Treating Personal Data Like Digital Pollution

Ivan Burke[1, 2] and Renier Pelser van Heerden[1, 3]
[1]CSIR, Pretoria, South Africa
[2]Rhodes University, Grahamstown, South Africa
[3]Nelson Mandela Metropolitan University, South Africa
iburke@csir.co.za

Abstract: During the past 10 years data breaches have become more frequent than ever. Large volumes of personal and corporate data are being leaked via these breaches. The corporate responses to these breaches, as well as, the remediation of these breaches are often not sufficient. Similarly to how production companies should be taken to task for polluting the physical environment due to bad business practises, IT institutions should be made aware of their contribution to Cyber pollution. In our article we define the concept of cyber pollution as unmaintained or obsolete devices connect to the internet and corporate networks. This paper breaks down the current state of data breach disclosures within the Europe by providing statistics on large scale data breach disclosures from 2013 till 2016. This paper attempts to model the increase of threat exposure over time similar to that of pollution breaches within the physical environment. Over time small openings or vulnerabilities within systems can lead to exploitation of whole systems. By modelling these breaches as pollution we aim to make the concept of cyber pollution a more tangible concept for IT managers to relay to staff and upper management. The model is validated using anonymised corporate network traffic and Open Source penetration testing software.

Keywords: cyber maintenance, governance, responsibility

1. Introduction

In the digital age data is everywhere and is constantly being generated by everyday tasks. The safeguarding of these data items is becoming increasingly more difficult. In the case of Internet of Things (IoT) and Machine to Machine (M2M) communications, humans are often completely absent from the data creation process. Bruce Schneier warns in his book, Data and Goliath, that: “Data is the pollution problem of the information age, and protecting privacy is the environmental challenge" (Schneier, 2015). This echoes the words of Corry Doctorow, "Every gram - sorry, byte - of personal information these feckless data-packrats collect onus should be as carefully accounted for as our weapons-grade radioisotopes, because once the seals have cracked, there is no going back” (Doctorow, 2011). This brings to mind the question, if the loss of data is truly this dangerous, why is the regulation to prevent data loss not as strict as that of pollution prevention.

The term data breach commonly refers to a security incident by which sensitive data becomes exposed to individuals which are not authorized to access the data. Data breaches can usually be categorised into one of three main types of data being leaked: Personal Health Information (PHI), Personally Identifiable Information (PII), or Intellectual Property (IP). A data breach is exactly the type of loss of control that Schneier and Doctorow were concerned about. Unlike physical pollution, digital pollution often leaves no trace and can go undetected for many years. The European Union has taken the first step towards raising public awareness by introducing a Data Breach Notification Policy into legislation. In this paper the current state of data breaches will be reviewed in Section 2, the review will focus on data breaches within Europe. In Section 3, current legislation regarding data breach disclosure and prevention will be discussed. In section 4, a basic experimental scenario will be defined on how to potentially model data breach detection within an organisation. In Section 5, the results of the experimental model will be discussed. In Section 6, a conclusion will be provided with recommendations for future expansion of the work.

2. Current state of data breaches

In the past few years large volumes of records have been leaked due to data breaches. Table 1 contains a list of processed data from www.breachlevelindex.com. Since the web service started tracking data breaches in 2013, approximately six billion records have been leaked worldwide. The data has been filtered to only include European countries and excluded data breaches where the data breach only consisted of publically available data. According to the Breach Level Index report (2017) approximately 52 percent of breaches that were reported have no data on the amount of records leaked. The web service only reports on publicised data breaches, any non-disclosed or unknown/unconfirmed breaches have been omitted. Of the reported breaches, only 4 percent of the stolen records had been encrypted and unusable by the perpetrator of the breach.

According to Table 1, Turkey has the highest volume of data leaks, with 132 312 866 records lost since 2013. The three largest reported breaches were General Directorate of Population and Citizenship Affairs the General Directorate of Land Registry and Cadaster (Greenberg, 2016), Country's Supreme Election Committee (YSK) (Daily News, Istanbul, 2013), and the Turkish State Hospitals Data breaches (Murdock, 2016). The majority of the attacks against Turkey according to the Breach Level Index were perpetrated by a malicious outsider against Government institutions.

Table 1: Records lost due to data breaches within Europe from 2013 till 2016 (Breach Level Index, 2017)

Armenia	64	**Malta**	4 000
Austria	21 000	**Moldova**	140 000
Azerbaijan	51 850	**Netherlands**	1 332 829
Belgium	92 731	**Norway**	81 000
Brussels	650	**Poland**	41
Bulgaria	2 832 312	**Romania**	27 000
Czech Republic	1 504 000	**Russia**	56 737 681
Denmark	6 203 195	**Serbia**	7 277 452
Estonia	13 486	**Slovakia**	157 700
Finland	3 622	**Spain**	11 154 702
France	5 879 535	**Sweden**	6 090 601
Germany	44 143 054	**Switzerland**	110 257
Iceland	77 000	**Turkey**	132 312 866
Ireland	350 324	**Ukraine**	7 000 000
Italy	440 100	**United Kingdom**	74 697 683
Kazakhstan	62 235		
Grand Total Number of Records Lost			358 798 970

