Department of Energy Cyber Security Incidents

Working paper: Marshall Kuypers (mkuypers@stanford.edu) and Dr. Elisabeth Pate-Cornell (mep@stanford.edu)

February 2016

Abstract

Despite significant interest in cyber security, data on cyber security incidents remains scarce. On April 16, 2015, the US Department of Energy released data on 1,131 cyber security incidents through a Freedom of Information Act Request. While only containing the date, location, and type of incident, several interesting insights can be kneaded from the data. In this paper, we analyze the DOE security incident data and perform a statistical analysis on the rate of incidents. We find that the rate of cyber security incidents is decreasing over time and that incidents can be modeled stochastically. We believe that this is further evidence that cyber risk can be accurately evaluated with data-driven models.

Introduction

Despite the large interest in cyber security, researchers have had difficulty in validating cyber attack models because of a lack of data. Several cyber incident sharing initiatives now exist, although many rely on crowdsourcing incidents that are publically disclosed through the media, mandatory disclosure laws, or are self-reported.[1] While these data repositories are very valuable to researchers, there is a concern that reporting biases skew conclusions that might be made if all of the incidents were available.

There are remarkably few examples of datasets that contain all cyber security incidents recorded at an organization. Condon, He, and Cukier published an analysis of cybersecurity incidents in 2008, although the data consisted primarily of malware incidents [1]. Kuypers and Pate-Cornell have studied a dataset of 60,000 cyber security incidents at a large organization to inform cyber risk models [2]. Kuypers, Maillart, and Pate-Cornell conducted a statistical analysis of these incidents and found that investigation times have heavy-tails, the rate of incidents is relatively constant over time, and the impact of certain security safeguard investments (i.e. full disk encryption) can be empirically observed to reduce incident investigation times [3]. Several other papers have analyzed data breaches from publically available data sources that aggregate security incidents from many organizations [4,5,6]. Other datasets (especially those that detail incidents at a single organization) are incredibly difficult to obtain. The availability of data is crucial for the field of information security to advance. Data allows new insights to be discovered, models to be validated, and points to new directions for research.

In this paper, we present a brief analysis of a recently published list of cyber security incident impacting the US Department of Energy. We analyze the data to infer information about cyber security at DOE, and find that cyber incidents occur at a relatively constant rate over time.[2]

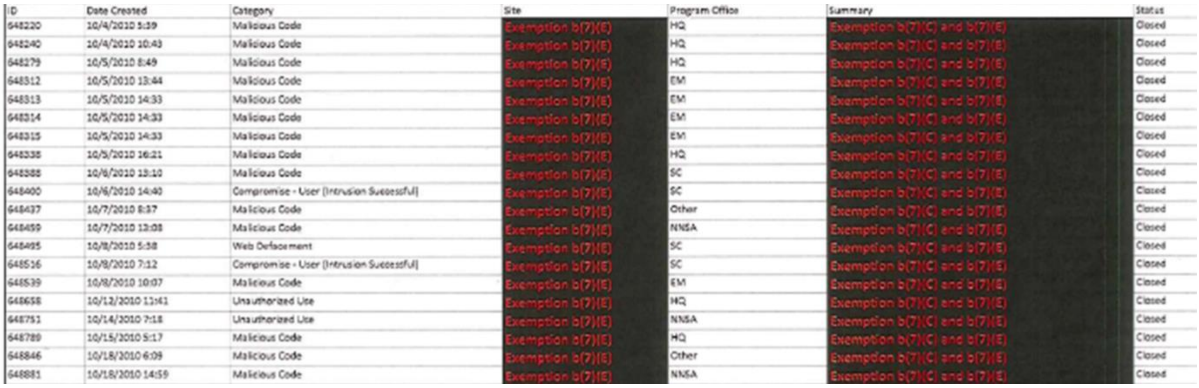
[1] For example, see the VERIS community framework at veriscommunity.net, or the Privacy Rights Clearinghouse at privacyrights.org.

[2] This is somewhat surprising, and while these data do not contain information about severity, it is clear that the rate of incidents is not dramatically increasing.

Data Description

The data come from the Joint Cybersecurity Coordination Center (JC3) for the Department of Energy, which receives cyber security incident reports from all DOE organizations. The dataset was obtained through a Freedom of Information Act request by Stephen Reilly, an investigative journalist from USA Today. [3] The data contain 1,131 cyber security incidents impacting 10 locations and are categorized into 7 types of attacks. The data contain an ID number, date, category, site (redacted), program office, summary (redacted), and status (closed for all).

Table 1: A sample of the data.



