Economics of Cybersecurity

Databreaches

Assignment block 2

Group 5

Hasan Abdullah - 4614097
Seu Man To - 4064976
Elsa Turcios Rodriguez - 4597818
Tim van Rossum - 4246306

I. Introduction

Companies manage more sensitive data nowadays, so ensuring its confidentiality is a concern and in some countries an obligation. Companies are putting in effort to reduce incidents related to data breaches, yet there is an estimation that the cost of data breach related incidents will be $150 million by 2020 (Levick, 2017). A data breach is "A security incident that involves the intentional or unintentional access, disclosure, manipulation or destruction of data; or meets specific definitions of a "Breach" as per state/province or federal laws or active contracts with clients, third parties or partners" (Fowler, 2016).

We have analyzed the data from Privacy Rights Clearinghouse which provide information of data breaches from 2005 to 2017. The data breach is classified based on the type of organization and based on the type of attack. It also provides the amount of breach records affected per breach. We used all records from 2016 as our dataset for this assignment.

Our first analysis of the data sets provided several findings. First, based on the type of organization, we found that 70% of the total data breaches happened in the Healthcare (MED) sector. See figure 1. Second, the data breaches are most often due to either Hacking (HACK) or Unintended Disclosure (DISC). See figure 2.

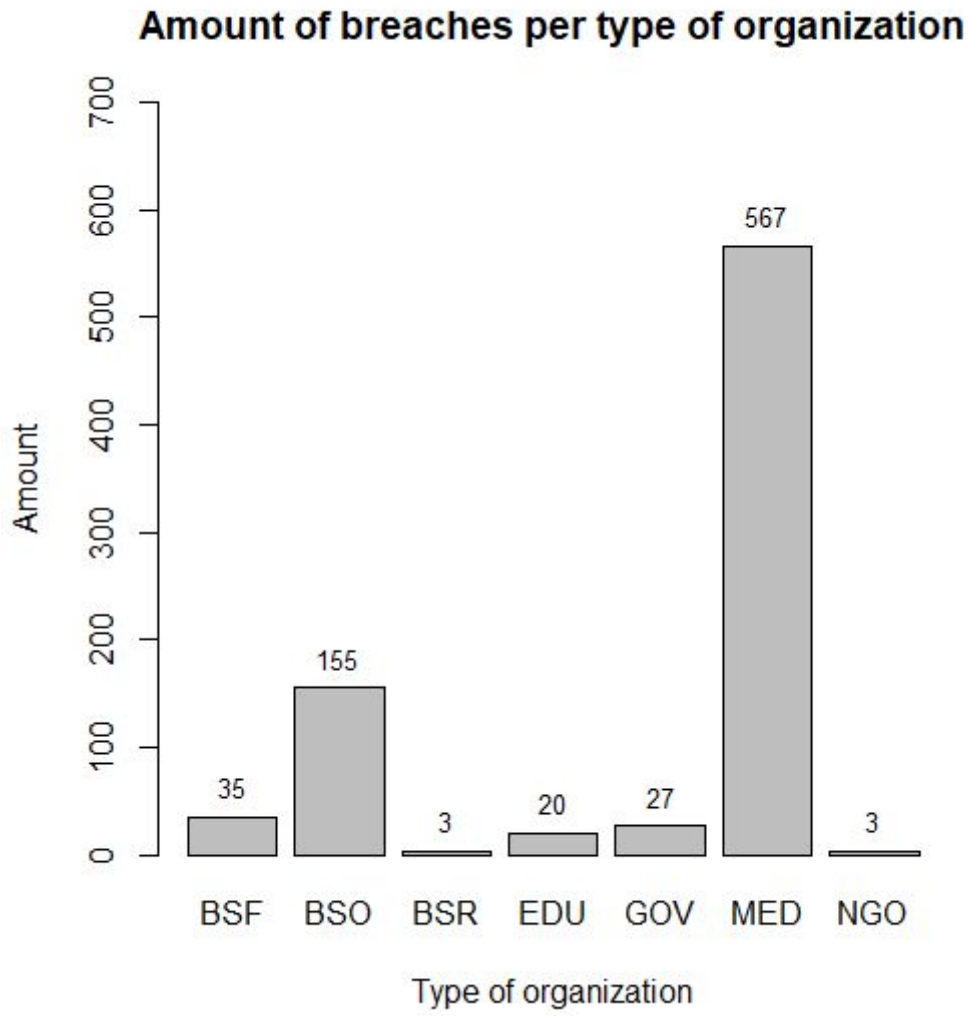

Figure 1. The amount of breaches per type of organization.