Figure 1 depicts a stacked bar graph which group the data breach incidences by industry affected by the breach. The values have been weighted by multiplying the number of records lost by the risk score associated by the breach, as was calculated by the Breach Level Index Report (2017). This weighting was applied since not all data breach are equally damaging, for instance United Kingdom (UK) has a total of 106 recorded breach incidences related to Governmental data whereas Turkey only has fourteen breaches that were reported. However the Turkish breach in 2016 leaked fifty million, highly confidential, sensitive voter records whereas the majority of UK breaches contained less than five thousand records relating to city council members. What is however concerning is that the UK has the highest volume of Accidental Data Breaches. 198 of the reported 471 breach incidences (42%) were due to accidental data loss.

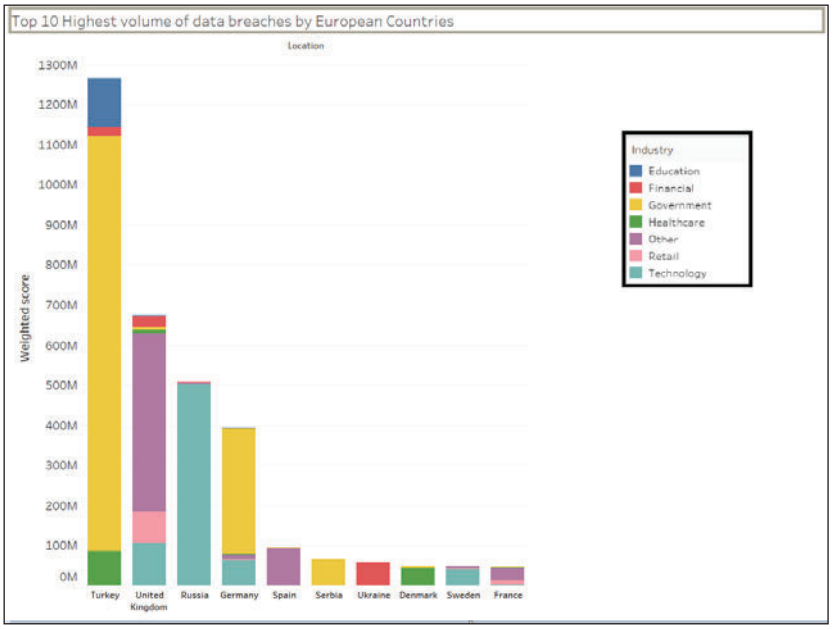


Figure 1: Top 10 European countries hit by data breaches grouped by industry (weighted by risk score)

A significant contributing factor to the large number of data breach incidences was the prevalence of unmaintained and unpatched devices with direct access to the internet. Even well-known and publicised vulnerabilities took a significant time to be patched. To illustrate this point the online resource Shodan (www.shodan.io/) was used to collect statistics on the Heartbleed vulnerability (CVE-2014-0160).

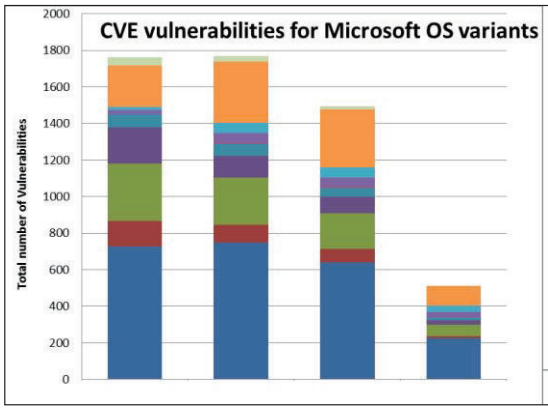
Shodan is an online search engine which indexes global vulnerability scan information. The Heartbleed vulnerability is a serious exploit which targets the OpenSSL cryptographic library. By exploiting this vulnerability attackers can read memory segments of vulnerable servers. This may result in sensitive data being breached. The vulnerability was disclosed 21 March 2014 and was patched officially by 9 April 2014 (Grubb, 2014). However, even through the vulnerability was widely publicised, due to the potential for unwanted data disclosure, there are still a number of online devices vulnerable to the Heartbleed vulnerability. Table 2 shows the current state of vulnerable devices scanned by Shodan. Only the top 10 vulnerable countries are listed within the table. Two scan reports were used for comparison. The first was a scan conducted on 26 March 2016 (Shodan, 2016) and the second a scan conducted 10 February 2017 (Shodan, 2017). Most countries have shown a significant improvement over the past year with regards to reducing the number of vulnerable devices. On average the number of vulnerable devices has decrease by 34%. However the Republic of Korea seems to have had a 44% increase in vulnerable devices.