ID	Date Created	Category	Site	Program Office	Summary	Status
648220	10/4/2010 5:39	Malicious Code	Exemption b(7)(E)	HQ	Exemption b(7)(C) and b(7)(E)	Closed
648240	10/4/2010 10:43	Malicious Code	Exemption b(7)(E)	HQ	Exemption b(7)(C) and b(7)(E)	Closed
648279	10/5/2010 8:49	Malicious Code	Exemption b(7)(E)	HQ	Exemption b(7)(C) and b(7)(E)	Closed
648312	10/5/2010 13:44	Malicious Code	Exemption b(7)(E)	EM	Exemption b(7)(C) and b(7)(E)	Closed
648313	10/5/2010 14:33	Malicious Code	Exemption b(7)(E)	EM	Exemption b(7)(C) and b(7)(E)	Closed
648314	10/5/2010 14:33	Malicious Code	Exemption b(7)(E)	EM	Exemption b(7)(C) and b(7)(E)	Closed
648315	10/5/2010 14:33	Malicious Code	Exemption b(7)(E)	EM	Exemption b(7)(C) and b(7)(E)	Closed
648338	10/5/2010 16:21	Malicious Code	Exemption b(7)(E)	HQ	Exemption b(7)(C) and b(7)(E)	Closed
648388	10/6/2010 13:10	Malicious Code	Exemption b(7)(E)	SC	Exemption b(7)(C) and b(7)(E)	Closed
648400	10/6/2010 14:40	Compromise - User (Intrusion Successful)	Exemption b(7)(E)	SC	Exemption b(7)(C) and b(7)(E)	Closed
648437	10/7/2010 8:37	Malicious Code	Exemption b(7)(E)	Other	Exemption b(7)(C) and b(7)(E)	Closed
648459	10/7/2010 12:08	Malicious Code	Exemption b(7)(E)	NNSA	Exemption b(7)(C) and b(7)(E)	Closed
648495	10/8/2010 5:58	Web Defacement	Exemption b(7)(E)	SC	Exemption b(7)(C) and b(7)(E)	Closed
648516	10/8/2010 7:12	Compromise - User (Intrusion Successful)	Exemption b(7)(E)	SC	Exemption b(7)(C) and b(7)(E)	Closed
648539	10/8/2010 10:07	Malicious Code	Exemption b(7)(E)	EM	Exemption b(7)(C) and b(7)(E)	Closed
648658	10/12/2010 11:41	Unauthorized Use	Exemption b(7)(E)	HQ	Exemption b(7)(C) and b(7)(E)	Closed
648751	10/14/2010 7:18	Unauthorized Use	Exemption b(7)(E)	NNSA	Exemption b(7)(C) and b(7)(E)	Closed
648789	10/15/2010 5:17	Malicious Code	Exemption b(7)(E)	HQ	Exemption b(7)(C) and b(7)(E)	Closed
648846	10/18/2010 6:09	Malicious Code	Exemption b(7)(E)	Other	Exemption b(7)(C) and b(7)(E)	Closed
648881	10/18/2010 14:59	Malicious Code	Exemption b(7)(E)	NNSA	Exemption b(7)(C) and b(7)(E)	Closed

There are 6 categories of incidents:

Malicious code
Successful DDOS
Unsuccessful DDOS
Compromise (user)
Compromise (root)
Web defacement
Unauthorized use

The current JC3 website[4] contains a detailed description of incident types and incident impacts, although the released data do not match the website perfectly.[5] There are several curiosities in the data, including a lack of incidents labeled 'Loss, Theft, or Missing' and 'Phishing'.[6]

[3] The article that he authored about the dataset can be found here [7].

[4] http://energy.gov/cio/office-chief-information-officer/services/incident-management/jc3-incident-reporting

[5] While some incident descriptions overlap with the dataset, the website contains additional types of incidents, and does not list several types found in the data.

[6] The probability that a laptop has never been lost at DOE is virtually 0. We would expect that these incidents should be included in the database, given that the JC3 website lists 'Loss, Theft, or Missing' as an incident category as far back as April 10, 2013 (obtained from the Internet Archive: https://web.archive.org/web/20130410050059/http://energy.gov/cio/office-chief-information-officer/services/incident-management/jc3-incident-reporting). Another possibility is that DOE filtered these incidents before releasing the data, but this too would be strange, given that the original FOIA request was for "Database, spreadsheet or list of all security incidents reported to the U.S. Department of Energy Joint Cybersecurity Coordination Center (JC3) between October 1, 2011 and October 1, 2014…".

Additionally, there are several program offices that are listed using 2-4 letter abbreviations. Using other DOE documents, we can identify the Program Office associated with each abbreviation.[7]

HQ	(Headquarters)
NNSA	(National Nuclear Security Administration)
SC	(Office of Science)
EM	(Office of Environmental Management)
EE	(EERE: Office of Energy Efficiency and Renewable Energy)
PA	(Possibly PMA: Power Marketing Administrations, or PA: Public Affairs)
NE	(Office of Nuclear Energy)
LM	(Office of Legacy Management)
FE	(Office of Fossil Energy)
Other	(Other)

Analysis

The dataset contains very little information, but there are still many interesting insights that can be gained from analyzing the data. Figure 1 shows the cumulative number of cyber security incidents over time, color coded by incident type. It is immediately apparent that the majority of incidents involve malicious code and that rate of incidents is relatively constant over time (since the graph of cumulative incidents is relatively strait.