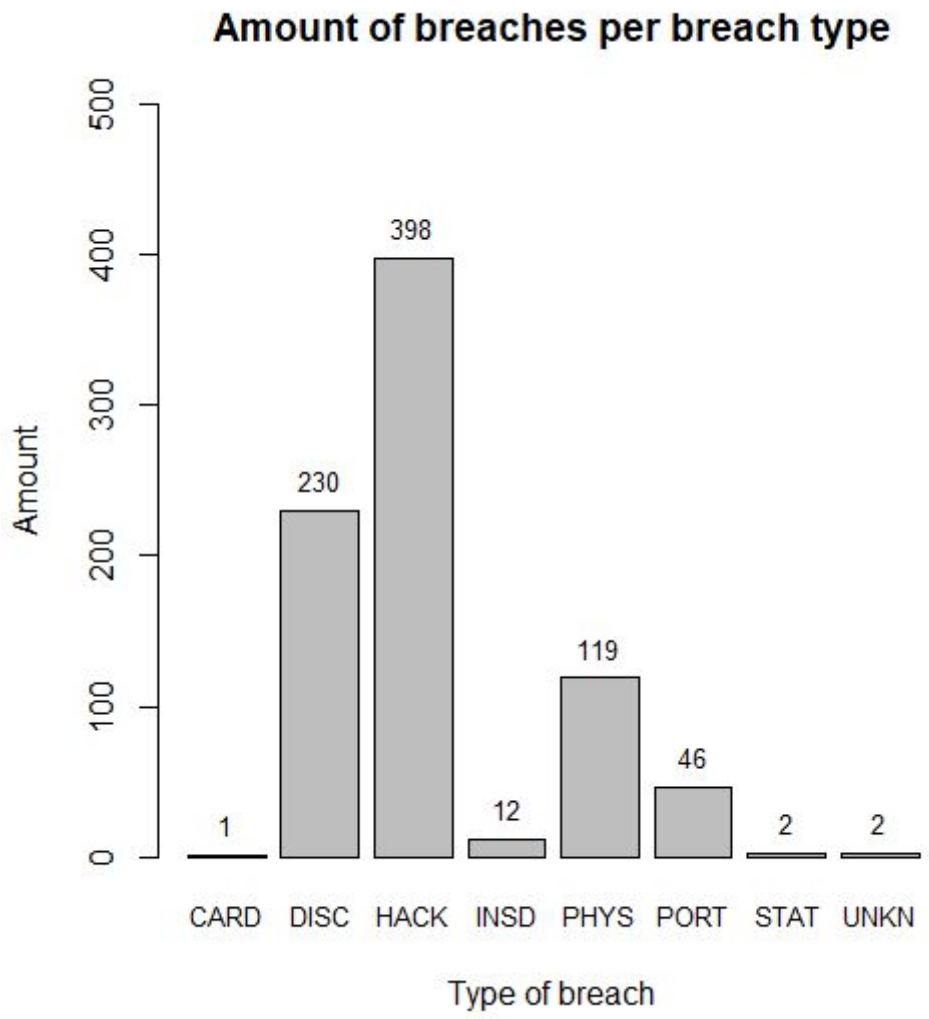

Figure 2. Amount of breaches per breach type.

The data set contains the type of industries affected by data breaches, and the type of breach they each suffered. Therefore, we are aiming to understand if certain industries are more likely to be affected by certain types of data breaches.