Table 2: Internet facing devices vulnerable to Heartbleed

	Measurement 26/03/2016		Measurement 10/02/2017		Improvement
1	United States of America	57598	United States of America	32731	43%
2	China	17455	Republic of Korea	10842	-44%
3	Germany	17273	Germany	10521	39%
4	France	10708	China	10084	42%
5	India	9427	France	6775	37%
6	United Kingdom	9268	Russian Federation	5273	43%
7	Russian Federation	7897	United Kingdom	5150	44%
8	Republic of Korea	7525	India	4159	56%
9	Brazil	7095	Brazil	3790	47%
10	Japan	5302	Japan	3690	30%

The example shown in **Table 2** was for a well-publicised and highly dangerous vulnerability; however there are large numbers of internet facing devices that are still running outdated software which have long since stopped having new patches released. According to a Shodan report there are still approximately 14 000 web servers running Microsoft's Internet Information Services module 2.0 (IIS 2.0) (Shodan, 2017). This service has been unsupported by Microsoft since 1997, hence no security patches have been released by Microsoft and these servers have gone without update for nearly twenty years. Though it is doubtful that these devices are still actively used, they could provide a pivoting point for an attacker into a corporate network. Mitre maintains a list the standard for Common Vulnerabilities and Exposures (CVE).

Table 3: CVE vulnerabilities for Microsoft OS variants



	Windows XP	Vista	Windows 7	Windows 10
# of exploits	44	33	17	0
File Inclusion	0	0	0	0
CSRF	0	0	0	0
Gain Privileges	227	333	319	108
Gain Information	16	57	57	35
Bypass something	22	54	52	29
Http Response Splitting	0	0	0	0
Directory Traversal	1	2	2	0
XSS	6	4	3	0
Sql Injection	0	0	0	0
Memory Corruption	67	63	47	13
Overflow	198	120	91	26
Code Execution	312	258	192	64
DoS	142	95	74	10
# of Vulnerabilities	726	750	641	225

Table 3 is a summary of all CVEs reported by Mitre since the release of each Operating System (OS) released by Microsoft since the year 2000. It should be noted that Mitre stopped tracking vulnerabilities for Windows XP

since 2014 due to the OS no longer being maintained by Microsoft. As is shown in Table 3, vulnerabilities tend to increase as technologies age. This shows that aged unmaintained devices on a network can increase the attack service of an organisation and increase the risk for data breaches.

The Ponemon institute conducted a global study on the data breaches since 2013 (Ponemon Institute LLC, 2016). The study included 383 companies from twelve countries. The report's key findings were as follows:

- Since first conducting this research, the cost of a data breach has not fluctuated significantly. This suggests that it is a permanent cost organizations need to be prepared to deal with and incorporate in their data protection strategies.
- Loss of business is the highest impact of data breaches to an organisation. Organisation need to regain the trust lost due to a data breach to recover clientele.
- The highest cause of data breaches is malicious outsiders and due to the delay in detection these breaches often lead to significant financial loss.
- Investments are being made in technologies and in-house expertise to reduce the time to detect and contain.
- Regulated industries, such as healthcare and financial services, have the most costly data breaches because of fines and the higher than average rate of lost business and customers.
- Awareness training and employing dedicated security personnel has led to a reduction in breaches and in costs associated with data breaches.
- The study revealed that there is a reduction in the cost when companies participated in threat sharing and deployed data loss prevention technologies.

Based on the results of the Ponemon report the costs of data breaches can be reduced by raising organisational awareness, sharing threat intelligence and providing in-house technologies to pro-actively.

Research has recently been conducted into achieving pro-active data breach detection (Botha, et al., 2016). The researchers' work focussed on the PII of South African citizens. South Africa recently signed into law the Protection of Legal Information (PoPI) Act (Government Gazette, 2013), which is similar to the European Data Breach Notification legislation. The authors used advance search heuristics to locate personal information of South African citizens using regular expressions and data carving. After compiling and indexing the results, the researchers constructed a geo-location tool which shows the location of servers leaking PII of South African citizens. Figure 2 shows a sample of the results produced by the tool. Though the results might have contained a number of false positives the concept was sound and helped identify accidentally disclosed PII. This work can be extended to citizens outside of South Africa be extending the heuristics and search criteria.