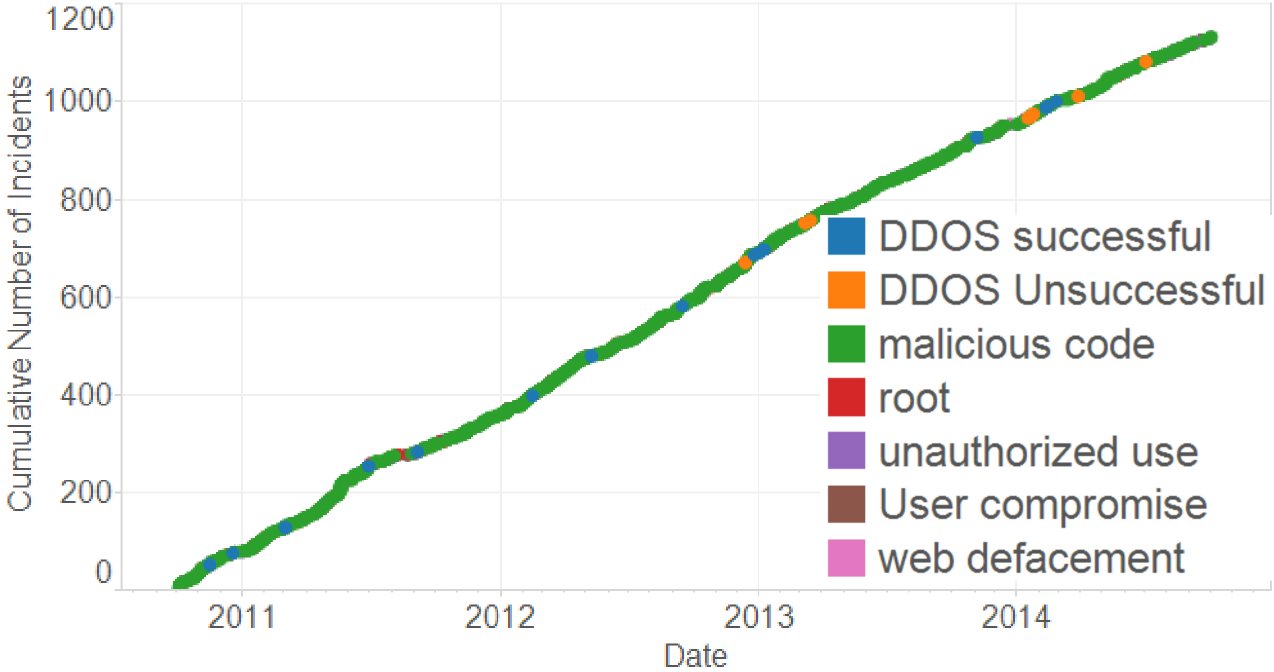


Figure 1: The cumulative number of incidents plotted over time. Note that the relatively constant slope suggests that the rate of cyber security incidents is relatively constant over time.

Figure 2 highlights the different types of incidents that occur over time. Again, note that while Web defacements occur somewhat clustered, all other incidents occur relatively constantly. Web

[7] See [8] for abbreviations.

defacements and DDoS attacks occur the most rarely while unauthorized use, root, and user compromises occur at roughly the same rate.

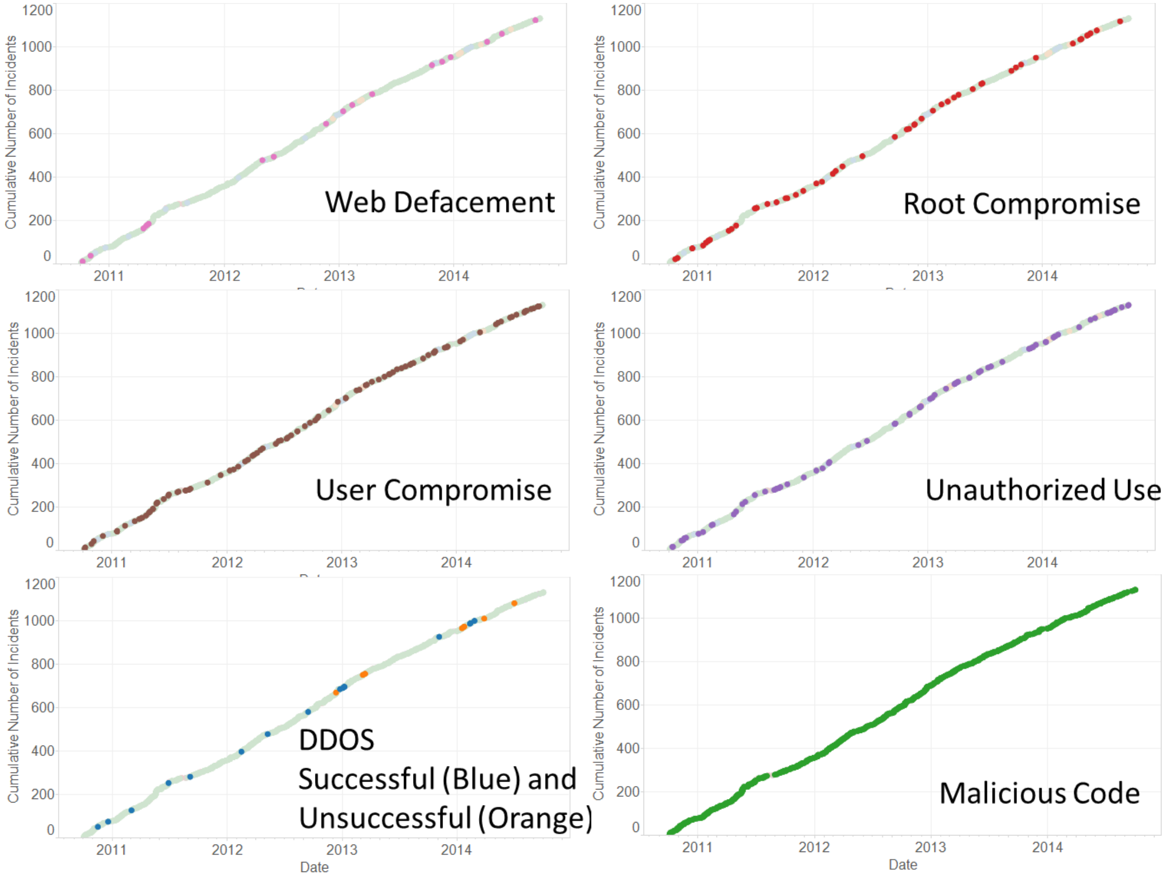


Figure 2: Incidents over time, highlighted by type. Note that Malicious Code incidents occur most often, followed by User Compromises.

