demonstrate two important facts: which factor has significant impact on the partial cost and what is the initial status of the factor (i.e., if the factor impacts a or b). However, we examine how a and b work in the power equation y = ax^b, where x is the number of affected records in the data breach. We use the binary cases of each a and b. We use the Logisim program to minimize the binary combinations to one case if possible, or as few as possible. Finally, we observe how a and b are impacted by the factors of the calculator. All the observations are mentioned under each partial cost. In Tables 5.5 and 5.6, xxxx means the four factors that impact or might impact the partial costs. The first digit of xxxx represents the

centralized data, the second represents actual fraud, the third represents PCI compliance, and the fourth represents action lawsuit. A one or zero in xxxx indicates the transmission impact of the factor, while x means the factor does not impact the partial cost. For instance, if a has 1xx0 and 1xx1, that means that the data is 1xxx - combination just for one value different, and that means it is centralized data has full impact on the partial cost.

Table 5.7: Power relations for the partial costs in the Hub International

| Partial Costs | Partial Costs Equations |
|---|---|
| ***Investigation Cost*** | **CC:** (474.83 * Centralized 1 factor) * (# of affected records ^ (0.53 * Centralized 2 factor)) (5.1)<br><br>Centralized 1 factor= 1, if the data is not centralized, 2.29 otherwise.<br>Centralized 2 factor=1, if the data is centralized, 1.23 otherwise. |
| | **PHI & SSN:** (652.19 * Centralized 1 factor) * (# of affected records ^ (0.53 * Centralized 2 factor)) (5.2)<br><br>Centralized 1 factor = 1, if the data is not centralized, 2.27 otherwise.<br>Centralized 2 factor = 1, if the data is centralized, 1.23 otherwise. |
| ***Crisis management Cost*** | **CC:** 41.5 * (# of affected records ^ 0.84) (5.3) |
| | **PHI & SSN:** 60.71 * (# of affected records ^ 0.84) (5.4) |
| ***Sanctions Cost*** | **CC:** (11308 * PCI complaince1 factor) * (# of affected records ^ (0.03 * PCI complaince2 factor)) (5.5)<br><br>PCI compliance1 factor= 1, if it is an issue, 56 otherwise.<br>PCI complaince2 factor= 1, if it is not an issue, 15.7 otherwise. |
| | **PHI & SSN:** (19145 * PCI complaince1 factor) * (# of affected records ^ (0.02 * PCI complaince2 factor)) (5.6)<br><br>PCI compliance1 factor= 1 if it is an issue, 45.22 otherwise.<br>PCI complaince2 factor= 1, if it is not an issue, 21.5 otherwise. |
| ***Lawsuit Cost*** | **CC:** (5.4 * Lawsuit 1 factor) *( # of affected record ^ (0.90 * Lawsuit 2 factor) (5.7)<br><br>Lawsuit 1 factor= 1, if lawsuit is filed, 1.11 otherwise.<br>Lawsuit 2 factor = 1, if lawsuit is filed, 1.11 otherwise. |
| | **PHI & SSN:** a * (# of affected records ^ b) (5.8)<br><br>a= 0, if lawsuit is not filed, 0.22 otherwise.<br>b= 0, if lawsuit is not filed, 1.15 otherwise. |

From the previous tables and observations, the partial costs equations for both types of data breaches are computed using the observation factor change status in Tables 5.5 and 5.6 and the amount of change from Tables 5.3 and 5.4. The partial costs equations in case of the factor are impacted in the following way in Table 5.7: incident investigation cost in equations (5.1) and (5.2), customer notification and crisis management cost in equations (5.3) and (5.4), regulatory and industry sanctions cost in equations (5.5) and (5.6), and the class action lawsuit cost in equations (5.7) and (5.8).

After we compute each partial cost in the Hub International calculator, the total cost of breach and, then, the cost per record can be computed in equations (5.9) and (5.10), respectively. The total cost will generally be affected by the type of data breach and the number of affected records. Here we have not examined the computation of probabilities.

$$\begin{aligned}\boldsymbol{Total\ Cost\ of\ a\ Breach} &= Incident\ investigation\ cost \\ &+ Customer\ notification\ or\ crisis\ management\ cost \\ &+ Regulatory\ and\ industry\ sanctions\ cost \\ &+ Class\ action\ lawsuit\ cost \end{aligned} \tag{5.9}$$

A widely used measure is cost per record, as defined below.

$$\boldsymbol{Cost\ per\ Record} = \frac{Total\ costs\ of\ breach}{Total\ number\ of\ affected\ records}, \tag{5.10}$$

where the number of affected records is entered by a user/organization.

#### 5.4.3.2 Symantec & Ponemon Institute Calculator

There are 13 factors impacting the cost and probability of a data breach. Four of these factors do not directly impact cost or probability in the computation; instead they are used to determine the currency of the cost and the number of affected records to be used when the total cost of breach is calculated. For most of the rest of factors (six out of nine), there is a correlation between the cost per record and the probability, such that the maximum cost per record matches the maximum probability and the minimum cost per record matches the minimum probability.

Ultimately, this calculation represents two costs and one probability: the average cost per record (taken by averaging the cost for each of the 9 factors), the average cost per breach (obtained by multiplying the average cost per record by the number of affected records), and the probability of experiencing a data breach in the next year (obtained the taking the average of the probabilities for each of the 9 factors).

#### 5.4.3.3 MegaPath & Ponemon Institute Calculator

The information for this calculator is limited since there are no data about the probabilities or the cost per record for each of the 10 factors. Therefore, we do not know exactly how the calculator estimates the overall data breach cost as a range from a lower cost to an upper cost of data breach. However, it appears that the calculator uses minimum and maximum costs per factor and calculates an average to provide the lower and upper costs based on 1,000 affected lost or stolen records as mentioned on the result report.

The media report published by MegaPath Corporation in 2013 mentioned that, in addition to calculating breach costs, the calculator could compute the probability of a data breach within the next 12 months [92]. To do so, the calculator generates five sub-steps of data breaches: detection and escalation, containment and remediation, information loss, business interruption, and reputational impact. From our experiments, we could not determine all the previous analyses related to the calculation of the data breach cost since we did not have actual data and we did not receive a response from MegaPath when we contacted it.

#### 5.4.3.4 IDentity Theft (IDT911) Calculator

This calculator is impacted by nine factors. Each factor includes several options for estimating cost through a one-part response. The calculator does not include a third-party estimation. It is visually organized like a lookup table for different factors or options. The total cost is determined by computing the cumulative costs of all factors (equation 5.11). However, there is no information on probability or the number of affected records to allow us to calculate the cost per record.

$$\boldsymbol{Estimate\ 1st\ party\ response\ amount} = summation\ of\ the\ the\ costs\ for\ all\ the\ \ factors \tag{5.11}$$

#### 5.4.3.5 CyberTab Calculator

As with the other calculators, CyberTab offers no information on the number of affected records or the probability of a data breach. This calculator is based on users' estimates of incident response, business expenses and losses, and returns on prevention. Users answer many factors to compute different costs. One of the costs is attack or breach cost, which contains the different costs when the data breach occurs (equation 5.12). The incident response expenses include technology costs and outsourced third-party services costs such as consulting and forensic services. The business expenses or costs include c-suite and executives cost, customer services cost, credit monitoring cost, legal services cost, crisis management expenses, notifications and fines. Meanwhile, lost business includes reputation loss such as grant loss, lost sales, customer loss, and customer retention.

The CyberTab calculator can be used in one of two modes: a planning mode, which gauges the potential cost of an attack to determine and understand the risks and the choices for security investment, and reporting mode, which discovers the specific attack costs that remain for a company and which assists in benchmarking.

$$\boldsymbol{Attack\ Cost} = Incident\ response\ expenses + business\ cost + lost\ business \tag{5.12}$$

#### 5.4.3.6 IBM & Ponemon Institute Calculator

This calculator is similar to the Symantec and Ponemon Institute calculators since their data come from the same source. However, IBM uses newer data from the 2015 Ponemon Institute study. IBM incorporates 16 impacted factors, 3 of which have no direct impact on the cost or probability of a data breach. These are used to determine the currency of the cost or the number of affected records—both of which are used to compute the total cost of the breach.

The calculator calculates the average cost per record, the average cost per breach, and the likelihood of a data breach in the next 12 months.

We observed that, for six factors, there is a correlation between the cost per record and the probability, such that the maximum cost per record has the maximum probability and the minimum cost per record has the minimum probability. In addition, three factors exhibit the opposite behavior, such that the maximum cost per record matches the minimum probability and the minimum cost per record matches the maximum probability.

### 5.5 Proposed Consolidated Model of Data Breach Costs

After examining the existing calculators, we built the framework for a new model (approach) to calculate the total cost of a breach. It incorporates the likelihood of a breach occurring within the next 12 months.

We constructed our model based on the factors that are numerous in different calculators. Some of these are significant, and some are not. Thus, we need to know the following information:

- What factors most impact the computations?
- Factor classification: If two factors are in the same category, are they correlated?
- Factors: What are the factors' relationships to cost/probability (linear, non-linear, logarithmic, exponential, or power)?

To achieve the first of these two goals, we use a systematic approach using the following steps. The third question requires further research, although this study provided some background.

First, the factors whose calculators have numbers of cost per breach/record or probability (not all calculators calculate probability) are combined in one table. We use the factors of the four

calculators only: Hub International/NetDiligence, Symantec/Ponemon, IDT911, and IBM/Ponemon. After that, the factors are classified into six categories (Table 5.8).

Table 5.8: Factors impacting the partial costs before and after consolidation

| Classification | Item | Factors | Source | Risk score | Significant factors | Insignificant factors |
|---|---|---|---|---|---|---|
| ***Total number of affected records*** | A | Total number of affected records? | All | - | X | |
| ***Type of data breaches*** | B1 | What is your organization's industry classification? | Symantec & IBM | 3.42 | X | |
| | B2 | What types of information do your employees handle? | Symantec & IBM | 4.81 | X | |
| | B3 | On whom do you collect sensitive or personally identifiable information (PII)? | IDT911 | Redundant, removed. | | |
| | B4 | What types of information does the business collect on employees? | Hub Int'l | Redundant, removed. | | |
| ***Incident investigation cost*** | C1 | Data is in a centralized system/location? | Hub Int'l | - | X | |
| | C2 | Actual fraud is expected already? | Hub Int'l | - | X | |
| | C3 | Federal class action lawsuit filed? | Hub Int'l | - | X | |
| | C4 | What do you think is the most likely cause of a data breach? | Symantec & IBM | 2.94 | X | |
| | C5 | Do your employees store sensitive data on laptops or on removable storage? | Symantec | 1.28 | | X |
| | C6 | Is sensitive data encrypted on all laptops or removable storage? | Symantec & IBM | 3.31 | X | |
| | C7 | Did your organization ever suffer a data breach caused by a third party's negligence or criminal acts? | IBM | 1.86 | | X |
| | C8 | Did your organization ever suffer a data breach caused by the loss or theft of a laptop, smart phone, tablet, USB thumb drive, or other portable data bearing device? | IBM | 1.22 | | X |
| | C9 | Did your organization suffer a data breach incident involving the loss or theft of 5,000 or more records over the past 24 months? | IBM | 1.84 | | X |
| | C10 | Did your organization ever suffer a data breach incident that was publicly disclosed to the media, advocates, and/or regulators? | IBM | 1.76 | | X |
| | C11 | Who has access to sensitive information on employees, customers, and patients? | IDT911 | - | | X |
| | C12 | How does the business store information? (paper or electronic) | IDT911 | - | | X |
| | C13 | How does the business store information? (on-site, off-site, cloud) | IDT911 | Redundant, removed. | | |
| | C14 | Is document, email, and device/hard drive encryption used to protect sensitive electronic information about employees, customers, and/or patients during transmission, storage, and retrieval of data? | IDT911 | Redundant, removed. | | |
| | C15 | How long does the business keep/retain sensitive information pertaining to employees, customers, and patients? | IDT911 | - | X | |
| | C16 | What best describes your organization's privacy and data protection program? | Symantec & IBM | 2.5 | X | |
| ***Crisis management cost*** | D1 | Number of Years for credit monitoring? | Hub Int'l | - | X | |
| | D2 | What is the global headcount of your organization? | Symantec & IBM | 1.93 | X | |
| | D3 | What is the global footprint of your organization? | Symantec & IBM | 1.62 | | X |
| | D4 | Does your organization have a dedicated information security leader, such as a chief information security officer (CISO)? | Symantec & IBM | 1.69 | | X |
| | D5 | Is your organization's business continuity management team involved in the data breach incident response process? | IBM | 1.73 | X | |
| | D6 | Approximately how many employee records do you store and retain that contain social security numbers, account numbers, medical records information, and/or other personal information? | IDT911 | - | | X |
| | D7 | How many states does the business have employees, customers, and patients in? | IDT911 | - | | X |
| ***Regulatory and sanction cost*** | E | Is PCI compliance an issue? | Hub Int'l | - | X | |
| ***Lawsuit cost*** | F1 | Actual fraud is expected already? (This factor impacts again here) | Hub Int'l | - | X | |
| | F2 | Federal class action lawsuit filed? (This factor impacts again here) | Hub Int'l | - | X | |