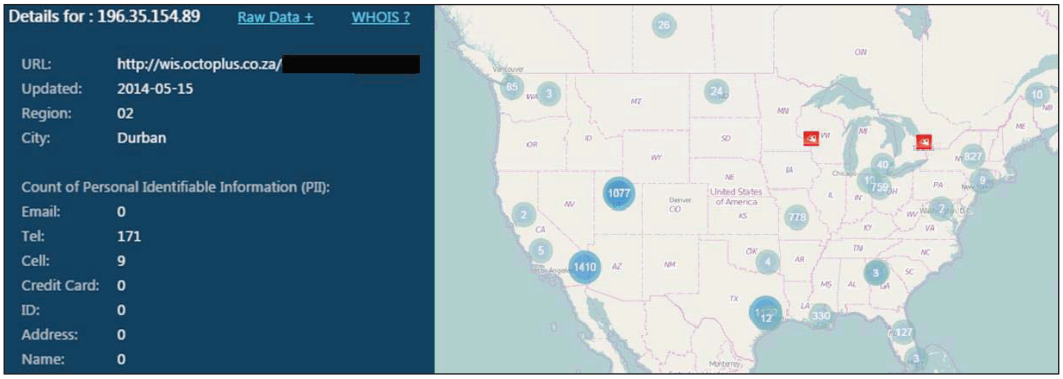


Figure 2: Sample of PII detected by cyber protect

3. Current legislation

The European Network and Information Security Agency (ENISA) introduced the European data breach notification requirement for the electronic communication sector in e-Privacy Directive (2002/58/EC, 2002). A consistent set of guidelines addressing the technical implementation measures and the procedures for Data Breach notification was provided within Article 4 of the Directive.

In November 2009, the European Parliament and the Council of Ministers reached an agreement on European Union Telecoms Reform, after negotiations brokered by the European Commission. The reform, proposed by the Commission in November 2007, aimed to strengthen competition and consumer rights on Europe's telecoms markets, facilitate high-speed internet broadband connections to Europeans and established a European Body of Telecoms Regulators to complete the single market for telecoms networks and services.

Data protection authorities (DPAs) take varied approaches to enforcing data protection and privacy. Some follow European Commission Directives closely, while others take on additional responsibilities beyond those outlined in the Directives (ENISA, 2010).

The Dutch DPA set forth a guide to assist their citizen in gaining a greater understanding of the in e-Privacy Directive (2002/58/EC). Dutch Data Protection Authority (2015) describes the organisational and indifidual obligations as set forth by Article 34a of the Dutch Data Protection Act.

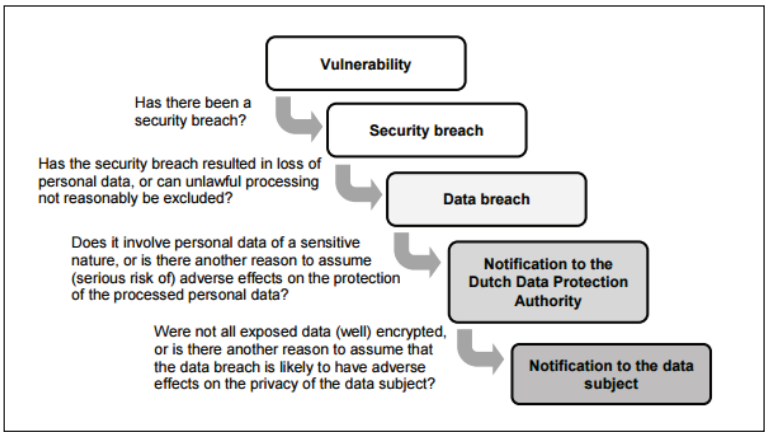


Figure 3: Dutch DPA disclosure sequence diagram (Dutch Data Protection Authority, 2015)

According to the Dutch DPA, after a vulnerability has been detected one must first determine if any security breach occurred due to the vulnerability. If a security breach was caused it must be determined if any data has been leaked due to the security breach. If any personal or sensitive data was exposed during the breach the Dutch DPA needs to be notified. If the data that was exposed was not secured prior to exposure, i.e. encrypted, the data subject needs to be informed of the breach. This sequence is depicted in Figure 3. Further details on how to proceed with each step of the Data Breach disclosure process is provided in detail within the Dutch DPA report (Dutch Data Protection Authority, 2015).

On 4 May 2016, the official texts of the Regulation and the Directive have been published in the EU Official Journal in all the official languages. While the Regulation will enter into force on 24 May 2016, it shall apply from 25 May 2018. The Directive enters into force on 5 May 2016 and EU Member States have to transpose it into their national law by 6 May 2018 (European Commision, 2016).

4. Experimental setup

In order to illustrate how a low cost open source solution can be used to gain valuable information regarding an organisations data breach potential, as simplistic experiment was set up to determine the data breach potential of three fictional organisations.

4.1 Data curation

In an attempt to obtain reasonably representative data for our experiments, data was collected from actual networks. The data was collected using easily accessible Free Open Source Software (FOSS) and Open Intelligence (OpenInt) solutions. The reason for using FOSS and OpenInt solutions is to provide a generalised solution for modelling the problem space. The proposed software is descriptive not prescriptive. If better or more accurate corporate data sets can be used it is recommended to us those data set but to perform similar analysis on those data sets.