The data span from October 4, 2010 and October 3, 2014, meaning that every month has been recorded four times.[8] Figure 3 shows that the summer months typically experience fewer incidents. It is clear from the data that security incidents are created M-F during normal working hours.[9] While it is not possible to say if for certain, these trends are likely indicative of the security investigators, and not the adversaries.[10] Further, it is possible to surmise that the time stamps are in Eastern Time (the location of JC3), not UTC.[11]

[8] Note that the FOIA request was from October 1, **2011**, but data was delivered from 2010 onwards.

[9] Interestingly, JC3 is listed as a 24 X 7 X 365. https://www.first.org/members/teams/jc3-circ

[10] Other studies [2] with incident data from organizations with a 24 hour security operations center (SOC) do not exhibit such large decreases during non-working hours.

[11] If the time stamps were in UTC, the JC3 center would perform the majority of its work between 1AM and 8AM local time (Maryland).

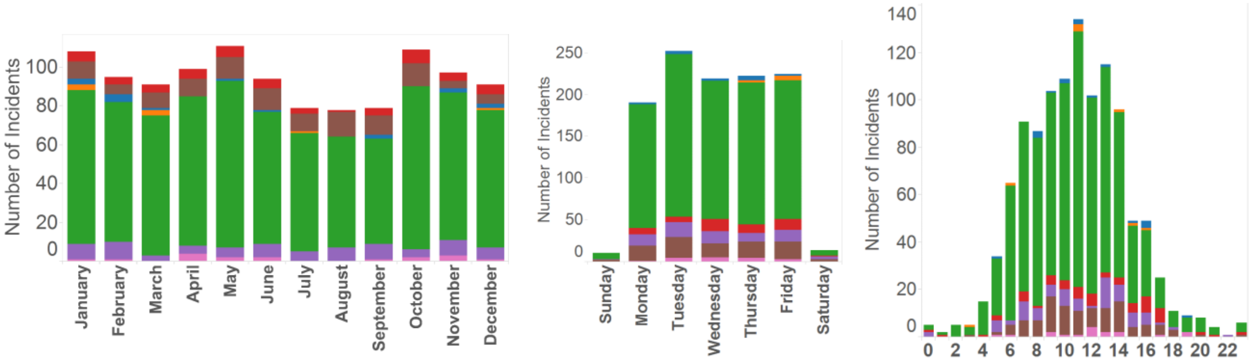


Figure 3: Graphs showing the number of incidents by month, day, and hour. The colors represent the type of incident.

It is difficult to tell if the data are machine generated or human generated. In some cases, multiple incidents are created in the same minute (see table 2). However, these could either be automatically created by a tracking program, created in bulk by an investigator, or created individually by an investigator. The incident shown in table 2 is particularly interesting because six incidents are created in a one minute time period, with another being generated the next minute.

Table 2: Seven incidents created in quick succession.

Date	Incident Number	Location	Type	ID
6/28/11 11:16	245	EM	Malicious code	661631
6/28/11 11:16	246	EM	Malicious code	661632
6/28/11 11:16	247	EM	Malicious code	661633
6/28/11 11:16	248	EM	Malicious code	661634
6/28/11 11:16	249	EM	Malicious code	661635
6/28/11 11:16	250	EM	Malicious code	661636
6/28/11 11:17	251	EM	Malicious code	661637

Table 2 suggests another interesting feature, which is the use of the ID field. For the six incidents that are rapidly generated, the ID's are sequential. However, while the ID's for other incidents are increasing, they are typically not sequential. This suggests that the ID may denote a unique incident number, with other incidents that are not considered full 'security incidents' in between. An old version of the JC3 website describes type 1 incidents (successful incidents) and type 2 incidents (reconnaissance and attempted intrusions), suggesting that type 2 incidents are given a unique ID but were not included in the dataset.[12]

To test the theory that each ID corresponds to a unique incident (some of which are not listed in this dataset), we could perform an analysis to see if the number of entries between the ID numbers correlates with the length of time between the entry time stamps. Alternatively, we can see how many ID numbers occur year to year (measured in 365 day increments).
Year 1: 18,926
Year 2: 26,397
Year 3: 24,614
Year 4: 21,342

[12] https://web.archive.org/web/20130410050059/http://energy.gov/cio/office-chief-information-officer/services/incident-management/jc3-incident-reporting

Since the number of IDs each year is relatively similar, it could be the case that the ID's that are not listed indeed specify unsuccessful intrusions, and the rate of unsuccessful incidents is approximately constant over time as well (or even decreasing).

Site Analysis

The majority of incidents in the dataset are attributed to Headquarters with the Office of Science having a bit more than half as many. Figure 4 shows the frequency of incidents at other locations. It is interesting to note that the Office of Science has many more User Compromise and Root Compromise incidents than would be expected from the number of Malicious Code incidents.