Second, all the repeated/correlated factors should be removed (such as items B3 and B4 in Table 5.8) after determining that the factors are closely related to others (items B1 and B2). We chose the new costs data like IBM, which uses cost per record, not like IDT911. In addition, items C13 and C14 are removed since there are similar to items C5 and C6, and we use them because we use the newer data and the factors provide more options.

Third, the type of data breach affects all the partial costs except the security upgrade. Thus, the two factors (B1 and B2) are significant.

The significance of a factor is determined using this approach:

(1) For each partial cost classification: make the average and standard deviation (STDEV) for all the averages of ratio for the cost and probability (max cost / min cost) and (max probability / min probability). However, we tested IDT911 alone since its cost is not per record.

(2) Choose significant factor if it is bigger than the following equation that distributes data normally: (average cost ratio or average probability ratio + 0.5 X STDEV).

(3) Choose insignificant factors if they are smaller than the following equation: average cost ratio or average probability ratio - 0.5 X STDEV (Table 5.9).

Table 5.9: Identifying the significant factors for some categories in the consolidated model

| Variation Attributes | Type of Data Breaches | | Incident investigation | | | Crisis management | | |
|---|---|---|---|---|---|---|---|---|
| | Avg. cost ratio | Avg. probability ratio | Avg. cost ratio | | Avg. probability ratio | Avg. cost ratio | | Avg. probability ratio |
| | *Calculators except IDT911* | | *IDT911* | *Other calculators* | | *IDT911* | *Other calculators* | |
| Avg.= | 1.8 | 2.29 | 6.33 | 1.46 | 1.4 | 10 | 1.35 | 1.29 |
| STDEV= | 0 | 0.54 | 5.13 | 0.34 | 0.24 | 0 | 0.12 | 0.06 |
| Avg. + 0.5 * STDEV = | 1.8 | 2.56 | 8.9 | 1.63 | 1.52 | 10 | 1.41 | 1.32 |
| Avg. - 0.5 * STDEV = | 1.8 | 2.02 | 3.77 | 1.29 | 1.28 | 10 | 1.29 | 1.26 |
| Significant factor | >1.8 | >2.56 | >8.9 | >1.63 | >1.52 | >10 | >1.41 | >1.32 |
| Insignificant factor | <1.8 | <2.02 | <3.77 | <1.29 | <1.28 | <10 | <1.29 | <1.26 |

(4) Between them: use equation (5.13) to compute the risk scores for the calculators [95] (Table 5.8).

$$\boldsymbol{Secuirty\ risk} = data\ security\ breach\ cost\ X$$
$$probability\ of\ chance\ data\ security\ breach\ within\ a\ year \quad (5.13)$$

(5) Use our judgment to determine which factor is significant (high risk) to remain and insignificant (low risk) to remove. That judgment is based on comparison between risk values in the same partial cost category. For instance, we determine each factor in incident investigation classification, and those with a risk score less than 2 are insignificant while those with a score higher than 2 are significant factors. While all factors in crisis management had a risk score under 2, we chose two factors as significant factors based on our subjective consideration since these factors were definitely important according to the cost ratio and probability ratio.

(6) Any insignificant should be removed.

(7) All the factors of Hub International are considered directly significant, and they were not tested in the previous steps because they are essential factors of one of the main sources of our model.

(8) For the factors of IDT911, any factor that has an average cost ratio larger than 8.9/10 and that we think is useful in our model is considered significant. There is no risk score since there is no probability for these factors.

Based on our study of existing approaches, we build a comprehensive model that incorporates all the factors affecting the cost while attempting to simplify the approach, as discussed in chapter (6).

# Chapter 6

# Quantitative Assessment of Data Security Breach Risk: Cost and Likelihood

## 6.1 Introduction

After taking into consideration the significant factors that impact data breach risks from all the available data breach calculators in the previous chapter, it is evident that a quantitative model that takes a systematic approach for total cost based on number of records is required. That model should include the concept of economy of scale to illustrate why the data breach cost per record figure is misleading. In addition, a quantitative model for probability of a breach in the next year is important part during estimation of data breach risk.

## 6.2 Applicability of Existing Models

As discussed in a previous chapter, there exit two major computational models with their own data sources: Ponemon, which created calculators with sponsorships from Symantac (2010), Megapath (2013), and IBM (2014), and NetDiligence, which created the Hub International calculator (2012) and contributed to the Verizon report. There have been no critical studies of the two approaches. This study presents a preliminary comparison using detailed data from two recent widely discussed breaches: 2013 Target breach and 2014 Home Depot breach (Tables 6.1 and 6.2). We also examine the recent claims made by Jay Jacobs of Verizon, who collaborated with NetDiligence [104] who has been critical of the computations model used by Ponemon because the Ponemon approach yields a cost per record that he believes is too high.

The difference between the two models is readily illustrated using the average cost/record implicit in the two approaches. Table 6.3 gives the cost per record according to the 2015 Ponemon study, which comes out to be about $217. Table 6.4 gives the cost per record as computed by our

analysis for Hub International/NetDiligence (regardless the average static loss), which is remarkably different: $12.57 for credit cards and $8.58 for personal information. The 2015 Verizon Data Beach Investigations Report (DBIR) finds the average cost per record to be $0.58 [96] by applying the Ponemon formula (all breach costs divided by compromised records: $400 million / 700 million records= $0.58). The two to three orders of magnitude difference raises several questions.

Table 6.1: Target data breach actual reported costs

| Years | Gross Expenses | Insurance receivable | Net Expenses (before tax deductions) | Net Expenses (after tax deductions) |
|---|---|---|---|---|
| 2013 | $61m | $44m | $17m | $11m |
| 2014 | $191m | $46m | $145 m | $94m |
| 2015 | N/A | N/A | $39 | $28 |
| Total | $252m | $90m | $201m | $133m |
| Raw cost per card= $6.30 (40 million cards affected) | | | | |

Table 6.2: Home Depot data breach actual reported costs

| Years | Gross Expenses | Insurance receivable | Net Expenses (before tax deductions) | Net Expenses (after tax deductions) |
|---|---|---|---|---|
| Q3, 2014 | $43m | $15m | $28m | N/A |
| Q4, 2014 | $20m | $15m | $5m | N/A |
| Total | $63m | $30m | $33m | N/A |
| Raw cost per card= $1.13 (56 million cards affected) | | | | |

Both organizations use significantly large and diverse data sets. Ponemon collected data from more than 1600 companies in several countries. NetDiligence data from the 2015 Verizon (DBIR) [96] uses data from 191 cyber insurance payouts.

We can compare the cost/record with the now well-known numbers for the Target and the Home Depot breaches, which come out to $6.30 and $1.13 per record, respectively (Tables 6.1, 6.2). The sources for these numbers are well documented. Target Brands Inc. had a major incident of data breach on December 2013 involving 40 million credit and debit card records [97]. The

total estimation of the data breach was about $252 million, as shown in Table 6.1, which uses Target's financial statements [98, 99]. These numbers are closer to the NetDiligence numbers. The results are similar to the Hub International calculation (Table 6.3) if we exclude the expected lawsuit cost ($12.57 – $7.09 = $5.48) since a lawsuit has not yet occurred. The cost per record in this breach is not at all close to Ponemon's cost per record.

The other notable example is that of the Home Depot data breach in 2014, which involved 56 million customer payment cards [97]. The available information about the cost of this breach is given in Table 6.2 [100, 101]. We found that the cost per card was $63m/56m = $1.13. This cost is much lower than` what would be estimated using the Ponemon calculators, and it is in fact closer to the Hub International estimate.

Table 6.3: Average cost per record according to Ponemon Study: 2015 cost of data breach in United States

| Partial costs | Avg. cost per breach | Avg. cost per record |
|---|---|---|
| Detection & escalation (includes investigation and crisis management) | $610,000 | $21.73 |
| Notification (includes notification and determination of regulatory) | $560,000 | $19.95 |
| Ex-post response (includes regulatory and lawsuit) | $1,640,000 | $58.43 |
| Lost business (includes reputation loss) | $3,720,000 | $132.53 |
| Total costs | $6,530,000 | $217 |
| Average number of records= 28070 | | |

Table 6.4: Average cost per record for two record types in Hub International calculator by our analysis

| Partial costs | Avg. cost per record for CC | Avg. cost per record for PHI&SSN |
|---|---|---|
| Incident investigation | $1.15 | $1.64 |
| Crisis management | $3.52 | $4.57 |
| Sanctions | $0.81 | $0.81 |
| Lawsuit | $7.09 | $1.56 |
| Total costs | $12.57 | $8.58 |

There are two apparent sources of discrepancy:

1.What the cost includes: Ponemon costs include intangible costs, such as reputation loss and its impact on business loss. NetDiligence does not include these costs. In addition, NetDiligence uses insurance claims as a measure of the costs. It has been argued that the insurance claims represent only the costs covered by insurance purchased [102]. Still, the insurance coverage should be of the same order of magnitude as the actual cost. The reputation loss cost can be hard to measure. Recent attempts to measure it using stock price as a metric appears to suggest that it may be minimal in many cases, and the effect may be masked by more significant causes in the stock price movement.

2.The sizes of the breaches: The major contributor to the cost per record differences is probably the fact that overall cost is not likely to be proportional to the number of records involved. The average number of records used in the 2015 Ponemon study was 28,070 and not in excess of 100,000 records in general, while the NetDiligence data involves breaches with a much higher number of records, with the average breach involving 3,166,600 records [103]. There are two reasons: some data breach costs are relatively fixed and are independent of the number of records. Thus, the cost per record would be lower with a larger number of records. Even when the cost does go up with the number of records, economics of scale enters the picture. An organization with a large number of records breaches should be able to handle them more cost effectively and should be able to negotiate better rates from organizations providing recovery services. The 2015 Verizon report points out that for 100 lost records, the average cost per record was $254 since the expected breach cost is $25,445, but for the loss of 100 million records, the cost was only $0.09/record since $9 million is the expected breach cost [96]. Thus, cost per record is a misleading metric.

Our analysis of the Hub International calculator, which uses the data breach cost data from NetDiligence, suggests that it assumes a linear trend. On the other hand, the calculators that depend on the Ponemon data, where the number of breaches do not exceed 100,000 records, are not linear [104].

We thus propose to a model that is non-linear relative to the number of records since a linear model implies that there is a cost per record that is meaningful.