The type of breaches in the data set are:

- Payment Card Fraud (CARD)
- Hacking or Malware (HACK)
- Insider (INSD)
- Physical Loss (PHYS)

- Portable Device (PORT)
- Stationary Device (STAT)
- Unintended Disclosure (DISC)
- Unknown (UNKN)

We will classify the type of breaches in two main categories: malicious and negligent (Edwards, Hofmeyr, Forrest, 2015). Unknown data breaches will not be considered in our report. Hack (HACK), Insider (INS), and Payment Card Fraud (CARD) will be categorized as malicious. On the other hand, Unintended Disclosure (DISC), Physical Loss (PHYS), Portable Device (PORT), Stationary Device (STAT) will be categorized as negligent data breaches. There is also a type of data breach classified as Unknown, and we will not consider this category in our analysis.

The industries given in the data are:

- Businesses-Financial and Insurance Services (BSF)
- Businesses-Other (BSO)
- Businesses-Retail/Merchant - Including Online Retail (BSR)
- Educational Institutions (EDU)
- Government and Military (GOV)
- Healthcare, Medical Providers and Medical Insurance Services (MED)
- Nonprofits (NGO)

In this report, we will be focusing our analysis from the perspective of the industries (or the "defender side") as our main actor. The asset we want to protect is the data records that are sensitive and that can cause losses in the bottom line of the companies, because customers can demand reimbursement for damage, fines, or recovery costs can be necessary.

There are many security issues that could cause a data breach incident: incompetence of the employees in an IT environment, the lack of security regarding the infrastructure of the company, or the fact that some records are simply more valuable on the black market than others and thus are more prone to being attacked. Nevertheless, the selected security issue for this report will be

the two main categories of attack that we previously defined, which are malicious and negligent data exposure

II. Ideal Metrics

According to Aghari, van Eten & Bauer (2016) getting trustable metrics is not an easy task, but this is a way of producing markets with a more efficient security. The ideal security metrics should include security cost, security levels and security benefits. However, these are very general metrics, so ideally metrics that are more specifically used for data breaches sector should be used.

A security cost metric specifically intended for data breaches was proposed by Algarni and Malaiya (2016). The paper proposes that the total cost of a data breach incident is the sum of the incident investigation cost, customer notification or crisis management cost, regulatory and industry sanctions cost, and the class action lawsuit cost. The paper also points out that a more used measure is cost per record, which is the total breach cost divided by the amount of breached records.

Another ideal metric would be the costs needed for security to prevent and recover from data breaches per industry. This includes the following:

- direct costs like purchasing, installing and administering security measures, and hiring security staff.
- indirect costs like the time to recover from a data breach per industry.

It would also be ideal if we can measure the details of what the cybersecurity spendings are to avoid data breaches. Instead of cybersecurity spending in general, it will be more complete if we have the detail on the specific spending. For example, how an industry spends in employee's cybersecurity or ethical training, network security, anti virus, et cetera to prevent the different types of data breaches.

Additionally, other ideal metrics would be:

- Type of breaches (malicious or negligent) versus losses.
- Prevented number of specific type of data breaches (malicious or negligent) versus prevented losses.

III. Metrics in practice

On the site of the VERIS Community Database VCDB (VCDB, 2017), a couple of metrics are used. In fact, VERIS is a set of metrics that looks at the following:

- Incident counts by year
- Incident counts by actor (external, internal, partner, unknown)
- Incident counts by attribute (confidentiality, integrity, availability)
- Incident counts by action (type of breach, e.g. malware, hacking etc.)
- Incident counts by asset (e.g. server, network, person etc.)
- Incidents by industry (e.g. healthcare, real estate, finance etc.)

Metrics that can be used to recover from a cyber event include cost, time, damage assessment, and number of incidents (NIST, 2016). These can be used to assess the damage and cost of an incident.

The Center for Internet Security (CIS, 2010) has defined twenty-eight metrics for seven business functions. Some that can be used are cost of incidents, number of incidents, mean-time to mitigate vulnerabilities, and mean cost to mitigate vulnerabilities.