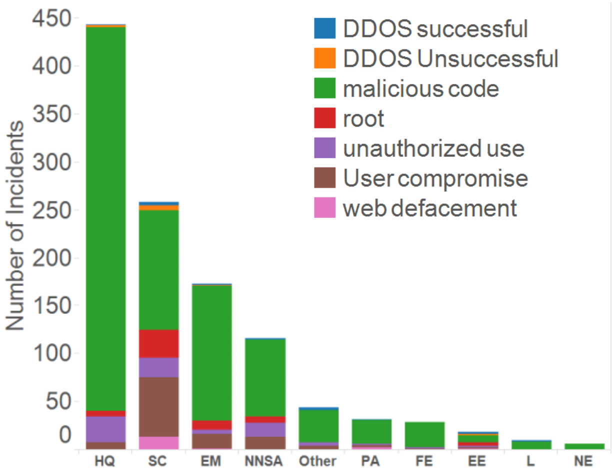
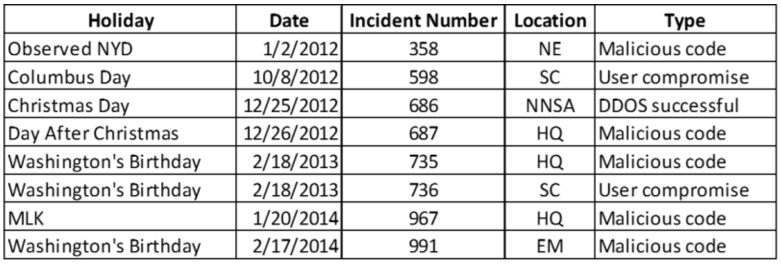


Figure 4: Number of incidents by location.

Holiday Schedule

Cyber security incidents are unlikely to be dated as a Federal Holiday. Table 3 shows the rare cases where an incident occurred on a day off. Christmas 2012 was likely an unhappy day for at least some security investigators, who opened a successful DDoS incident at the National Nuclear Safety Administration on Christmas day 2012.

Table 3: Incidents opened on Federal Holidays.



Holiday	Date	Incident Number	Location	Type
Observed NYD	1/2/2012	358	NE	Malicious code
Columbus Day	10/8/2012	598	SC	User compromise
Christmas Day	12/25/2012	686	NNSA	DDOS successful
Day After Christmas	12/26/2012	687	HQ	Malicious code
Washington's Birthday	2/18/2013	735	HQ	Malicious code
Washington's Birthday	2/18/2013	736	SC	User compromise
MLK	1/20/2014	967	HQ	Malicious code
Washington's Birthday	2/17/2014	991	EM	Malicious code

The observation that incidents are unlikely to be created on Federal Holidays suggests that the field labeled ‘event created’ notes when an incident is created by an analyst, and is not automatically created from network monitoring software.

Major Incidents

Several major hacking incidents have occurred at DOE over the past five years. One of the largest occurred in July of 2013 and involved the loss of 104,000 records of Personally Identifiable Information (PII) [9]. Costs involved a reported $1.6M in credit monitoring and $2.1M in lost productivity, as DOE employees were given 4 hours off of work to remediate damage to their personal identity. The Office of Inspector General’s report offers some details about the incident and documents serious problems with IT networks at DOE facilities [9]. The document further states that the incident took place against the DOE headquarters (page 18), and offers a timeline of events (page 17).

Viewing the data records, it is still difficult to determine which data entry corresponds to the described incident. The document states that the Energy Information Technology Services (EITS) group was alerted on July 2nd, 2013 that someone was trying repeatedly to gain access to a system. On July 24th, the system was breached according to logs, and on July 25th, EITS was alerted again. On August 8th, the breach was identified. Table 4 shows the incidents in the surrounding time period. Note that no incidents are listed for any location on August 8th. It is possible that the incident was not created until the following day, but this is not clear.

Table 4: Incidents opened around the July 2013 breach.

Date	Incident Number	Location	Type
7/2/13 5:42 PM	835	NNAS	malicious code
7/5/13 3:21 PM	836	SC	malicious code
7/8/13 6:22 AM	837	HQ	malicious code
7/8/13 10:36 AM	838	PA	malicious code
7/11/13 1:59 PM	839	SC	User compromise
7/12/13 5:52 AM	840	EM	malicious code
7/12/13 12:21 PM	841	SC	unauthorized use
7/15/13 11:28 AM	842	PA	malicious code
7/17/13 10:49 AM	843	HQ	malicious code
7/17/13 11:53 AM	844	Other	malicious code
7/18/13 7:48 PM	845	NNAS	malicious code
7/22/13 1:12 PM	846	SC	unauthorized use
7/24/13 2:04 PM	847	PA	User compromise
7/29/13 1:05 PM	848	HQ	malicious code
7/30/13 7:35 AM	849	HQ	malicious code
7/30/13 9:45 AM	850	HQ	malicious code
7/30/13 10:37 AM	851	HQ	malicious code
7/30/13 12:13 PM	852	EM	malicious code
8/5/13 2:55 PM	853	EM	malicious code
8/6/13 8:43 AM	854	Other	malicious code
8/7/13 4:32 AM	855	EM	malicious code
8/7/13 9:36 AM	856	EM	User compromise
8/9/13 2:10 PM	857	EM	malicious code
8/9/13 4:07 PM	858	HQ	User compromise
8/12/13 6:09 PM	859	SC	User compromise

Another major incident occurred in mid-January of 2013 involving PII loss of 14,000 individuals [10]. Bill Gertz of the Washington Free Beacon reported that officials stated that a breach occurred two weeks earlier in an article published Feb 4th, 2013 [11]. Again, the incident was reported to have impacted HQ. Table 5 shows the incidents surrounding this time period. Note that HQ experienced several Malicious Code incidents during this time. Bill Gertz reported that 14 servers and 20 workstations were penetrated, which would be consistent with the multiple incidents recorded in the data.