### 6.3 Economy of Scale

Since the cost per record is misleading when it comes to estimating the data breach cost, the economy of scale concept will assist in making the cost per record inconstant instead of the constant cost per record that is obtained by dividing the total breach cost by the total breach size. Therefore, using the economy of scale is important in order to remove the correlation between the big breach cost and breach size and ensure that the relationship relies not only on the cost per record factor only but also on other factors.

The current issue is that the total breach cost increases when the size of the data breach increases. To investigate this issue, the concept of economy of scale is needed for analysis, along with actual data. This concept is defined as a decrease in the average long-term costs resulting from an increase in the size of the operating unit [105]. The actual data that identifies that concept is taken from the following reports: Ponemon 2013 [106], Ponemon 2014 [107], and Verizon/NetDiligence [96] (we used a software to digitize a plot in Verizon report). Therefore, we present two hypotheses:

(1) The overall cost rises with breach size.

(2) For larger breaches, the breach cost per record will decline. Thus, the overall breach cost will rise less than linearly.

To test these hypotheses, we plot the data from the three datasets as given by Figures (6.1, 6.2, and 6.3). The two hypotheses are supported by both the Ponemon and the Verizon/NetDiligence data sets. The per-record cost declines as the size of the data breach increases. Table 6.5 gives the results of the regression. A residual analysis suggests that the trend appears to change slightly for record sizes greater than about 25,000. That suggests the model may be amenable to further refinement. Perhaps a piecewise regression may yield better accuracy, which may be addressed in future research. It is notable that while the parameter values for the Ponemon data sets are in the same range, the values are different for the Verizon/NetDiligence data. The reason is that the two data collection approaches are different. The Verizon/NetDiligence data is based on insurance payments, while the Ponemon data includes more complete costs such as opportunity costs.

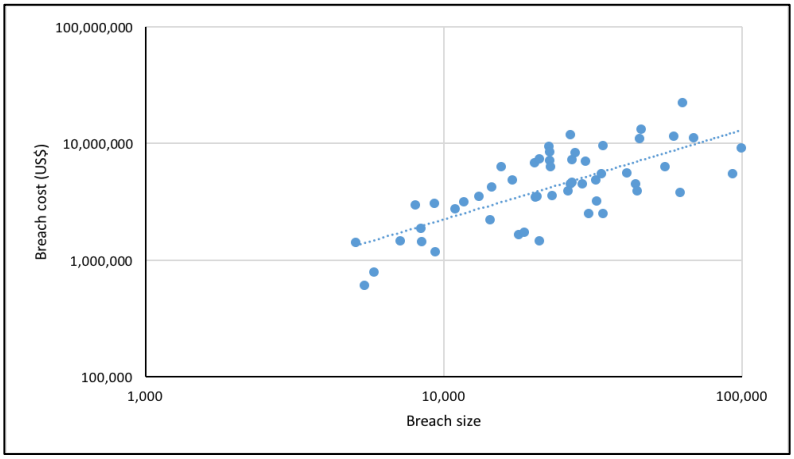


Figure 6.1: Ponemon 2013 data - the breach cost vs breach size (ranges from 5,000 to 100,000 records)

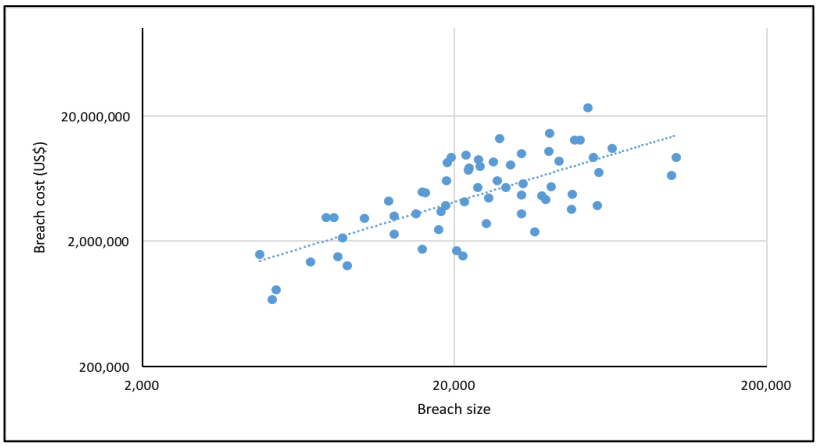


Figure 6.2: Ponemon 2014 data -the breach cost vs breach size (ranges from 4,700 to 103,000 records)

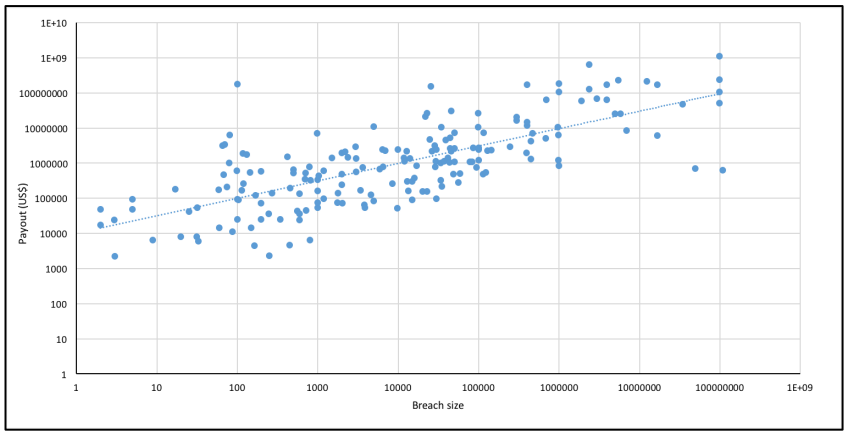


Figure 6.3: Verizon 2015 data -the claim amount vs breach size (ranges from single digits to 108 million records)

Since there are several factors that impact the overall breach cost, it is to be expected that there would be significant variation that is not explained by breach size alone. For very small breaches, the fixed costs would dominate, and thus the trend would not be clearly visible.

Based on the available data sets, a model of the total breach cost after incorporating economy of scale can be formulated as given in equation (6.1):

$$\boldsymbol{Total\ breach\ cost} = a\ *\ size\ \hat{}\ b \qquad (6.1)$$

In this equation, a and b are applicable parameters, and size refers to breach sizes bigger than or equal to 1000 records. (This equation is not applicable in cases where a small number of

records are affected.) Then, the cost per record after incorporating economy of scale is obtained by dividing the previous equation by breach size as in the equation (6.2):

$$\boldsymbol{Cost\ per\ record} = a * (size) \wedge (b - 1) \quad (6.2)$$

Table 6.5 gives the values of the two parameters for the Ponemon 2013, Ponemon 2014, and NetDiligence data. As observed above, the parameter values for the two Ponemon data sets are close, suggesting that the two data sets, while distinct, were collected using the same approach. The NetDiligence data yields somewhat different values, which is likely to be the result of the fact that their numbers were collected differently. However, the two hypotheses mentioned above are both supported by the three data sets. The parameter values in equation (6.2) should conform with how the numbers are to be interpreted.

Table 6.5: The breach cost/payout regression models for the three datasets

| Data sets | Size of breaches | Data points | Regression | |
|---|---|---|---|---|
| | | | Breach Cost Model | $R^2$ |
| Ponemon 2013 | 5000 - 100,000 | 54 | $y = 1924.2x^{0.7662}$ | 0.52 |
| Ponemon 2014 | 4700 - 103,000 | 61 | $y = 2439.9x^{0.7499}$ | 0.50 |
| NetDiligence (Verizon report) | 2 -108 million | 183 | $y = 10002x^{0.4971}$ | 0.54 |

## 6.4 A Comprehensive Cost Computation Model

Figure 6.4 presents a comprehensive model for the cost of a breach and hence the annual security cost. It incorporates the concepts from the existing computational models. Ponemon Institute terms some data breach costs direct, meaning that they require direct financial expenses, while indirect costs include the time, effort, and other. The total cost for a breach includes five partial costs: incident investigation, crisis management, regulatory and industry sanctions related to governmental procedures, class action lawsuit, and opportunity cost.

The total security cost consists of two components: the total cost directly attributable to a data breach and the cost of security maintenance and upgrade, which would be necessary even if the breach had not occurred. The two costs contribute to the expected annual security cost (equation 6.3) for an organization that is affected by a security data breach.

$$\boldsymbol{Expected\ Annual\ Security\ Cost} = \text{Annual expected costs due to breaches} + \text{Costs regardless of any breaches} \quad (6.3)$$

The expected annual cost because of possible data breaches would depend on the likelihood of a breach of a specific type (equation 6.4). Thus,

$$\boldsymbol{Annual\ Expected\ Cost\ due\ to\ Breach} = \Sigma\ \text{Probability of a breach of data type i} \times \text{Total cost per breach for type i} \quad (6.4)$$

The expected costs due to breach could cover the past data breach that occurred or the data breach that could occur in the future. In the past breach, the data breach probability will be 1 and the actual cost of data breach will compute normally. But, for the future data breach, the data breach probability will be less than 1.

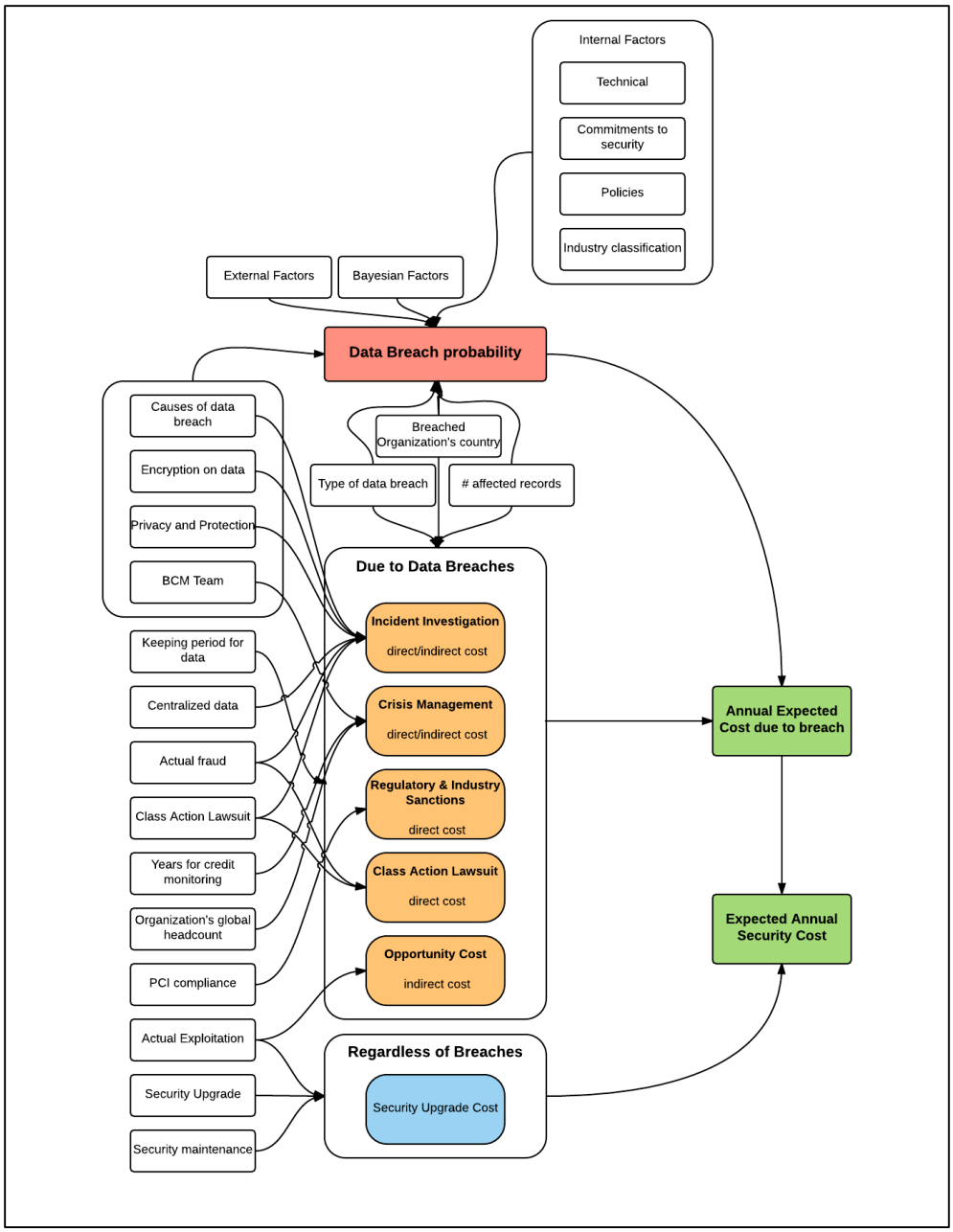


Figure 6.4. Overall risk evaluation model (Data breach cost and probability)