For network traffic analysis Wireshark was used. Wireshark is simply a network packet analyser capable of capturing network packets, which flow through the network interface device. The analyser was setup to capture all traffic flowing through the corporate proxy. To maintain anonymity most of the captured data was stripped of any personally identifiable information. The only data that was logged is: the protocols used, possible encryption applied to those protocols, the browser user-agents used during website access. The user agent strings were grouped into simplified groupings, using regular expressions. The groupings were based on the variant of browser used, browser major version and rendering engine used. Figure 4 depicts a sample User-Agent string.

```
Mozilla/5.0 (Linux; U; Android 1.5; de-de; HTC Magic Build/CRB17) AppleWebKit/528.5+ (KHTML, like Gecko) Version/3.1.2 Mobile Safari/525.20.1

Mozilla/5.0 (Linux; U; Android 2.1-update1; en-au; HTC_Desire_A8183 V1.16.841.1 Build/ERE27) AppleWebKit/530.17 (KHTML, like Gecko)
Version/4.0 Mobile Safari/530.17

Mozilla/5.0 (Linux; U; Android 4.2; en-us; Nexus 10 Build/JVP15I) AppleWebKit/534.30 (KHTML, like Gecko) Version/4.0 Safari/534.30
```

Figure 4: Sample user-agent for android web browser

The IP addresses were also captured but they were normalized to reflect IP addresses within the private network IP address space rather than the corporate IP address space. Using a tool called TCPrewrite the corporate IP addresses were mapped to the private IP ranges of 192.168.0.0/16 and all external addresses were mapped to 10.0.0.0/8 (Turner, 2012).

Media Access Control (MAC) addresses were however not altered since they could be useful for future research. The MAC addresses contain Organisational Unique Identifiers (OUI). An OUI is a 24-bit unique identifier contained within the first 3 octets of a MAC address. OUIs are standardised and maintained by the IEEE (IEEE Standards Association). Vendor vulnerabilities were not used as a metric for this experimental setup, but the OUI values was used to provide an overview of the manufacturers of devices on the corporate network. **Table 4** shows a sample of the OUI data collected.

Table 4: Sample OUI data

Device Count	MAC Address	OUI Lookup
16	00:0c:29:ff:35:df	00:0C:29 VMware, Inc.
132	00:0f:1f:a7:8b:b6	00:0F:1F Dell Inc.
7	00:0f:f7:6e:39:cb	00:0F:F7 Cisco Systems, Inc
4	00:13:60:1c:17:9c	00:13:60 Cisco Systems, Inc
4	00:15:2b:06:a7:fe	00:15:2B Cisco Systems, Inc
4	00:16:47:9d:f2:d6	00:16:47 Cisco Systems, Inc
3	00:18:f8:90:f1:e0	00:18:F8 Cisco-Linksys, LLC
1	00:1d:72:8c:a5:69	00:1D:72 Wistron Corporation

Table 5: NSE scripts executed

NSE Scan	Description of Scan
http-git	Checks for a Git repository found in a website's document root /.git/<something>) and retrieves as much repo information as possible, including language/framework, remotes, last commit message, and repository description.
http-passwd	Checks if a web server is vulnerable to directory traversal by attempting to retrieve /etc/passwd or \boot.ini.
http-phpmyadmin-dir-traversal	Exploits a directory traversal vulnerability in phpMyAdmin 2.6.4-pl1 (and possibly other versions) to retrieve remote files on the web server.
http-shellshock	Attempts to exploit the "shellshock" vulnerability (CVE-2014-6271 and CVE-2014-7169) in web applications.
http-slowloris-check	Tests a web server for vulnerability to the Slowloris DoS attack without actually launching a DoS attack.
http-vmware-path-vuln	Checks for a path-traversal vulnerability in VMWare ESX, ESXi, and Server (CVE-2009-3733).
http-vuln-cve2012-1823	Detects PHP-CGI installations that are vulnerable to CVE-2012-1823, This critical vulnerability allows attackers to retrieve source code and execute code remotely.
http-vuln-cve2013-0156	Detects Ruby on Rails servers vulnerable to object injection, remote command executions and denial of service attacks. (CVE-2013-0156)

NSE Scan	Description of Scan
http-vuln-cve2014-2126	Detects whether the Cisco ASA appliance is vulnerable to the Cisco ASA ASDM Privilege Escalation Vulnerability (CVE-2014-2126).
http-vuln-cve2014-2127	Detects whether the Cisco ASA appliance is vulnerable to the Cisco ASA SSL VPN Privilege Escalation Vulnerability (CVE-2014-2127).
http-vuln-cve2014-8877	Exploits a remote code injection vulnerability (CVE-2014-8877) in Wordpress CM Download Manager plugin. Versions <= 2.0.0 are known to be affected.
smb-vuln-conficker	Detects Microsoft Windows systems infected by the Conficker worm. This check is dangerous and it may crash systems.
smb-vuln-ms08-067	Detects Microsoft Windows systems vulnerable to the remote code execution vulnerability known as MS08-067. This check is dangerous and it may crash systems.
ssl-cert-intaddr	Reports any private (RFC1918) IPv4 addresses found in the various fields of an SSL service's certificate. These will only be reported if the target address itself is not private. Nmap v7.30 or later is required.
ssl-heartbleed	Detects whether a server is vulnerable to the OpenSSL Heartbleed bug (CVE-2014-0160). The code is based on the Python script ssltest.py
ssl-known-key	Checks whether the SSL certificate used by a host has a fingerprint that matches an included database of problematic keys.
ssl-poodle	Checks whether SSLv3 CBC ciphers are allowed (POODLE)
sslv2-drown	Determines whether the server supports SSLv2, what ciphers it supports and tests for CVE-2015-3197, CVE-2016-0703 and CVE-2016-0800 (DROWN)