IV. Metrics defined for this dataset

The main metrics that can be designed using the current dataset are:

- Amount of breaches (malicious or negligent) per type of breach
- Amount of breaches (malicious or negligent) per type of industry
- Breached records (size) per type of breach(malicious or negligent)

V. Graphical representations of the metrics

Amount of breaches per type of breach

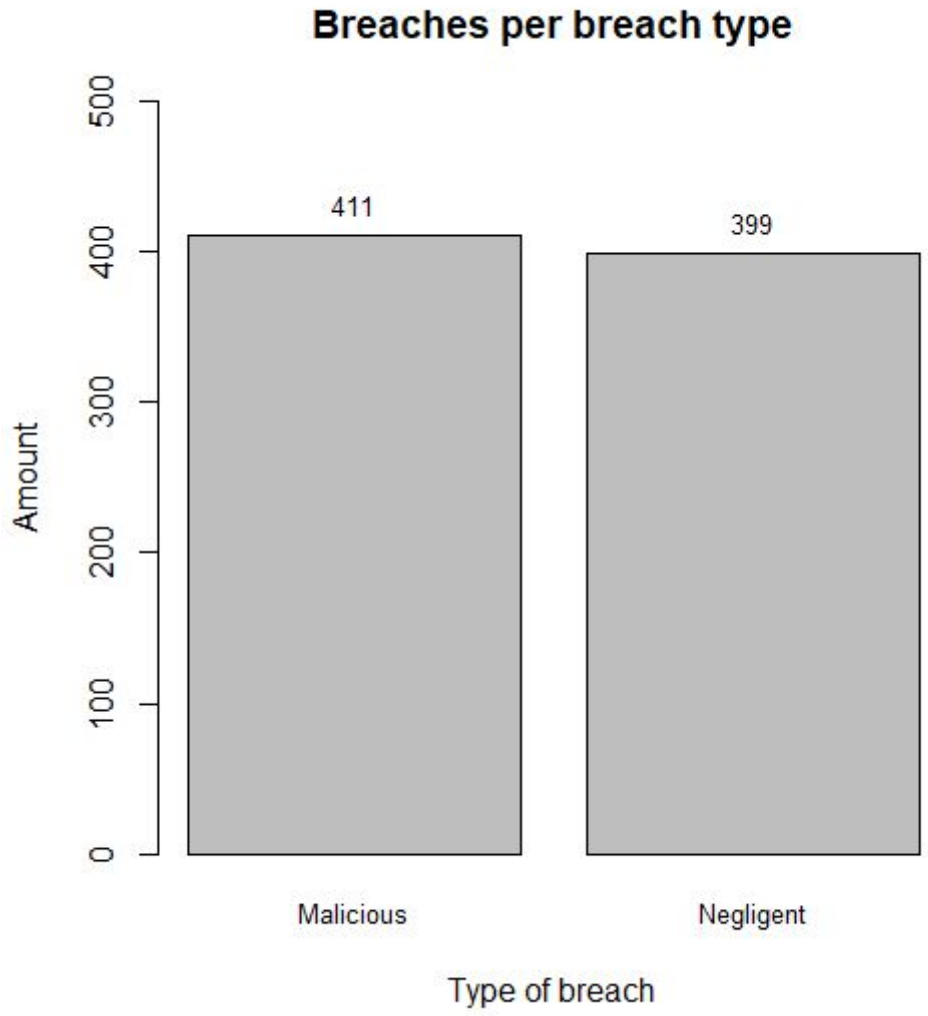

Figure 3. The amount of breaches per type of breach