### 6.4.1 Compiled Cost Data

As mentioned in a chapter (5), we analyzed the available calculators that estimate data breach risks. We studied the Hub International Calculator to collect actual data such as a and b parameters by using the power regression equations that computed partial costs, but did not

calculate the probability. For IDT911's calculator, the presented costs for each option have been recorded in the tables below. Those costs are part of the total breach cost, but such a small figure may not be the appropriate method for estimation of breach cost. With the IBM/Ponemon calculator, the details of costs per records and the probability of breach occurrence within next 12 months are sent out via email after filling out a survey. IBM/Ponemon does not publish how it estimates each cost per record for each option that is chosen for the questions (factors). Our model is based on the numbers of the Hub International calculator, since it was obtained by our analysis, and the IBM/Ponemon calculator provides the multiplier factors values that are represent the variation between options. The interface, a, b, cost per record, partial cost per breach, and the probability are shown in Table (6.6, 6.7, 6.8, 6.9, 6.10, and 6.11).

Only two types of data breach are used: Personal Health information (PHI) and Social Security Number (SSN), and Credit Cards (CC).

We ignore three factors during estimation of partial costs per record for several reasons. For instance, factor 3 is ignored, since we think that its data about costs and probability is redundant/overlapped with factor 2, which regards industry classification. In addition, we ignore factor 10 in incident investigation cost classification during the estimation of that cost because the estimation of the breach cost is small and might be insignificant, but we use this factor when we compute the cost per record (CPR). Moreover, factor 12 is redundant, as we believe it is similar to factor 1. Therefore, we ignore it, because its impact of more headcount will equal the impact of more breach size.

Table 6.6: First factor to enter the size of breach that impacts on the data breach cost

| Data Source | Significant Factors | Option |
|---|---|---|
| Hub Int'l, IBM/ Ponemon | (1) Total Number of Affected Records? | User's Input |

Table 6.7: The data breach cost and the probability for the factors of data breach types

| Factors that impact Types of Data Breaches | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data Source | Significant Factors | Options with cost and probability | | | | | | | | | | | |
| IBM/ Ponemon | (2) What is your organization's industry classification? | Communications | Consumer Products | Education | Financial Services | Government Services | Healthcare and Pharmaceuticals | Industrial | Retail | Services: professional and general services | Technology and software | Transportation | All others |
| | cost/record (US$) | 219 | 191 | 184 | 273 | 169 | 289 | 174 | 182 | 243 | 267 | 195 | 217 |
| | Probability % | 11 | 12.50 | 13.10 | 9.90 | 16.50 | 12.70 | 7.80 | 17.10 | 14.90 | 12.70 | 8.70 | 10.10 |
| IBM/ Ponemon | (3) What types of information do your employees handle? | Consumer data | Customer data including credit card information | Customer data excluding credit card information | Employee records | Citizen records | Student information | Patient health data | All other types of data | | | | |
| | cost/record (US$) | 167 | 243 | 213 | 250 | 169 | 195 | 289 | 210 | | | | |
| | Probability % | 11.00 | 12.40 | 12.90 | 9.10 | 15.40 | 11.40 | 16.80 | 9.00 | | | | |

Table 6.8: The parameters of a and b, and the data breach cost and the probability for incident investigation cost

| Incident Investigation Cost | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Data Source | Significant Factors | Options with cost and probability | | | | | | | |
| Hub Int'l | (4) Data is in a centralized system/location? | Yes | Yes | Yes | No | Yes | No | No | No |
| | (5) Actual Fraud is expected already? | Yes | Yes | No | Yes | No | Yes | No | No |
| | (6) Federal Class Action Lawsuit filed? | Yes | No | Yes | Yes | No | No | Yes | No |
| For PHI&SSN | a | 1532.8 | 1473.2 | 1452.4 | 901.37 | 1473.2 | 459.15 | 789.08 | 459.15 |
| | b | 0.57 | 0.50 | 0.56 | 0.64 | 0.50 | 0.66 | 0.65 | 0.66 |
| For CC | a | 1108.1 | 1093.5 | 1054.2 | 650.41 | 1093.5 | 338.74 | 338.74 | 1.4 |
| | b | 0.57 | 0.49 | 0.56 | 0.64 | 0.49 | 0.66 | 0.64 | 0.66 |
| IBM/ Ponemon | (7)What do you think is the most likely cause of a data breach? | Malicious or criminal attack | | Negligence or mistakes (Human error) | | System glitch | | Don't know | |
| | cost/record (US$) | 291 | | 163 | | 169 | | 245 | |
| | Probability % | 16.60 | | 10.30 | | 9.50 | | 12.60 | |
| | (8)Is sensitive data encrypted on all laptops or removable storage? | Yes | | No | | Not sure | | | |
| | cost/record (US$) | 130 | | 267 | | 254 | | | |
| | Probability % | 8.80 | | 14.20 | | 13.80 | | | |
| | (9)What best describes your organization's privacy and data protection program? | A formal privacy and data protection program that is enterprise-wide | | A formal privacy and data protection program that is not enterprise-wide | | An informal privacy and data protection program that is enterprise-wide | | An informal privacy and data protection program that is not enterprise-wide | No privacy or data protection program in place |
| | cost/record (US$) | 156 | | 202 | | 228 | | 241 | 258 |
| | Probability % | 10.10 | | 10.40 | | 11.30 | | 13.50 | 16.00 |
| IDT911 | (10)How long does the business keep/retain sensitive information pertaining to employees, customers and patients? | Less than 3 months | | More than 3 months but less than 3 years | | More than 3 years but less than 5 years | | More than 5 years | Don't know |
| | cost/breach | 250 | | 1000 | | 2000 | | 3000 | 3000 |

Table 6.9: The data breach cost and the probability for crisis management cost

| Crisis Management Cost | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Data Source** | **Significant Factors** | **Options with cost and probability** | | | | | | | |
| Hub Int'l | (11)Number of Years for Credit Monitoring? | 0 | 1 | 2 | 3 | 4 | 5 | 10 | 20 |
| For PHI&SSN | a | 31.25 | 60.71 | 85.19 | 108.71 | 131.87 | 154.86 | 268.94 | 496.06 |
| | b | 0.83 | 0.84 | 0.84 | 0.85 | 0.85 | 0.85 | 0.86 | 0.86 |
| For CC | a | 21.11 | 41.50 | 58.73 | 75.23 | 91.52 | 107.65 | 187.58 | 346.62 |
| | b | 0.84 | 0.84 | 0.85 | 0.85 | 0.85 | 0.85 | 0.86 | 0.86 |
| IBM/ Ponemon | (12) What is the global headcount of your organization? | Fewer than 500 | 501 to 1,000 | 1,001 to 5,000 | 5,001 to 10,000 | 10,001 to 25,000 | 25,001 to 75,000 | More than 75,000 | |
| | cost/record (US$) | 167 | 180 | 230 | 243 | 269 | 224 | 206 | |
| | Probability % | 11.00 | 11.20 | 13.40 | 13.50 | 12.80 | 12.50 | 11.40 | |
| IBM/ Ponemon | (13)Is your organization's business continuity management team involved in the data breach incident response process? | Yes | | No | | Not sure | | | |
| | cost/record (US$) | 184 | | 243 | | 224 | | | |
| | Probability % | 10.50 | | 13.80 | | 12.50 | | | |

Table 6.10: The parameters of a and b for regulatory and industry sanctions cost

| Regulatory and Industry Sanctions Cost | | | |
|---|---|---|---|
| **Data Source** | **Significant Factors** | **Options with cost and probability** | |
| Hub Int'l | (14)Is PCI compliance an issue? | Yes | No |
| For PHI&SSN | a | 19145 | 865754 |
| | b | 0.43 | 0.02 |
| For CC | a | 11308 | 610611 |
| | b | 0.47 | 0.03 |

Table 6.11: The parameters of a and b for class action lawsuit cost

| Class Action Lawsuit Cost | | | | | |
|---|---|---|---|---|---|
| **Data Source** | **Significant Factors** | **Options with cost and probability** | | | |
| Hub Int'l | (15)Actual Fraud is expected already? | Yes | Yes | No | No |
| | (16)Federal Class Action Lawsuit filed? | Yes | No | Yes | No |
| For PHI&SSN | a | 0.36 | 0 | 0.09 | 0 |
| | b | 1.16 | 0 | 1.16 | 0 |
| For CC | a | 5.12 | 6 | 5.68 | 6 |
| | b | 1.04 | 1 | 1.01 | 1 |

Below, the different cost components are described briefly.

### 6.4.2 Security costs due to data breach

There are several partial costs that contribute to the total cost per breach. These partial costs are generally affected by the number of records that might be impacted by the breach as well as the data breach type. The partial costs per record equations are presented after incorporating the economy of scale. Then, we can obtain the partial cost per breach by multiplying the cost per record with breach size.

The partial costs include the following:

#### 6.4.2.1 Incident investigation cost

The cost of incident investigation includes costs of all the activities that assist the organization to discover the data breach [108]. For example, forensic, investigation, and consulting services, assessment and audit services, and technology staff cost. We use the factors' values to note the variation between options in the data of IBM/Ponemon. We discuss more about that in the section on Computation of Factors. The equation that estimates the incident investigation cost is show in (6.5):

***Investigation cost per record*** =

$$[a*(size)^{(b-1)} \text{ for factors } 4,5,6] * F_{breach\_cause} * F_{encryption} * F_{privacy} \quad (6.5)$$

Where the values of a, b, and the cost per record can use them from Table 6.8. In addition, the factors that use in this equation are analysis as the follows:

**Cost factor: Data Breach Causes $F_{breach\_cause}$:**

The causes that lead to data breach have different impacts on the cost of data breach. Some causes have minimal impact on the cost and some have a greater impact. Table 6.12 is represented the values of this factor.

Table 6.12: Cost factor - Data breach causes $F_{breach\_cause}$

| Data breach causes | Multiplier |
|---|---|
| Malicious or criminal attack | 1.19 |
| Negligence or mistakes (Human error) | 0.67 |
| System glitch | 0.69 |
| don't know | 1 (default) |

**Cost factor: Sensitive Data Encryption $F_{\text{encryption}}$:**

The encryption of sensitive data on laptops or removable storage (if applicable) costs the organization less money if the organization has a data breach, but more money if the data is not encrypted. The values for this factor are shown in Table 6.13.

Table 6.13: Cost factor - Sensitive data encryption $F_{\text{encryption}}$

| Encryption of sensitive data | Multiplier |
|---|---|
| Yes | 0.51 |
| No | 1.05 |
| Not sure | 1 (default) |

**Cost factor: Organization's Privacy $F_{\text{privacy}}$:**

The privacy and the protection of data have a huge effect on the data breach cost based on if they are applied or not. The factor values are shown in Table 6.14.