Table 5: Incidents opened around the January 2013 breach.

Date	Incident Number	Location	Type
1/15/13 1:55 PM	702	SC	web defacement
1/15/13 2:37 PM	703	EM	User compromise
1/16/13 10:18 AM	704	NNAS	unauthorized use
1/16/13 3:34 PM	705	SC	malicious code
1/17/13 2:15 PM	706	SC	root
1/17/13 4:01 PM	707	SC	malicious code
1/18/13 12:37 PM	708	HQ	malicious code
1/22/13 8:25 AM	709	HQ	malicious code
1/22/13 11:08 AM	710	HQ	malicious code
1/22/13 11:17 AM	711	HQ	malicious code
1/22/13 11:32 AM	712	HQ	malicious code
1/22/13 11:42 AM	713	HQ	malicious code
1/22/13 11:57 AM	714	SC	malicious code
1/22/13 2:16 PM	715	HQ	malicious code
1/23/13 2:00 PM	716	SC	unauthorized use
1/28/13 4:23 PM	717	EM	malicious code

The US government attributed the January 2013th attack to Chinese hackers, although an Iranian hacker group named 'PARASTOO' posted data to pastebin.com on January 20th 2013 claiming to have hacked DOE.[13] Other sources expressed skepticism that the group was behind the attack.

Insights

While the data to not contain any information on the severity of different incidents, the rate at which incidents occur can be studied. Figure 5 shows the number of incidents in each month.[14] A linear regression is performed in table 6. We find that the rate of incidents is falling over time, but that this is mostly driven by a decrease in malicious code incidents. The rate of Root Compromises, User Compromises, DDoS attacks, Web Defacements, and Unauthorized use incidents is either constant or decreasing.

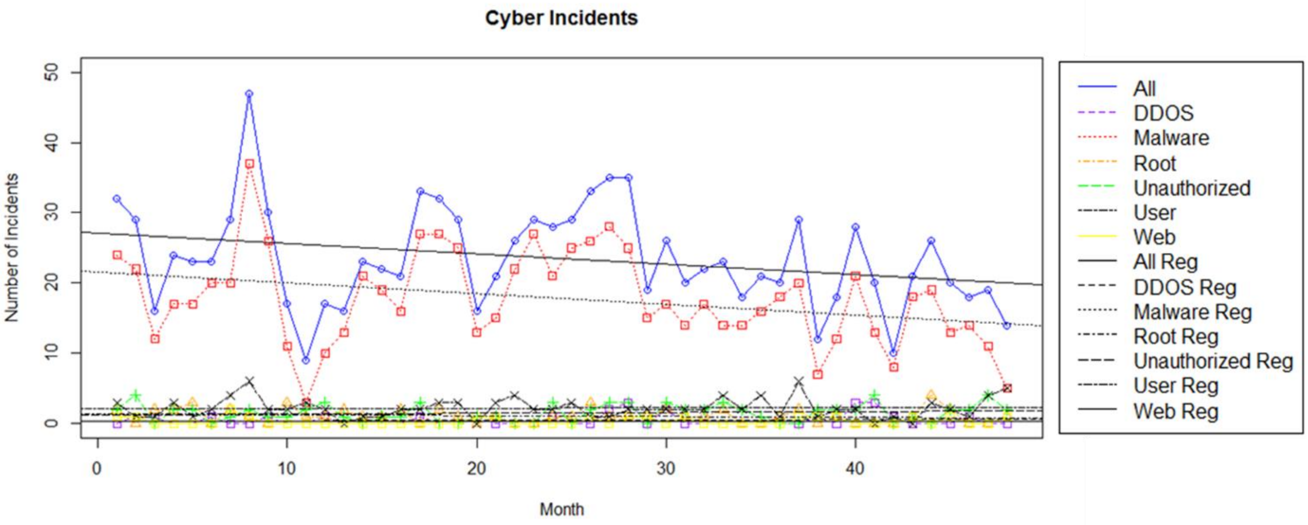


Figure 5: A plot of the number of cyber security incidents per month, along with a linear regression.

Table 6: Parameters for the rate of cyber security incidents by incident type and location.

Type	Average Per Month	Standard Deviation	Regression	Standard Error
All	23.50	7.30	-0.15	7.08
DDOS	0.50	0.85	0.01	0.85
Malicious Code	17.81	6.73	-0.15	6.44
Root	1.10	0.99	-0.01	0.99
Unauthorized Use	1.52	1.22	0.01	1.23
User Compromise	2.21	1.41	0.00	1.43
Web Defacement	0.35	0.53	0.00	0.53

Location	Average Per Month	Standard Deviation	Regression	Standard Error
HQ	9.25	4.53	-0.09	4.39
SC	5.33	2.00	-0.03	1.98
EM	3.58	2.95	-0.04	2.93
NNAS	2.42	1.93	-0.01	1.94

Figure 6 shows that the rate of incidents is decreasing at each of the four locations that exhibit the most number of incidents.