An internal assessment of corporate network assets was also performed using Nmap scan results. Nmap is a free and open source utility for network discovery and security auditing. The Nmap Scripting Engine (NSE) is a LUA scripting engine which has been built into the Nmap software to perform security audits. Table 5 list all the vulnerability detection scripts included in the assessments. The majority of the NSE scripts relate to involuntary data disclosure but it also includes scripts for known vulnerability detection. For example: Shellshock (Symantec, 2014), Heartbleed (Grubb, 2014), Conficer (McAfee, 2015) and Poodle (US-CERT, 2014).

The Nmap scans that were used were part of a collection of scans that have been conducted over several years of various companies. To maintain anonymity the IP address were also mapped to the private IP address space, similar to how the Wireshark traffic was mapped to the private address space.

After anonymization and parsing of the data the curated data set consisted out of:

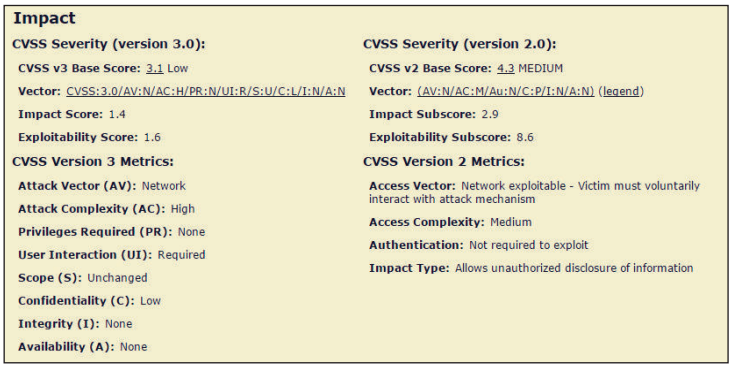
- 841 unique devices
- 37 User agent string groupings
- 62 unique manufacturer OUI numbers

The 841 devices were then further split into three distinct groups. Each device was randomly assigned to one of the three groups. For the purpose of the experiment each group represented a fictional organisation. The potential data leak exposure for each of the fictional organisations was computed using the steps discussed in Section 4.2. The results of these calculations and comparisons will be presented in Section 5.

4.2 Calculation of potential data exposure

Since not all vulnerabilities are equally dangerous, the Common Vulnerability Scoring System (CVSS) was used to apply a weighting to the vulnerabilities detected via the NSE vulnerability scans. The CVSS is a scoring system is a system proposed by the Forum of Incident Response and Security Teams (FIRST) to standardize IT vulnerability risk scores (FIRST, 2015). Figure 5 shows a sample CVSS entry for the Poodle vulnerability.

Each vulnerability count was multiplied by its CVSS score. In case the scan result does not match a CVSS entry the CVSS value was assumed to be 1. In case the CVSS had multiple values assigned to it, i.e. Poodle, the average of the scores was used. The total sum of all weighted vulnerability scores were used to assign a score to each of the fictional organisations.



Impact

CVSS Severity (version 3.0):

CVSS v3 Base Score: 3.1 Low

Vector: CVSS:3.0/AV:N/AC:H/PR:N/UI:R/S:U/C:L/I:N/A:N

Impact Score: 1.4

Exploitability Score: 1.6

CVSS Version 3 Metrics:

Attack Vector (AV): Network

Attack Complexity (AC): High

Privileges Required (PR): None

User Interaction (UI): Required

Scope (S): Unchanged

Confidentiality (C): Low

Integrity (I): None

Availability (A): None

CVSS Severity (version 2.0):

CVSS v2 Base Score: 4.3 MEDIUM

Vector: (AV:N/AC:M/Au:N/C:P/I:N/A:N) (legend)

Impact Subscore: 2.9

Exploitability Subscore: 8.6

CVSS Version 2 Metrics:

Access Vector: Network exploitable - Victim must voluntarily interact with attack mechanism

Access Complexity: Medium

Authentication: Not required to exploit

Impact Type: Allows unauthorized disclosure of information

Figure 5: National vulnerability database score for poodle vulnerability

The user-agent strings were used to determine if users were using out-of-date browsers to access the internet. As was stated earlier, aged technologies tend to have a higher vulnerability probability. A technical report, by Manners & Vanderbrink (2011) from the SANS intitute, documented how user-agents can be used to determine portentail threats to an organisation. For our purposes only the age of the user agent is taken into account. The potential data loss by browser was calculated as follows:

```
(Latest_release_version - User_version) = Potential_data_loss_score
```

It should be noted that due to the fact that rather old data sets were used this value ended up being skewed slightly.

As for the protocols used during the Wireshark capture, the protocol analyser within Wireshark was used to determine which protocols were the most used by each fictional organisation. Protocol usage could be an indicator of how secure communications are within an organisation.