Amount of breaches per type of industry

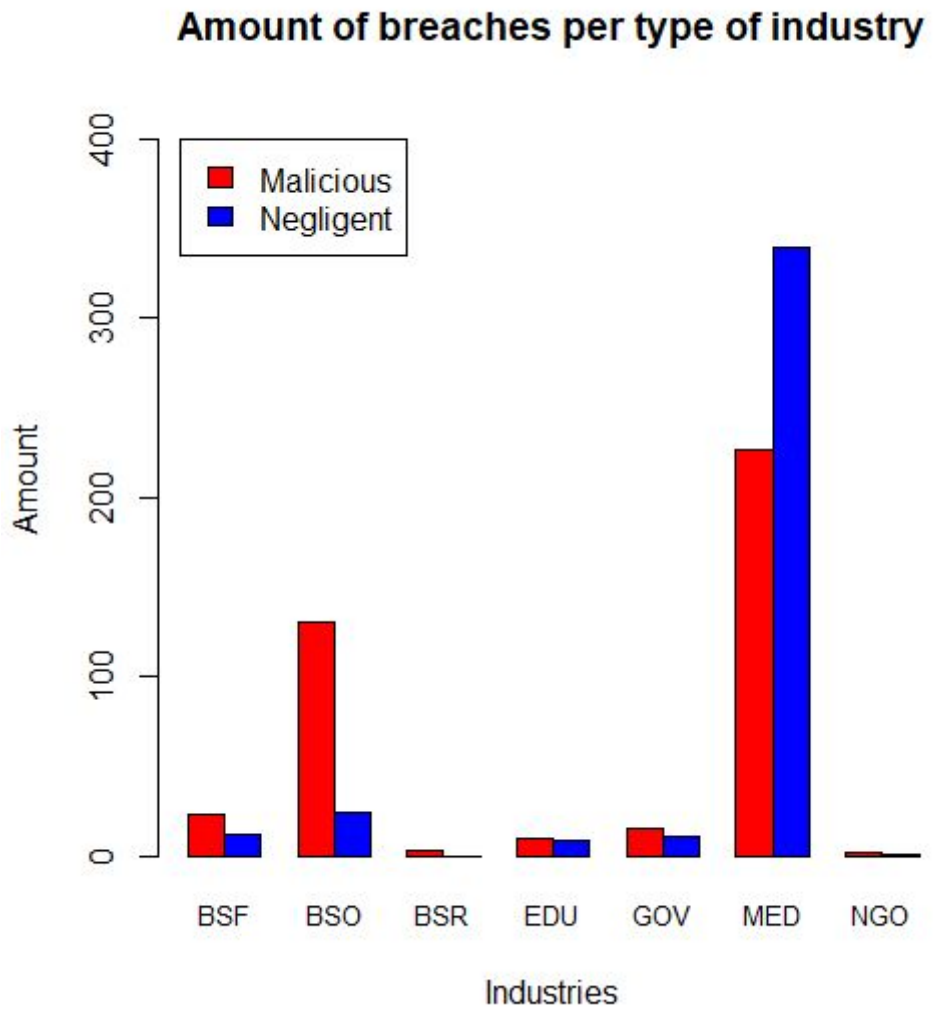

Figure 4. The amount of breaches per type of industry

	BSF	BSO	BSR	EDU	GOV	MED	NGO
CARD	0	1	0	0	0	0	0
DISC	5	14	0	6	9	195	1
HACK	20	130	3	10	12	221	2
INSD	3	0	0	0	3	6	0
PHYS	2	4	0	1	0	112	0
PORT	5	5	0	2	2	32	0
STAT	0	1	0	0	0	1	0
UNKN	0	0	0	1	1	0	0
TOTAL	**35**	**155**	**3**	**20**	**27**	**567**	**3**
Malicious	**23**	**131**	**3**	**10**	**15**	**227**	**2**
Negligent	**12**	**24**	**0**	**9**	**11**	**340**	**1**