[13] Media reported that the post was on January 21, although the post is dated January 20th. The post can be seen here: https://cryptome.org/2013/01/parastoo-hacks-doe.htm

[14] Note that the last month (October 2014) is not included because a full month of data is not present.

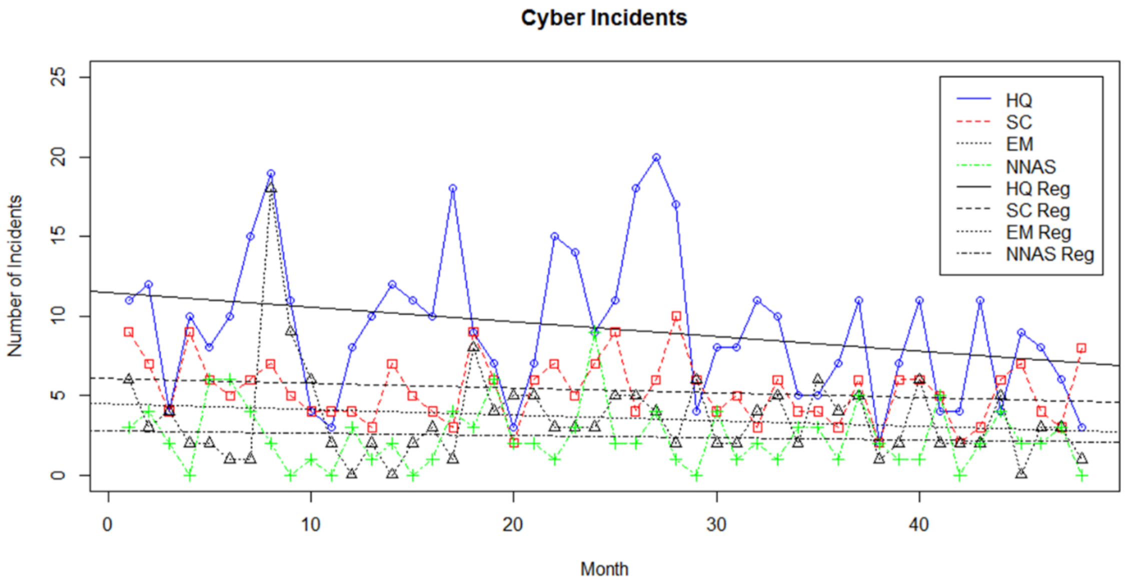


Figure 6: A plot of the number of cyber security incidents per month for several locations, along with a linear regression.

The arrival process of security incidents can be studied as well. We might expect that the number of incidents follows a Poisson distribution, given that security incidents arise from a process where many endpoints (servers, workstations, etc.) have a small probability of being compromised.

To gain more insight, the interval between incidents can be calculated. Here, we limit the analysis to intervals less than ten hours, because the effect of the workday can be seen in data and distort the results.

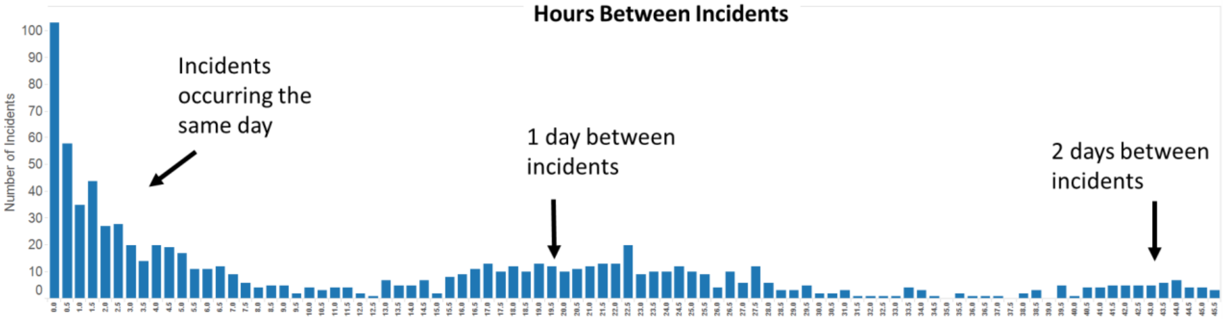


Figure 7: A graph of the time between incidents. Note that the time decreases exponentially (shown in figure 8), but additional peaks occur that correspond to roughly a work day between incidents.

For processes that exhibit a Poisson distribution, the time between arrivals follows an exponential distribution. We test several distributions and find that the data are best explained by an exponential distribution with $\frac{1}{\lambda} = 2.5869$, meaning that the expected time between incidents is roughly 2 hours and

35 minutes. The arrival process was tested against many other distributions, including gamma, logistic, normal, Weibull, negative binomial, and many others.[15]