Table 6.14: Cost factor - Organization's Privacy $F_{\text{privacy}}$

| Organization's Privacy | Multiplier |
|---|---|
| A formal privacy and data protection program that is enterprise-wide | 0.68 |
| A formal privacy and data protection program that is not enterprise-wide | 0.89 |
| An informal privacy and data protection program that is enterprise-wide | 1 (default) |
| An informal privacy and data protection program that is not enterprise-wide | 1.06 |
| No privacy or data protection program in place | 1.13 |

### 6.4.2.2 Crisis management cost

Crisis management refers to the activities that allow the company to inform the public that personal information has been lost or stolen [108] and manage the impact of the current data breach. It also includes notification activities, credit monitoring and reissuing credit cards (if any). Note that the cost of reissuing the cards may be borne by the issuing bank and not the

organization experiencing the breach (this may however change because of recent changes in the rules). The equation that estimates the crisis management cost is show in (6.6):

***Crisis Management Cost per Record*** = [a * (size) ^ (b-1) for factor 11] * $F_{BCM}$ (6.6)

Where the values of a, and b can use them from Table 6.9. In addition, the factor values of the team of business continuity management are discussed as the following:

**Cost factor: Business Continuity Management Team $F_{BCM}$:**

This team usually knows to detect the data security risk in the organization and has an emergency plan to deal with a potential breach. Therefore, finding this team in the organization will reduce the data breach cost. Table 6.15 shows the values of these factors.

Table 6.15: Cost factor - Business continuity management team $F_{BCM}$

| BCM involved in incident response plan | Multiplier |
|---|---|
| Yes | 0.82 |
| No | 1.08 |
| Not sure | 1 (default) |

#### 6.4.2.3 Regulatory and industry sanctions cost

This cost depends on the PCI compliance. If the organization is not compliant, fines and sanctions may be imposed on it. The equation that estimates the regulatory and industry sanctions cost is show in (6.7):

***Sanctions cost per record*** = a * (size) ^ (b-1) for factor 14 (6.7)

Where the values of a and b can use them from Table 6.10.

#### 6.4.2.4 Class action lawsuit cost

If a federal class action lawsuit is filed, the organization will incur the costs for litigation, legal defense, and damages. The equation that estimates the class action lawsuit cost is show in (6.8):

***Class Action Lawsuit Cost per record*** = a * (size) ^ (b-1) for factor 15 and 16 (6.8)

Where the values of a and b can use them from Table 6.11.

#### 6.4.2.5 Opportunity cost

These are also known as lost business costs that result from lost business opportunities and reputation after a data breach has been reported to victims and publicly in the media [108]. This can be hard to evaluate because assessing the opportunity loss resulting specifically from the breach can be difficult.

It has been argued that the impact of a breach would be reflected in the stock price of the organization. While some earlier studies suggested that there is in fact a significant impact on stock price, a more recent study has questioned this impact. This could be due to the perception that data breaches are common [109]. The relative impact on the stock price depends on the relation between the total data breach cost and the company's annual revenue. If the breach cost is low, the breach itself will have little impact on stock prices.

### 6.4.3 Security costs Regardless of Data Breach

All organizations take measures to reduce the possibility of data breaches. These depend on the size of the organization and perceived security risks.

These costs include recurring costs of the security measures and security upgrades that may be needed. Upgrading means protecting the organization against the damage or loss of information by bridging existing security gaps. Such security upgrades reduce the likelihood of a data breach and hence the cost of cyber insurance at the same time.

Generally, it is hard to find information on the costs of security upgrades since the upgrades occur internally and organizations do not publish the details. Therefore, creating a preliminary economics model to estimate the security upgrade cost for an organization remains an open problem. Some information on upgrade costs can be found in articles that are occasionally

published. For instance, LinkedIn spent between $2–3 million in 2012 to prevent password theft [110].

For relatively small organization departments, these costs may be difficult to quantify, as each staff member spends time, which may be difficult to measure, participating in activities such as installing security patches, configuring systems and applications to achieve better security, and managing system behavior in response to a security breach.

Equation (6.9) shows how to compute the elements that comprise security upgrade costs.

$$\boldsymbol{Cost\ regardless\ of\ data\ breach} = \text{Security Maintance Cost} + \text{Security Upgrade Cost} + \text{insurance premium} \quad (6.9)$$

### 6.4.4 Cyber Liability Insurance Coverage

Cyber liability insurance is also known as data breach insurance. It provides the required coverage after the data breach occurred and the data loss. This coverage by insurance companies is partial.

In the current days, the demand of cyber insurance has increase since the data breaches number has increased also [108]. Therefore, markets have a lot of cyber insurance companies that coverage the 1st party costs and 3rd party costs of data breach. The cyber insurance is becoming the main element during the cyber risk management especially through the data breach risks. We think that cyber insurance is an essential part of data breach incident response plan that assist to minimize the organization's damage, liability, and performance. We can consider that cyber insurance can coverage any business loss and reduce the impacts of it. The cost of cyber insurance and its coverage vary depending on how the cost is accounted for by the organization's agenda. Sometimes, the insurance cost is considered a security cost, and other times it is simply

considered a cost of doing business. Insurance coverage needs further analysis and investigation in the future.

### 6.4.5 Computation of Factors

To calculate all the related data breach and security costs, and the probability, we use the available data from the IBM/Ponemon calculator that is shown in Tables (6.6, 6.7, 6.8, 6.9, 6.10, and 6.11) in order to make the equations of our model to compute the data breach risk. Therefore, the model of data breach cost per record and the probability will use the method of "multiplicative model" similar to the other quantitative models such as the defect density models by Chulani and Boehm [111] and Malaiya and Denton [112], software cost estimation model by Barry Boehm [113], and MIL-HDBK-217 Chip failure rate model [114]. The model of the cost per record is shown in equation (6.10):

***Cost per record for type i (CPR)=*** $F_{\text{country}}$ * $F_{\text{Industry}}$ * $F_{\text{duration}}$ **(Σ** four partial costs per record: incident investigate, crisis management, regulatory and industry sanctions, class action lawsuit**)** (6.10)

Where the three factors are the country of organization that had a data breach $F_{\text{country}}$, the organization's industry classification $F_{\text{Industry}}$, and the duration that the business keeps the sensitive information of employees, customers and patients $F_{\text{duration}}$. These factors will be multiplied by the cost per record of the four partial costs that include incident investigation, crisis management, regulatory and industry sanctions, and class action lawsuit. Each factor has a default value that is equal to one. The sub-models for each factor are proposed as the following:

**Cost factor: Organization's Country** $F_{\text{country}}$

Table 6.16 presents 11 countries whose organizations have a data breach in one of them. This factor is not determined as a significant factor in our approach, but we believe that the country has a large impact on the costs per record mentioned in IBM/Ponemon 2015 Global

analysis [77]. We use the US as the default. Then, we use weighted cost/factor for the rest of the countries based on the cost per record for the US, which is $217.

Table 6.16: Cost factor - Organization's Country $F_{\text{country}}$

| Country Name | Multiplier |
|---|---|
| USA | 1 (default) |
| Germany | 0.97 |
| Canada | 0.95 |
| France | 0.86 |
| UK | 0.75 |
| Italy | 0.67 |
| Japan | 0.62 |
| Australia | 0.61 |
| Arabian Cluster (Saudi Arabia and United Arab Emirates) | 0.56 |
| Brazil | 0.36 |
| India | 0.26 |

**Cost factor: Organization's Industry Classification $F_{\text{Industry}}$**

This factor takes into account the different types of industry classifications. Some of the classifications have a bigger effect on the cost per record than others. The values of the factor of industry classifications are shown in Table 6.17, and the default value is one.

Table 6.17: Cost factor - Organization's Industry classification $F_{\text{Industry}}$

| Industry classification | Multiplier |
|---|---|
| Communications | 1.01 |
| Consumer Products | 0.88 |
| Education | 0.85 |
| Financial Services | 1.26 |
| Government Services | 0.78 |
| Healthcare and Pharmaceuticals | 1.33 |
| Industrial | 0.80 |
| Retail | 0.84 |
| Services: professional and general services | 1.12 |
| Technology and software | 1.23 |
| Transportation | 0.90 |
| All others | 1 (default) |

**Cost factor: Sensitive Information Keeping $F_{duration}$**

The business keeps some information about their employees, customers, and patients for different lengths of time. These times will cost the organizations depending on the length of the keeping time. We determine how many months approximately are in each duration. The durations are as follows: 3, 12, 48, 72. Then, we make 48 months as the default one. After that, we plot the months with cost, make a trend line, and obtain the equation: y=37 * months + 317. In addition, we normalize that equation by dividing all values by 2000, which is the cost of 48 months (the default). The equation becomes: y= 0.0185 * months + 0.158. The values of a factor are found by dividing the cost of each duration by 2000 that is the cost of 48 months. The values are presented in Table 6.18.

Table 6.18: Cost factor - Sensitive information keeping $F_{duration}$

| Duration | Multiplier |
|---|---|
| 3 months | 0.125 |
| 1 year | 0.5 |
| 4 years | 1 (default) |
| 6 years | 1.5 |

After calculation of the cost per record, the total cost due to breach can be computed by multiplication of the cost per record by the number of affected records as shown in equation (6.11):

***Total cost due to breach for type i*** = Cost per records for type i * Breach size (6.11)

## 6.5 Modeling Data Breach Probability

The likelihood of data breach for an organization can depend on factors that may be termed internal [vulnerabilities (if any) that could lead to a breach and whether they are still available, which would imply that there is insufficient security], external (attacker motivation and

capabilities), or Bayesian (past breaches may imply weaker security unless security is significantly upgraded because of a breach) (Figure 6.4).

It should be noted that some of the factors considered by the Ponemon Institute to impact the cost can be considered as factors impacting the probability, such as BCM team and data encryption.

There are two main factors to predict the data breach probability: number of affected records lost or stolen, and industry classification of organizations that is considered a factor under the classification of data types of breach [77]. From the 2015 Ponemon report [77], we used a software to extract some data. The probability is computed by size of data breach and by country as in Figures (6.5 and 6.6). The expression for the probability of data breach based on the breach size given in the equation (6.12) is based on the data points in Figure (6.5) using a trend line for the data.

**The data probability over a 12 month period (y)** = $\alpha \, e^{-\beta x}$

(6.12)

Where: $\alpha = 0.4405$, $\beta = 4E\text{-}05$, and x the breach size. After studying the probability of data breach, we found that most researchers estimate the probability based on limited methods, such as survey and collecting experts' opinions like the Ponemon estimation, using the dataset of Privacy Rights Clearinghouse (PRC) [79] between 2005 to 2014 to predict the probability of some event for the next year [115]. In addition, there are some researchers who do not explain how to compute the probability when they discuss it.