Table 1. The amount of breaches per type of industry and their totals

Breached records (size) per type of breach

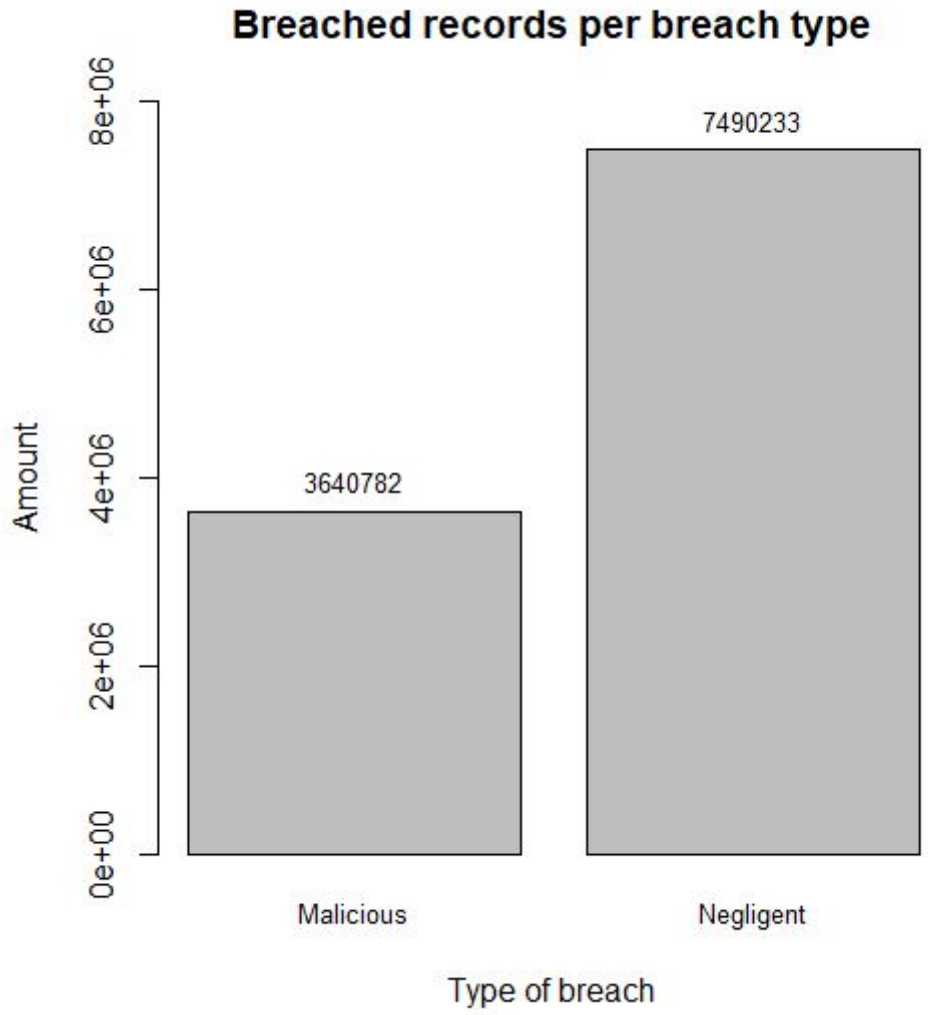

Figure 5. The number of breached records per breach type

	CARD	DISC	HACK	INSD	PHYS	PORT	STAT	UNKN
Unknown	1	224	358	8	116	43	2	1

Table 2. The amount of breaches with unknown breached records.

Table 2 lists the amount of breaches where the records breached are unknown. Therefore, these unknown amount of records are not counted in the barplot.

VI. Conclusion and evaluation of metrics

The metrics that we defined for this dataset give an idea on what methods are used to breach data (whether data was breached actively or due to negligence). The final histogram shows that, even though hacks are the most common data breach cause, the most records are actually breached through negligent methods, so not through hacks. Further, we could observe the Medical Industry is more prone to negligent data exposure while the rest of the industries were more affected by malicious attacks.

However, since the dataset is really limited (it only includes data breaches from the US, and even then does not include data like costs for the company because of the breached data), we are limited to using very simple metrics. perhaps in future assignments we can use more complex metrics if we can use multiple data sources.

References

Asghari, Hadi, Michel van Eeten, Johannes M. Bauer. "Economics of Cybersecurity", In Handbook on the Economics of the Internet (2016), Edgar Elgar, pp. 262–287, 2016.

Böhme, R. (2010, November), "Security Metrics and Security Investment Models". In *IWSEC* (pp. 10-24).

Edwards B., Hofmeyr S., Forrest S. 2015. Retrieved from http://www.econinfosec.org/archive/weis2015/papers/WEIS_2015_edwards.pdf

CIS (Center for Internet Security) (2010), "The Center for Internet Security. The CIS Security Metrics".

Fowler, K. (2016). Chapter 1 - An Overview of Data Breaches. En Data Breach Preparation and Response (pp. 1-26). Boston: Syngress. https://doi.org/10.1016/B978-0-12-803451-4.00001-0

LevickChange In Corporate Mindset Needed To Combat Cyber Attacks. Retrieved September 22, 2017, from https://www.forbes.com/sites/richardlevick/2017/02/13/change-in-corporate-mindset-needed-to-combat-cybersecurity/\

NIST (National Institute of Standards and Technology) (2016). Retrieved from http://csrc.nist.gov/publications/drafts/800-184/sp800_184_draft.pdf

VCDB (2017). Retrieved from http://vcdb.org/explore.html