5. Results

In this section a score card for each organisation is presented.

Table 6: Data breach potential score card for Organisation A

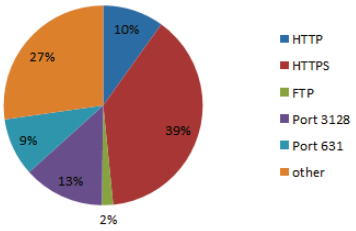
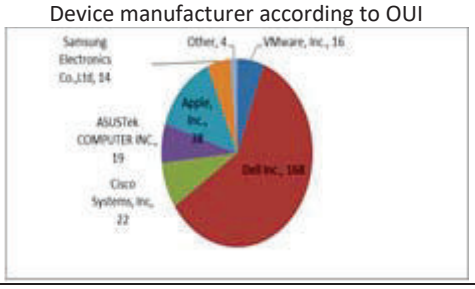
Number of Devices in network	281
Number of vulnerabilities detected	30
Cumulative CVSS score for all vulnerabilities detected	76.8
Protocol spread according to Wireshark 	Device manufacturer according to OUI 
Average data loss potential score for User-Agents	6.2
Most common User Agent-string	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; .NET CLR 1.1.4322; .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET CLR 3.5.30729; LEN2; .NET4.0C)

Table 7: Data breach potential score card for Organisation B

Number of Devices in network	280
Number of vulnerabilities detected	31
Cumulative CVSS score for all vulnerabilities detected	90.8
Protocol spread according to Wireshark:	Device manufacturer according to OUI

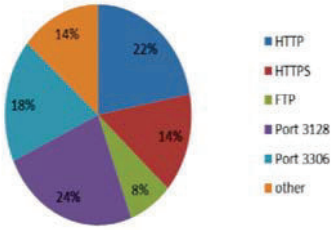
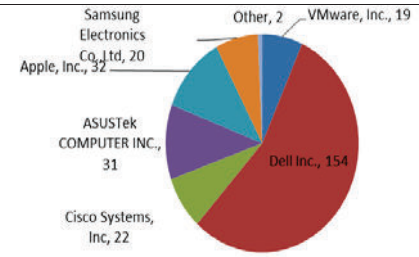
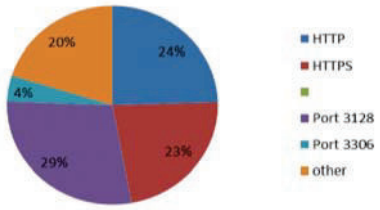
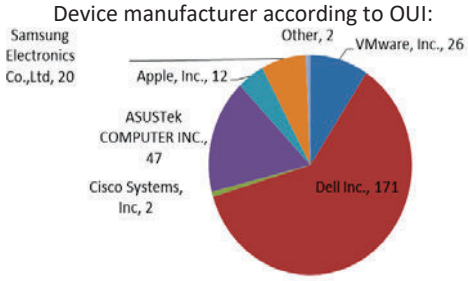
	
Average data loss potential score for User-Agents	6.1
Most common User Agent string	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)

Table 8: Data breach potential score card for Organisation C

Number of Devices in network	280
Number of vulnerabilities detected	18
Cumulative CVSS score for all vulnerabilities detected	55.4
Protocol spread according to Wireshark 	Device manufacturer according to OUI: 
Average data loss potential score for User-Agents	6.1
Most common User Agent string	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)

From these score cards the following interesting pieces of information can be obtained:

- Only organisation A has a higher utility of HTTPS (39) over HTTP (10%)
- Though organisation A and B have roughly the same number of potential vulnerabilities, based on the CVSS score Organisation B is vulnerable to far more dangerous vulnerabilities.
- The user-agent values were skewed due to the age of scans used for analysis, yet some valuable data could be obtained. The user-agents that were detected were nearly five years old at the time of analysis even though the average age of the scans was about three years old.
- One other interesting find within the User-Agent data was the User-agent string “mercuryboard_user_agent_sql_injection.nasl'. This user agent string is linked to CVE-2005-2028. This vulnerability provides unauthorized access and allows partial confidentiality, integrity, and availability violation; Allows unauthorized disclosure of information; Allows disruption of service.
- The OUI information reflects the company’s device manufacturers. These values seem consistent with the purchasing policies for the companies that were scanned. The companies relied on a Dell support contract with Cisco networking equipment. The ASUS devices were predominantly personal devices on the corporate network and the Apple and Samsung devices were mobile devices connected to the corporate networks. The only contentious OUIs detected were 00:90:96 ASKEY COMPUTER CORP and F0:DE:F1 Wistron Infocomm (Zhongshan) Corporation. Neither of these OUIs should have been detected on the network and only a single instance of each was detected. This may point to unauthorised devices on the network or a network compromise.

6. Conclusions

In this paper the growing problem of data breaches within Europe was highlighted in Section 2 and the European legal response to these breach were discussed in Section 3. The remainder of the paper focussed on establishing a low cost standardised set of measurements and tools which can be used to quantify the organisational risk of a data breach. The tools used were merely descriptive and more accurate results can be achieved by incorporating commercial tools or by enhancing the metrics used (Manners & Vanderbrink, 2011). The results obtained by these basic tools still provided useful insight into potential data loss.