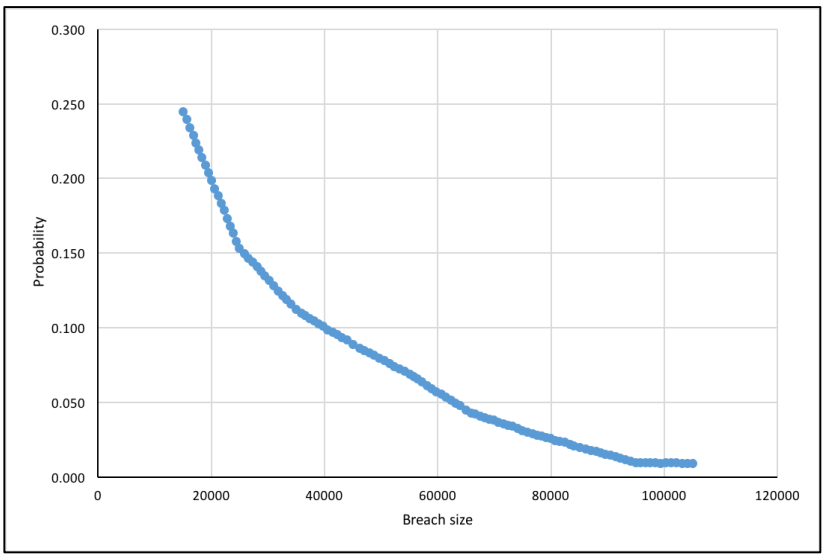


Figure 6.5: Data breach probability based on the breach size (Ponemon data 2015) [77]

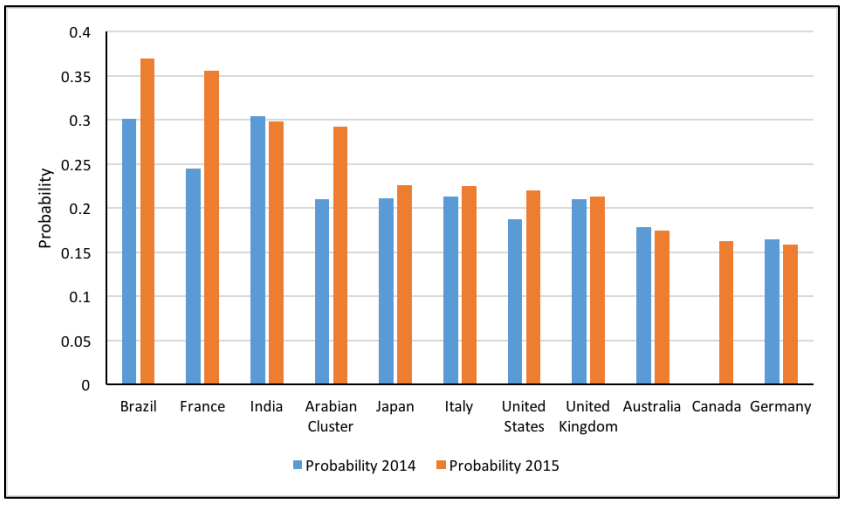


Figure 6.6: Data breach probability by country (Ponemon data 2015) [77]

However, cost and probability of data breach of Ponemon calculators (Symantec and IBM) use a survey (questions or factors) to examine the cost and probability incurred by organizations *after* experiencing data breach incidents. Therefore, we do not expect that the calculators will estimate the cost and probability for the organizations that do not face any data breaches beforehand. Equation (6.13) is represented by the data breach probability for the organization in next 12 months:

***The Probability of a breach of data type i for next 12 months =***

$$F_{country} * F_{BCM} * F_{industry} * F_{breach\_cause} * F_{encryption} * F_{privacy} * \alpha\, e^{-\beta x} \quad (6.13)$$

Where $\alpha$ = 0.4405, $\beta$ = 4E-05, x the breach size, the six factors are the country of organization that had a data breach $F_{country}$, the organization's business continuity management team is involved in the data breach incident response process $F_{BCM}$, the organization's industry classification $F_{Industry}$, and the most likely cause of a data breach $F_{breach\_cause}$, the sensitive data encrypted on al laptops or removable storage $F_{encryption}$, the organization's privacy and availability of data protection program $F_{privacy}$. The factors are multiplied with the probability (equation 6.12). Generally, each factor has a default value that is equal to one at the beginning. The sub-models for each factor have proposed as the following:

**Probability factor: Organization's Country $F_{country}$**

As the data breach cost, the probability is impacted by the different countries of the organizations that have the breach experience. The data of the probability based on the country is taken from the 2015 Ponemon report [77]. We make the US the default. Then, we use a weighted factor for the rest countries based on the probability for US that is 0.22%, as presented in Table 6.19.

Table 6.19: Probability factor - Organization's Country $F_{country}$

| Country Name | Multiplier |
|---|---|
| USA | 1 (default) |
| Germany | 0.72 |
| Canada | 0.74 |
| France | 1.62 |
| UK | 0.97 |
| Italy | 1.02 |
| Japan | 1.03 |
| Australia | 0.79 |
| Arabian Cluster (Saudi Arabia and United Arab Emirates) | 1.33 |
| Brazil | 1.68 |
| India | 1.35 |

**Probability factor: Business Continuity Management Team $F_{BCM}$**

This factor is essential to identify the potential threat that faces the organizations and the impacts that come from these threats. Therefore, this factor has different impacts based on whether or not the team is involved in the data breach incident response plan or not. Table 6.20 shows the values of these factors.

Table 6.20: Probability factor - Business continuity management team $F_{BCM}$

| BCM involved in incident response plan | Multiplier |
|---|---|
| Yes | 0.84 |
| No | 1.1 |
| Not sure | 1 (default) |

**Probability factor: Organization's Industry Classification $F_{Industry}$**

The different industry classifications also making for different probabilities of data breach, as well as the cost of data breach. The factor values are shown in Table 6.21.

Table 6.21: Probability factor - Organization's Industry classification $F_{Industry}$

| Industry classification | Multiplier |
|---|---|
| Communications | 1.09 |
| Consumer Products | 1.24 |
| Education | 1.30 |
| Financial Services | 0.98 |
| Government Services | 1.63 |
| Healthcare and Pharmaceuticals | 1.26 |
| Industrial | 0.77 |
| Retail | 1.69 |
| Services: professional and general services | 1.48 |
| Technology and software | 1.26 |
| Transportation | 0.86 |
| All others | 1 (default) |

**Probability factor: Data Breach Causes $F_{breach\_cause}$**

The probability of data breach varies based on the reason that the data is breached. Table 6.22 shows the values of this factor.

Table 6.22: Probability factor - Data breach causes $F_{\text{breach\_cause}}$

| Data breach causes | Multiplier |
|---|---|
| Malicious or criminal attack | 1.32 |
| Negligence or mistakes (Human error) | 0.82 |
| System glitch | 0.75 |
| don't know | 1 (default) |

**Probability factor: Sensitive Data Encryption $F_{\text{encryption}}$**

If the sensitive data on the laptops or removable storage is encrypted (if applicable), that lowers the probability of data breach compared to data that is not encrypted. The values of this factor are presented in Table 6.23.

Table 6.23: Probability factor - Sensitive data encryption $F_{\text{encryption}}$

| Encryption of sensitive data | Multiplier |
|---|---|
| Yes | 0.64 |
| No | 1.03 |
| Not sure | 1 (default) |

**Probability factor: Organization's Privacy $F_{\text{privacy}}$**

Table 6.24 shows this factor's values. The forms of applied privacy in an organization will impact data breach probabilities. Therefore, if the organization has strict privacy, the probability of data breach is less. Likewise, if the data has perfect protection, the lower the probability for the data to be breached.

Table 6.24: Probability factor - Organization's Privacy $F_{\text{privacy}}$

| Organization's Privacy | Multiplier |
|---|---|
| A formal privacy and data protection program that is enterprise-wide | 0.89 |
| A formal privacy and data protection program that is not enterprise-wide | 0.92 |
| An informal privacy and data protection program that is enterprise-wide | 1 (default) |
| An informal privacy and data protection program that is not enterprise-wide | 1.19 |
| No privacy or data protection program in place | 1.42 |

## 6.6 Challenges and Limitations

Often the companies that have encountered data breaches do not publish details on the real numbers of damage costs, although some numbers do emerge in reports. The Ponemon Institute and NetDiligence collect proprietary data and publish an annual summary. Therefore, our analysis has focused on the reports they have published as well as news reports. The data breach cost estimates—and what those estimates include—vary from one source to another. We have tried to resolve and possibly explain the apparent differences. We develop computational components of our model to ensure that it makes realistic assumptions backed by data from multiple sources.

The data breach cost calculators represent a major first step towards systematic estimation of breach costs. However, they are mainly intended for online estimation for specific cases only and may be intended to promote the security-related services offered. In some cases, we need to obtain the computational results by filling the calculator inputs at various steps, and the calculator then sends the results back by email. The methodology used for computations is not disclosed. Most of the calculators provide the cost per breach only for specifically chosen values for a factor. The values returned are often not broken into cost components for example, the Hub International calculator only generates the cost per breach. Furthermore, some of the calculators do not use any underlying data for costs of security breaches, such as CyberTab, where users need to enter the cost of data breaches themselves to calculate the cost. Thus, some of the calculators provide little insight that would allow the construction of an accurate model for data breach costs.

# Chapter 7

# Conclusion and Future Work

This dissertation has focused mainly on the software vulnerabilities markets, their discoverers, and their potential risks regarding security and economics. We analyze the current types of vulnerability markets, the main players of these markets, and how the vulnerability markets produce huge impacts on security and economics in the same time. In addition, we identify the significant factors that impact the risk of data breaches, and we estimate the cost and probability of a data breach for any organization that has faced a cybercrime. We have faced several challenges and limitations during our research and results, which is natural in many areas, but the important point is how to overcome these challenges to expand our research in the future.

## 7.1 Research Challenges

Studying vulnerability markets is still a new direction in the software vulnerability field. Therefore, there are not many academic references or specific articles in newspapers that concentrate on the vulnerability markets and their transactions. In addition, there is not yet enough data to start the development of key hypotheses regarding the mechanics of multiple vulnerability markets. Additionally, datasets that contain vulnerability market transaction data are not available, missing, or available only to large software companies, not for academic organizations. Specifically, for some vulnerability markets, such as black markets, the transactions between sellers and buyers are unknown.

However, security data breach does not pay attention a lot by the academic researches. Some specific magazines, newspapers' articles, and research labs focus on some aspects of

data breach of organizations or countries. Therefore, the academic references are not available much about the topic of data breach risks. Moreover, the data of security breach is too limited. For instance, most of the companies that have encountered data breaches do not publish details on the real numbers of damage costs, although some numbers do emerge in reports. Our estimation model mainly focusses on different sources of data such as Ponemon Institute and NetDiligence. They collect proprietary data and publish annual reports. Therefore, our analysis accuracy depends on the accuracy of these two reports.

### 7.2 Summary and Conclusions

In our work, we study and analyze several topics related to the software vulnerability markets, whether inside these markets or risks produced from some types of markets if vulnerabilities have been sold and exploited.

We have examined the motivation and methods of vulnerability discoverers by studying the motivation and the methods of discoverers and the vulnerability market. The most successful vulnerability discoverers are identified, and their motivation and techniques have been examined.

While vulnerability discoverers use some tools— including those that they have developed themselves—they rely on their expertise and insight to a considerable extent. It must be kept in mind that tools for finding known vulnerabilities are completely different, and are not of use for discovering new vulnerabilities.

We find that a large fraction of the discoverers is from outside of the software development organizations, and their key motivation is a monetary reward. The vulnerabilities are disclosed in a proper and responsible way when they are traded though the

legitimate markets. Reward programs and contract-based software review services are the major components of the legitimate markets. Organizations that act as vulnerability brokers may deal in either the legitimate or the black market. The vulnerability discoverers acknowledge that the black markets can often be attractive. Our work suggests that government agencies may make up a significant part of the black market buyers.

Furthermore, we have identified multiple vulnerability markets where the exchange between the discoverers and the buyers of the vulnerabilities takes place. We find that legitimate vulnerability markets and illegitimate markets. We believe that all legitimate markets should be regulated to be more attractive for vulnerability discoverers than illegitimate markets since they are providing a good income for discoverers, but produce many risk against the economy and society because the buyers in most cases have special agendas and they are not trusted people or organizations.

However, the potential cost of information loss to businesses and society is increasing yearly. Therefore, there is a need for an economic model to calculate the costs related to data breach incidents with strong predictive capabilities.