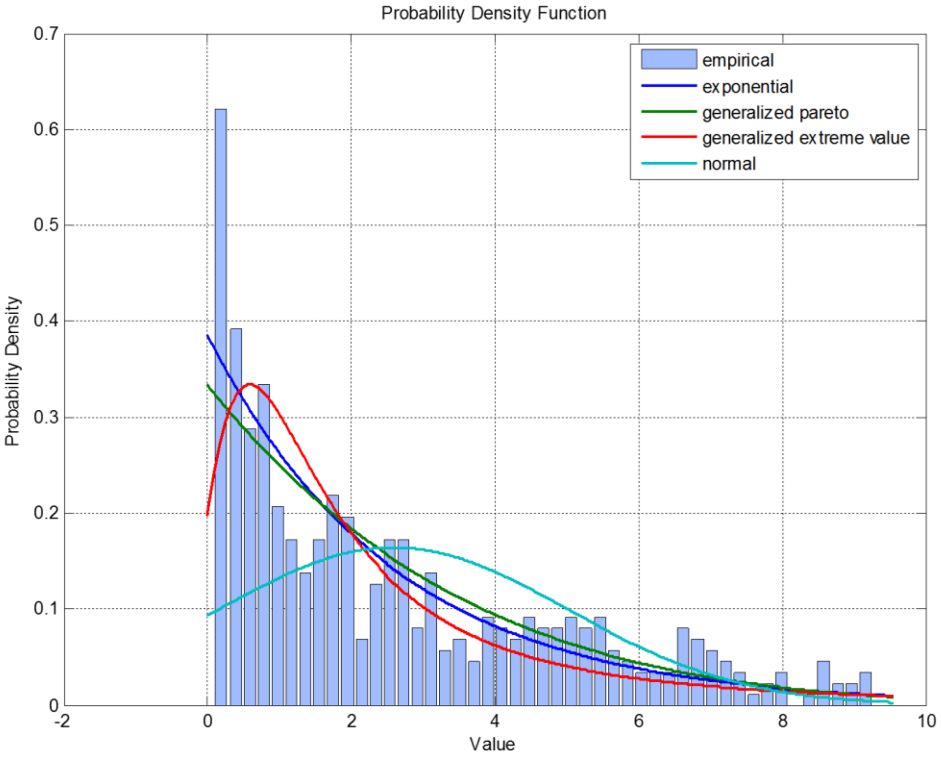


Figure 8: Probability density function for the time between incidents, along with several fits. The exponential distribution gives the best fit.

Conclusions

In many ways, cyber security is still in its infancy. Rigorous methods for assessing cyber risk are just beginning to emerge, and basic ground truths about cyber security incidents are still largely unknown. This paper demonstrates how even heavily redacted data can be used to determine interesting insights about cyber security at a large organization. In particular, while the impact of cyber incidents has not been assessed, it is clear from these data that cyber security incidents are not currently accelerating in frequency. In fact, cyber security incidents are shown to occur relatively consistently over time and may even be decreasing.

It is the author's hope that these datasets will encourage other organizations to release non-sensitive data about cyber security incidents. The community of information security would benefit enormously from unbiased, publically available, and complete data sources.

[15] We use the matlab function 'allfitdist' by Mike Sheppard, available here: http://www.mathworks.com/matlabcentral/fileexchange/34943-fit-all-valid-parametric-probability-distributions-to-data

Acknowledgements

The authors would like to thank Alex Keller and Elise Kuypers for help in cleaning the data from pdf format to text format. In the spirit of encouraging other research in this area, please email mkuypers@stanford.edu if you would like to request the cleaned dataset.

References

[1] Condon, Edward, Angela He, and Michel Cukier. "Analysis of computer security incident data using time series models." Software Reliability Engineering, 2008. ISSRE 2008. 19th International Symposium on. IEEE, 2008.

[2] Kuypers, M.A., and Pate-Cornell, M.E., "Quantitative Cyber Risk," Society for Risk Analysis Annual Meeting. Arlington, Virginia. December 7-9, 2015.

[3] Kuypers, M.A., Maillart, T., and Pate-Cornell, M.E. "An Empirical Analysis of Cyber Security Incidents at a Large Organization," Submitted to the Workshop on the Economics of Information Security (WEIS) 2016.

[4] Wheatley, Spencer, Thomas Maillart, and Didier Sornette. "The Extreme Risk of Personal Data Breaches & The Erosion of Privacy." arXiv preprint arXiv:1505.07684 (2015).

[5] Maillart, T., and D. Sornette. "Heavy-tailed distribution of cyber-risks." The European Phys-ical Journal B 75.3 (2010): 357-364.

[6] Edwards, Benjamin, Steven Hofmeyr, and Stephanie Forrest. "Hype and Heavy Tails: A Closer Look at Data Breaches. "Workshop on Economics of Information Security"

[7] Reilly, Stephen. "Records: Energy Department struck by cyber attacks," USA Today. 11 September 2015. http://www.usatoday.com/story/news/2015/09/09/cyber-attacks-doe-energy/71929786/

[8] United States Department of Energy Corporate Overview. 2012. (Page 6). http://energy.gov/sites/prod/files/DOE_Corporate_Overview-2012.pdf

[9] Special Report: The Department of Energy's July 2013 Cyber Security Breach. U.S. Department of Energy Office of Inspector General, Office of Audits and Inspections. December 2013. http://energy.gov/sites/prod/files/2013/12/f5/IG-0900.pdf

[10] King, Rachael. "Department of Energy Hacked." Wall Street Journal. February 4 2013. http://blogs.wsj.com/cio/2013/02/04/department-of-energy-hacked/

[11] Gertz, Bill. "Cyber Breach: Energy Department networks hit by sophisticated cyber attack." The Washington Free Beacon. Feburary 4 2013. http://freebeacon.com/politics/cyber-breach/