The assessment of the potential data breach can be used to assess the risk of potential data pollution. Before one can start regulating the leaking of personal information one must first be able to measure the potential of a breach before one occurs.

In future work the researchers plan to investigate attribution, rights and responsibilities with regards to the data items and data subjects. This will include assigning responsibility of the users to maintain good governance by ensuring that all software is patched and up to date, and that vendors release security patches within a reasonable time frame for users to apply.

References

2002/58/EC, 2002. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector. [Online] Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:EN:HTML

Botha, J., Eloff, M. & Swart, I., 2016. Pro-active data breach detection: examining accuracy and applicability on personal information detected. Boston, USA, Proceedings of the 11th International Conference on Cyber Warfare and Security, ICCWS.

Breach Level Index, 2017. Data Breach Statistics by Year, Industry and More - Breach Level Index. [Online] Available at: http://www.breachlevelindex.com/ [Accessed 2 February 2017].

Daily News, Istanbul, 2013. Russian hackers stole 54 million Turkish citizens' ID data: Claim. [Online] Available at: http://www.hurriyetdailynews.com/russian-hackers-stole-54-million-turkish-citizens-id-data-claim.aspx?pageID=238&nID=59644&NewsCatID=338weak-state-servers-breach-causes-mass-identity-theft-in-turkey-haberi [Accessed 16 January 2017].

Doctorow, C., 2011. Context. 2nd ed. s.l.:Tachyon Publications.

Dutch Data Protection Authority, 2015. The data breach notification obligation as laid down in the Dutch Data Protection Act. [Online] Available at: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/policy_rules_data_breach_notification_obligation.pdf

ENISA, 2010. Data breach notifications in the EU. [Online] Available at: https://www.enisa.europa.eu/publications/dbn/at_download/fullReport

European Commision, 2016. Protection of personal data. [Online] Available at: http://ec.europa.eu/justice/data-protection/

FIRST, 2015. Common Vulnerability Scoring System, V3 Development Update. [Online] Available at: https://www.first.org/cvss

Government Gazette, 2013. Protection of Personal Information Act. [Online] Available at: http://www.gov.za/sites/www.gov.za/files/37067_26-11_Act4of2013ProtectionOfPersonalInfor_correct.pdf

Greenberg, A., 2016. Hack Brief: Turkey Breach Spills Info on More Than Half Its Citizens. [Online] Available at: https://www.wired.com/2016/04/hack-brief-turkey-breach-spills-info-half-citizens/ [Accessed 16 January 2017].

Grubb, B., 2014. Heartbleed disclosure timeline: who knew what and when. [Online] Available at: http://www.smh.com.au/it-pro/security-it/heartbleed-disclosure-timeline-who-knew-what-and-when-20140414-zqurk.html

IEEE Standards Association, n.d. Registration Authority. [Online] Available at: http://standards.ieee.org/develop/regauth/oui/ [Accessed 5 January 2017].

Manners, D. & Vanderbrink, R., 2011. The user agent field: Analyzing and detecting the abnormal or malicious in your organization. [Online] Available at: https://www.sans.org/reading-room/whitepapers/hackers/user-agent-field-analyzing-detecting-abnormal-malicious-organization-33874

McAfee, 2015. How to combat the W32/Conficker worm. [Online] vailable at: https://kc.mcafee.com/corporate/index?page=content&id=KB60909

Murdock, J., 2016. Anonymous hacker claims to leak hospital data of millions of Turkish citizens. [Online] Available at: http://www.ibtimes.co.uk/anonymous-hacker-claims-leak-hospital-data-more-10-million-turkish-citizens-1560985 [Accessed 16 January 2017].

Ponemon Institute LLC, 2016. 2016 Cost of Data Breach Study:, s.l.: Ponemon Institute LLC sponsored by IBM.

Schneier, B., 2015. Data and Goliath: The hidden battles to collect your data and control your world. s.l.:WW Norton & Company.

Shodan, 2016. Devices Vulnerable to Heartbleed. [Online] Available at: https://www.shodan.io/report/89bnfUyJ

Shodan, 2017. Countries vulnerable to Heartbleed. [Online] Available at: https://www.shodan.io/report/9F6Zj0R4

Shodan, 2017. IIS 2.0. [Online] Available at: https://www.shodan.io/report/1jePq5mb

Symantec, 2014. ShellShock: All you need to know about the Bash Bug vulnerability. [Online] Available at: http://www.symantec.com/connect/blogs/shellshock-all-you-need-know-about-bash-bug-vulnerability

Turner, A., 2012. TCPRewrite. [Online] Available at: http://tcpreplay.synfin.net/wiki/tcprewrite

US-CERT, 2014. SSL 3.0 Protocol Vulnerability and POODLE Attack. [Online] Available at: https://www.us-cert.gov/ncas/alerts/TA14-290A