We examine the current state of existing approaches, which often disagree in terms of methodologies and results. We collected all calculators' factors that can impact on the cost of data breach or the probability. Then, we choose the significant factors and merge/remove the redundant factors or insignificant based on scientific methodology.

After that, we build an open quantitative model of data breach costs and probability that can be explained and calibrated using existing data as well as newer data as it becomes available. The concept of economy of scale is incorporated in the model in addition to

identification of the fixed and variable costs. The model is validated using actual data. Such a model is allowing more accurate estimation of the economic risks, allowing both software vendors and users to optimally invest in security. The computation of probability, which is provided by some calculators, is explored in some detail.

## 7.3 Future Work

The future work will continue to investigate and study the reasons that make the vulnerability markets are not fully organized, which result in risks that harm economies or societies.

For the vulnerability markets, we should continue to improve our understanding of the factors that affect the vulnerability markets and the main players in them. Moreover, we must organize these markets and encourage the players in them to enter legitimate markets and reduce their use of illegitimate markets. This will assist in minimizing the security and economic risks. Therefore, our next work should be expanded further by looking at a larger number of individual discoverers. There is a need to study the legitimate and the black markets so that the processes can be modeled accurately. Because this is a dynamically changing field, studies such as this need to be repeated in order to see if there are any observable trends in terms of the vulnerabilities that end up in the legitimate and black market periodically, and the subsequent risks to society. Furthermore, our work needs for further examination of the markets in more detail. There is a need to collect data about the transactions in the regulated and the unregulated markets so that the processes can be modeled accurately.

For the risk assessment model of data breach, we need to review the data breach factors that impact the risk year by year and compare our model with the actual data breach cost or probability to ensure that our risk model is improved since there are no perfect/accurate model to estimate the data breach risks, but there is a good or better model when we compare several model in the same criteria.

We believe some issues should be investigated immediately; some ideas should be launched in the near future to mitigate vulnerability risks. For example, we need to create new vulnerability markets that are suitable for the different types of vulnerability discoverers. These markets should be legitimate, attractive, and easy to deal with so they are a good income source for both sellers and buyers. In addition, studying the rewards buyers give to sellers should be reasonable, depending on supply and demand and other commercial market concepts.

# Bibliography

[1] C. P. Pfleeger and S. L Pfleeger. Security in Computing, 3rd ed. Prentice Hall PTR, 2003.

[2] C. Miller, “The legitimate vulnerability market: the secretive world of 0-day exploit sales,” in Workshop on the Economics of Information Security (WEIS), 2007, pp. 7–8.

[3] D. McKinney, “Vulnerability Bazaar,” IEEE Security Privacy, vol. 5, no. 6, pp. 69–73, 2007.

[4] Andy Greenberg, Meet The Hackers Who Sell Spies the Tools to Crack Your PC, Forbes, March 21, 2012, bit.ly/11cbLC6

[5] Sherstobitoff, Ryan. "Anatomy of a data breach." Information Security Journal: A Global Perspective 17.5-6 (2008): 247-252.

[6] Algarni, A. and Malaiya Y, “Most Successful Vulnerability Discoverers: Motivation and Methods”. International Conference on Security and Management, pp. 3-9, 2013.

[7] Algarni, A., and Y. Malaiya. " A Consolidated Approach for Estimation of Data Security Breach Costs." the 2nd International Conference on Information Management (ICIM2016) (2016).

[8] Algarni, A. and Malaiya Y, “Software Vulnerability Markets: Discoverers and Buyers”. International Journal of Computer, Information Science and Engineering, pp. 71-81, 2014.

[9] "What is a Zero-Day Vulnerability?". Pctools by Symantec. [Online]. Available: http://www.pctools.com/security-news/zero-day-vulnerability/.

[10] Böhme, Rainer. "Vulnerability markets." Proc. of 22C3, vol. 27, p. 30, 2005.

[11] Frei, Stefan, Dominik Schatzmann, Bernhard Plattner, and Brian Trammell. "Modeling the security ecosystem-the dynamics of (in) security." In Economics of Information Security and Privacy, pp. 79-106. Springer US, 2010.

[12] Shahzad, Muhammad, Muhammad Zubair Shafiq, and Alex X. Liu. "A large scale exploratory analysis of software vulnerability life cycles." In Software Engineering (ICSE), 2012 34th International Conference on, pp. 771-781. IEEE, 2012.

[13] Allodi, Luca. "The dark side of vulnerability exploitation: a proposal for a research analysis.⋆." In ESSoS Doctoral Symposium. 2012.

[14] Bilge, Leyla, and Tudor Dumitras. "Before we knew it: an empirical study of zero-day attacks in the real world." In Proceedings of the 2012 ACM conference on Computer and communications security, pp. 833-844. ACM, 2012.

[15] Mell, Peter, Karen Scarfone, and Sasha Romanosky. "A complete guide to the common vulnerability scoring system version 2.0." In Published by FIRST-Forum of Incident Response and Security Teams, pp. 1-23. 2007.

[16] Canfield, Casey, Frankie Catota, and Nirajan Rajkarnikar. "A National Cyber Bug Broker: Retrofitting Transparency." (2015).

[17] Finifter, Matthew, Devdatta Akhawe, and David Wagner. "An empirical study of vulnerability rewards programs." In USENIX Security. 2013, 273-288.

[18] Arora, Ashish, Ramayya Krishnan, Rahul Telang, and Yubao Yang. "An empirical analysis of software vendors' patch release behavior: impact of vulnerability disclosure." Information Systems Research 21, no. 1 (2010): 115-132.

[19] Frei, Stefan, Bernhard Tellenbach, and Bernhard Plattner. "0-day patch exposing vendors (in) security performance." BlackHat Europe, Amsterdam, NL (2008).

[20] Arbaugh, William A., William L. Fithen, and John McHugh. "Windows of vulnerability: A case study analysis." Computer 33, no. 12 (2000): 52-59.

[21] Marconato, G. Vache, Mohamed Kaâniche, and Vincent Nicomette. "A Vulnerability Life Cycle-Based Security Modeling and Evaluation Approach." The Computer Journal 56, no. 4 (2013): 422-439.

[22] Frei, Stefan, Martin May, Ulrich Fiedler, and Bernhard Plattner. "Large-scale vulnerability analysis." In Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense, pp. 131-138. ACM, 2006.

[23] Okamura, Hiroyuki, Masataka Tokuzane, and Tadashi Dohi. "Optimal security patch release timing under non-homogeneous vulnerability-discovery processes." In Software Reliability Engineering, 2009. ISSRE'09. 20th International Symposium on, pp. 120-128. IEEE, 2009.

[24] National Vulnerability Database. Available: http://nvd.nist.gov/.

[25] The open Source Vulnerability Database. Available: http://osvdb.org./

[26] Wita, Ratsameetip, Nattanatch Jiamnapanon, and Yunyong Teng-Amnuay. "An Ontology for Vulnerability Lifecycle." In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, pp. 553-557. IEEE, 2010.

[27] Mitre: CVE- Common Vulnerabilities and Exposures. Available: http://cve.mitre.org

[28] Schechter, Stuart. "How to Buy Better Testing Using Competition to Get the Most Security and Robustness for Your Dollar." In Infrastructure Security, pp. 73-87. Springer Berlin Heidelberg, 2002.

[29] Camp, L. Jean, and Catherine Wolfram. "Pricing security." In Economics of information security, pp. 17-34. Springer US, 2004.

[30] Karthik Kannan and Rahul Telang, Market for Software Vulnerabilities? Think Again, Management Science, Vol. 51, No. 5 (May, 2005), pp. 726-740.

[31] S. Ransbotham, S. Mitra, and J. Ramsey, “Are markets for vulnerabilities effective?,” MIS Quarterly-Management Information Systems, vol. 36, no. 1, p. 43, 2012.

[32] A. Ozment, “Bug auctions: Vulnerability markets reconsidered,” in Third Workshop on the Economics of Information Security, 2004.

[33] “WabiSabiLabi may close 0day auction site.” [Online]. Available: http://www.networkworld.com/news/2008/103008-wabisabilabi-may-close-0day-auction.html.

[34] Radianti, Jaziar, and Jose J. Gonzalez. "A preliminary model of the vulnerability black market." In 25th International System Dynamics Conference at Boston, USA. 2007.

[35] Radianti, Jaziar, Jose J. Gonzalez, and Eliot Rich. "A quest for a framework to improve software security: Vulnerability black markets scenario." In Proceedings of the the 27th International Conference of the System Dynamics Society. 2009.

[36] L. Allodi, W. Shim, and F. Massacci, “Quantitative assessment of risk reduction with cybercrime black market monitoring”. The 2013 IEEE Security and Privacy Workshops, pp. 165-172, 2013.

[37] Losses, Net. "Estimating the Global Cost of Cybercrime." McAfee, Centre for Strategic & International Studies (2014).

[38] Layton, Robert, and Paul A. Watters. "A methodology for estimating the tangible cost of data breaches." Journal of Information Security and Applications 19.6 (2014): 321-330.

[39] “Teen Exploits Three Zero-Day Vulns for $60K Win in Google Chrome Hack Contest | Threat Level | Wired.com,” Threat Level. [Online]. Available: http://www.wired.com/threatlevel/2012/03/zero-days-for-chrome/. [Accessed: 06-Oct-2013].

[40] “Bug brokers offering higher bounties.” [Online]. Available: http://www.securityfocus.com/news/11437. [Accessed: 06-Oct-2013].

[41] “Be a Millionaire: The Market for Zero-Day Software Exploits | Critical Start.” [Online]. Available: http://www.criticalstart.com/2012/04/be-a-millionaire-the-market-for-zero-day-software-exploits/. [Accessed: 06-Oct-2013].

[42] “White hat,” Search security. [Online]. Available: http://searchsecurity.techtarget.com/definition/white-hat [Accessed: 06-Oct-2013].

[43] “HacK, CouNterHaCk | New York Times Magazine,” [Online]. Available: http://www.nytimes.com/library/magazine/home/19991003mag-hackers.html. [Accessed: 06-Oct-2013].

[44] Arora, A.; Rahul Telang, "Economics of software vulnerability disclosure," Security & Privacy, IEEE, vol.3, no.1, pp.20, 25, Jan.-Feb. 2005.

[45] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. van Eeten, M. Levi, T. Moore, and S. Savage, “Measuring the cost of cybercrime,” in 11th Workshop on the Economics of Information Security, 2012.

[46] “Shopping For Zero-Days: A Price List For Hackers’ Secret Software Exploits - Forbes,” Forbes. [Online]. Available: http://www.forbes.com/ sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/. [Accessed: 06-Oct-2013].

[47] “Google throws stacks of cash at hackers to publicly crack its Chrome browser,” VentureBeat. [Online]. Available: http://venturebeat.com/ 2012/03/08/hackers-crack-chrome-in-publi/. [Accessed: 06-Oct-2013].

[48] “Cyber-security: The digital arms trade | The Economist.” [Online]. Available: http://www.economist.com/news/business/21574478-market-software-helps-hackers-penetrate-computer-systems-digital-arms-trade. [Accessed: 06-Oct-2013].

[49] Vulnerability Reward Program for Google web properties. [Online]. Available: http://www.google.com/about/appsecurity/reward-program/. [Accessed: 21-Jan-2014].

[50] Chrome Vulnerability Rewards Program. [Online]. Available: http://www.chromium.org/Home/chromium-security/vulnerability-rewards-program. [Accessed: 21-Jan-2014].

[51] The Mozilla Security Bug Bounty Program. [Online]. Available: http://www.mozilla.org/security/bug-bounty.html. [Accessed: 21-Jan-2014].

[52] Facebook rewards program. [Online]. Available: https://www.facebook.com/whitehat/bounty/. [Accessed: 21-Jan-2014].

[53] Wordpress rewards program. [Online]. Available: http://www.whitefirdesign.com/about/wordpress-security-bug-bounty-program.html. [Accessed: 06-Oct-2013].

[54] CCBill Vulnerability Reward Program. [Online]. Available: http://www.ccbill.com/developers/security/vulnerability-reward-program.php. [Accessed: 21-Jan-2014].

[55] Microsoft Bounty Programs. [Online]. Available: http://technet.microsoft.com/en-US/security/dn425036. [Accessed: 21-Jan-2014].

[56] "Microsoft Says No to Paying Bug Bounties," Threatpost. [Online]. Available: http://threatpost.com/microsoft-says-no-paying-bug-bounties-072210/. [Accessed: 06-Oct-2013].

[57] "The Shadowy World Of Selling Software Bugs - And How It Makes Us All Less Safe," ReadWrite. [Online]. Available: http://readwrite.com/ 2012/10/04/the-shadowy-world-of-selling-software-bugs-and-how-it-makes-us-all-less-safe. [Accessed: 06-Oct-2013].

[58] Secunia Vulnerability Coordination Reward Program (SVCRP). [Online]. Available: http://secunia.com/community/research/svcrp/. [Accessed: 21-Jan-2014].

[59] ZDI Rewards Program. [Online]. Available: http://www.zerodayinitiative.com/about/benefits/. [Accessed: 21-Jan-2014].

[60] Ryan Gallagher, "Cyberwar's Gray Market- Should the secretive hacker zero-day exploit market be regulated?" Slate, Jan. 16, 2013.

[61] Michael Riley and Ashlee Vance "Cyber Weapons: The New Arms Race" BloomsbergBusinessWeek, July 20, 2011.

[62] "Schneier on Security: The Vulnerabilities Market and the Future of Security." [Online]. Available: https://www.schneier.com/blog/archives/ 2012/06/the_vulnerabili.html. [Accessed: 06-Oct-2013].

[63] "Stuxnet was work of U.S. and Israeli experts, officials say - The Washington Post." [Online]. Available: http://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6U_story.html. [Accessed: 06-Oct-2013].

[64] "Black hat greed reducing software vulnerability report rate • The Register." [Online]. Available: http://www.theregister.co.uk/2013/02/ 26/grey_market_cuts_vulnerability_reporting/. [Accessed: 06-Oct-2013].

[65] "Welcome to the Malware-Industrial Complex | MIT Technology Review." [Online]. Available: http://www.technologyreview.com/news/ 507971/welcome-to-the-malware-industrial-complex/. [Accessed: 06-Oct-2013].

[66] H. Joh and Y. K. Malaiya, "Defining and Assessing Quantitative Security Risk Measures Using Vulnerability Lifecycle and CVSS Metrics," SAM'11, The 2011 International Conference on Security and Management, pp.10-16, 2011.

[67] "Report: Eastern European Hackers More Sophisticated Than Asian Counterparts". [Online]. Available: http://blogs.wsj.com/digits/2012/ 09/18/report-eastern-european-hackers-more-sophisticated-than-asian-counterparts/. [Accessed: 06-Oct-2013].

[68] "GCHQ Establishes Cyber Unit to Detect Software Vulnerabilities - IBTimes UK." [Online]. Available: http://www.ibtimes.co.uk/articles/ 448951/20130321/gchq-establishes-cyber-research-unit-search-software.htm. [Accessed: 06-Oct-2013].

[69] O. H. Alhazmi and Y. K. Malaiya, "Application of Vulnerability Discovery Models to Major Operating Systems," IEEE Trans. Reliability, March 2008, pp. 14-22

[70] S.-W. Woo, H. Joh, O. H. Alhazmi and Y. K. Malaiya, "Modeling Vulnerability Discovery Process in Apache and IIS HTTP Servers", Computers & Security, January 2011, Pages 50-62.

[71] R, Anderson, University of Cambridge, Home page. [Online]. Available: http://www.cl.cam.ac.uk/~rja14/ [Accessed: 27-Apr-2013].

[72] H.-C. Joh and Y. K. Malaiya, "Seasonal variation in the vulnerability discovery process," in Software Testing Verification and Validation, 2009. ICST'09. International Conference on, 2009, pp. 191–200.

[73] Alhazmi, O.H.; Malaiya, Y.K., "Quantitative vulnerability assessment of systems software," Reliability and Maintainability Symposium, 2005. Proceedings. Annual, vol., no., pp.615, 620, Jan. 24-27, 2005.

[74] Ross Anderson and Tyler Moore, The Economics of Information Security, Science, 27 October 2006: 314 (5799), 610-613.

[75] Ponemon Institute. "Is Your Company Ready for a Big Data Breach?". The Second Annual Study on Data Breach Preparedness. 2014. [Online]. Available: http://www.experian.com/assets/data-breach/brochures/2014-ponemon-2nd-annual-preparedness.pdf.

[76] ITRC Breach Report 2015. [Online]. Available: http://www.idtheftcenter.org/images/breach/ITRCBreachReport2015.pdf.

[77] Ponemon Institute. "2015 Cost of Data Breach Study: Global Analysis". Sponsored by IBM. 2015. [Online]. Available: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=SEW03053WWEN.

[78] Ponemon Institute. “2013 Cost of Data Breach Study: Global Analysis”. Sponsored by Symantec. 2013. [Online]. Available: https://www4.symantec.com/mktginfo/whitepaper/053013_GL_NA_WP_Ponemon-2013-Cost-of-a-Data-Breach-Report_daiNA_cta72382.pdf

[79] A Chronology of Data Breaches, Privacy Rights Clearinghouse. [Online]. Available: http://www.privacyrights.org/data-breach.

[80] Open Security Foundation’s formerly Attrition.org. [Online]. Available: http://datalossdb.org.

[81] VERIS. VERIS Community Database (VCDB). [Online]. Available: http://veriscommunity.net/vcdb.html.

[82] The Web Hacking Incident Database (WHID ). [Online]. Available: http://projects.webappsec.org/w/page/13246927/FrontPage.

[83] Hub International’s data breach cost calculator. [Online]. Available: http://www.hubinternational.com/data-breach-cost-calculator/.

[84] Symantec and Ponemon Institute Data Breach Calculator. [Online]. Available: https://databreachcalculator.com/GetStarted.aspx.

[85] MegaPath Data Breach Risk Calculator. [Online]. Available: http://www.megapath.com/security/ponemon-breach-calculator/.

[86] IDT911’s Data Breach Response Expense Calculator. [Online]. Available: http://idt911.com/expensecalc/start.aspx.

[87] CyberTab - Cyber crime & cyber attack calculator. [Online]. Available: https://cybertab.boozallen.com.

[88] IBM Data Breach Risk Calculator. [Online]. Available: http://www.ibmcostofdatabreach.com.

[89] “Hub International Launches Portal to Help Protect Business Against Cyber Losses” NetDiligence. [Online]. Available: http://www.netdiligence.com/PartnerPR071812.php.

[90] NetDiligence. “NetDiligence: Cyber Liability & Data Breach Insurance Claims 2011”. 2011. [Online]. Available: http://netdiligence.com/files/CyberLiability-0711sh.pdf.

[91] Ponemon Institute. “2011 Cost of Data Breach Study: United States”. Sponsored by Symantec. 2012. Available online on: http://www.ponemon.org/local/upload/file/2011_US_CODB_FINAL_5.pdf.

[92] MegaPath Launches Data Breach Risk Calculator. [Online]. Available: https://www.megapath.com/about/press-releases/megapath-launches-data-breach-risk-calculator/.

[93] IDT911’s History. [Online]. Available: http://idt911.com/company/history.

[94] “CyberTab: Free Tool Estimates Damages from Attacks” Tripwire. 2014. [Online]. Available: http://www.tripwire.com/state-of-security/latest-security-news/cybertab-free-tool-estimates-damages-attacks/.

[95] Watters, Paul A. Cyber Security: Concepts and Cases. CreateSpace Independent Publishing Platform, 2012.

[96] Team, Verizon RISK. "2015 Data Breach Investigations Report." 2015. [Online]. Available: http://www.verizonenterprise.com/DBIR/2015/.

[97] N.E. Weiss and R.S. Miller, "The Target and Other Financial Data Breaches : Frequently Asked Questions," Congressional Rese ch Service, 2015. [Online]. Available: https://fas.org/sgp/crs/misc/R43496.pdf.

[98] Target Reports Fourth Quarter and Full-Year 2014 Earnings. [Online]. Available: http://investors.target.com/phoenix.zhtml?c=65828&p=irol-newsArticle&ID=2019880.

[99] Target Reports Fourth Quarter and Full-Year 2015 Earnings. [Online]. Available: http://investors.target.com/phoenix.zhtml?c=65828&p=irol-newsArticle&ID=2142619.

[100] “Data breaches may cost less than the security to prevent them”. TechRepublic. 2015. [Online]. Available: http://www.techrepublic.com/article/data-breaches-may-cost-less-than-the-security-to-prevent-them/.

[101] “How Much Did The Target, Home Depot Breaches Really Cost”. PYMNTS. 2015. [Online]. Available: http://www.pymnts.com/news/2015/target-home-depot-reveal-full-breach-costs/.

[102] “Why Ponemon Institute’s Cost of Data Breach Methodology Is Sound and Endures”. Ponemon Institue. 2015. [Online]. Avaliable: https://www.ponemon.org/news-2/65.

[103] NetDiligence. “2015 Cyber Claims Study”. 2015. [Online]. Available: http://netdiligence.com/downloads/NetDiligence_2015_Cyber_Claims_Study_093015.pdf.

[104] Jacobs, Jay. "Analyzing ponemon cost of data breach." Data Driven Security 11 (2014). [Online]. Available: http://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/.

[105] “Defination of economy of scale”. Business Dictonary. 2015. [Online]. Available: http://www.businessdictionary.com

[106] Ponemon Institute. “2013 Cost of Data Breach Study: United States”. Sponsored by Symantec. 2013. [Online]. Available: http://www.symantec.com/content/en/us/about/media/pdfs/b-cost-of-a-data-breach-us-report-2013.en-us.pdf

[107] Ponemon Institute. “2014 Cost of Data Breach Study: United States”. Sponsored by Symantec. 2014. [Online]. Available: http://public.dhe.ibm.com/common/ssi/ecm/se/en/sel03017usen/SEL03017USEN.PDF?

[108] Ponemon Institute. “2015 Cost of Data Breach Study: United States”. Sponsored by IBM. 2015. [Online]. Available: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=SEW03055USEN.

[109] “3 reasons why cyberattacks don’t hurt stock prices”. Market Watch. 2015. [Online]. Available: http://www.marketwatch.com/story/3-reasons-why-cyberattacks-dont-hurt-stock-prices-2015-04-03.

[110] “LinkedIn: Breach Cost Up to $1M, Says $2-3 Million in Security Upgrades Coming”. Security Week. 2012. [Online]. Available: http://www.securityweek.com/linkedin-breach-cost-1m-says-2-3-million-security-upgrades-coming.

[111] Chulani, Sunita, and Barry Boehm. Modeling software defect introduction and removal: COQUALMO (COnstructive QUALity MOdel). Technical Report USC-CSE-99-510, University of Southern California, Center for Software Engineering, 1999.

[112] Malaiya, Yashwant K. "Software reliability and security." Encyclopedia of Library and Information Science (2005).

[113] “COnstructive COst Model II (COCOMO® II)”. [Online]. Available: http://csse.usc.edu/csse/research/COCOMOII/cocomo_main.html

[114] “MIL-HDBK-217F(N2) Parts Count Prediction Calculator”. [Online]. Available: http://reliabilityanalyticstoolkit.appspot.com/mil_hdbk_217F_parts_count

[115] Edwards, Benjamin, Steven Hofmeyr, and Stephanie Forrest. "Hype and heavy tails: A closer look at data breaches." WEIS, 2015.