White Paper

Data Loss by the Numbers

Table of Contents

Executive Summary 3

Analysis 4

Findings 4

- External versus internal breaches 4
- Incidents by vertical 5
- Incidents by data type 6
- Data types by vertical 7

Conclusion 7

Executive Summary

Virtually every day there are stories about data loss in the news. Scammers, fraudsters, hackers, and malicious insiders are making off with sensitive business and personal information. In some cases, the events are accidental, but in many other cases, nefarious activities are uncovered. The reasons aren't opaque. Data is valuable, for legitimate purposes and criminal purposes alike, and there have been plenty of high profile examples of data loss such as:

Sony Corporation (2011)	77 million records compromised
Heartland Payment Systems (2009)	130 million records compromised
RockYou Inc. (2009)	32 million records compromised
TJX Companies Inc. (2007)	94 million records compromised

In many data loss cases, the results can be devastating:

- Brand damage and damage to shareholder faith
- Legal fees, class action lawsuits, and public relations costs
- Regulatory fines and victim notification for state disclosure laws
- Credit monitoring services and other free goods and services to retain customers
- Service downtime, breach investigations, lost customers, and lost revenue

Much of the data resides in data centers and that's where the data loss occurs. A lack of security controls on servers, storage, content, and networks, across these data centers is often attributed to data loss. Further, data centers have been going through a generational change over the last few years with consolidation, virtualization, and the use of cloud services. Given the complexity of these next-generation data centers coupled with the massive amounts of data they store and process, if security isn't factored in—preferably at the design and architecture phase of data center construction—the data loss numbers will continue to rise.

To apply some statistical analysis to data loss, this paper focuses on raw data obtained through the Open Security Foundation's Data Loss Database or DataLossDB. This data is considered to be accurate, current, and unbiased. We have analyzed their data and applied a number of statistical measures to produce a handful of key statistics. It is our intention to help raise awareness while providing usable statistics for supporting improved decision making. Here are some of the high-level results.

- External attacks accounted for 52 percent of the incidents and 60 percent of the compromised records
- Malicious insiders had the largest average number of compromised records per breach at 72,325
- Across all verticals, the average number of records affected per breach is 58,600
- The highest numbers of reported attacks by vertical were retailers, universities, and healthcare providers
- The lowest numbers of reported attacks by vertical were the federal government, industry, and K-12 schools
- The most common data types compromised were names and/or addresses, Social Security numbers and equivalents, and credit card numbers

After the Breach: Epsilon

In 2011, a data breach of email marketer Epsilon exposed the names and email addresses of millions of customers. Because Epsilon services are used by a large number of highly visible companies, there was a rash of notifications sent out to warn customers of potential fraud. Companies having to engage in customer notification warnings included: Barclaycard US, Capital One, Best Buy, JPMorgan, Citigroup, TiVo, Disney Destinations, New York & Company, Walgreens, Marriott, and many others.

(Source: http://mcaf.ee/5342e)

Analysis

We investigated at a total of 386 incidents between January 2009 and April 2011 that impacted 23 million records. To derive more generalized averages and minimize extreme statistical variances, we excluded some of the larger breaches like Heartland Payment Systems. We broke attack sources into four categories: external, malicious insiders, accidental insiders, and unknown. We limited our breach categories to areas that are most directly associated with data centers such as information exposure through malware and computer-based intrusion, and excluded incidents associated with categories like improper computer disposal, improper document disposal, lost laptops, and stolen backup tapes.

We further analyzed the data across 10 business verticals including:

- Retailers
- Financial services
- Technology companies
- US Federal Government
- State governments
- Universities
- K-12 schools
- Healthcare providers
- Insurance companies
- Industry (critical infrastructure, manufacturing, and related fields)

Our goal was to address four specific questions related to data loss:

1. How do the number of breaches and the number of records compromised per breach differ when comparing external and internal incidents?
2. Which verticals suffered the highest and lowest levels of reported breaches?
3. What type of data was most often compromised?
4. By vertical, what were the most common types of data compromised?

After the Breach: Ellery Systems Inc.

In the mind 1990s, Ellery Systems, based in Boulder Colorado, suffered a breach when an employee stole source code and gave it to an international competitor. This theft directly resulted in the company going out of business and the US losing its competitive advantage in a strategically important emerging industry. The prosecution had to be dropped and the malicious insider went free because the US did not, at that time, have an effective law for the prosecution of economic espionage. This case was in part responsible for the passing of the 1996 Economic Espionage Act. More than a decade later in 2010, the US-based firm CyberSitter announced that it was suing the Chinese government and several organizations for software theft; they cited the 1996 Economic Espionage Act.

(Source: http://mcaf.ee/bknqi)

Findings

External versus internal breaches

When securing data centers, and investigating methods to mitigate data loss, one of the most common discussions is around the greatest source of risk. Is the risk highest from an external attacker, a legitimate insider such as a privileged user, for example, a system administrator or a database administrator, or perhaps just carelessness?

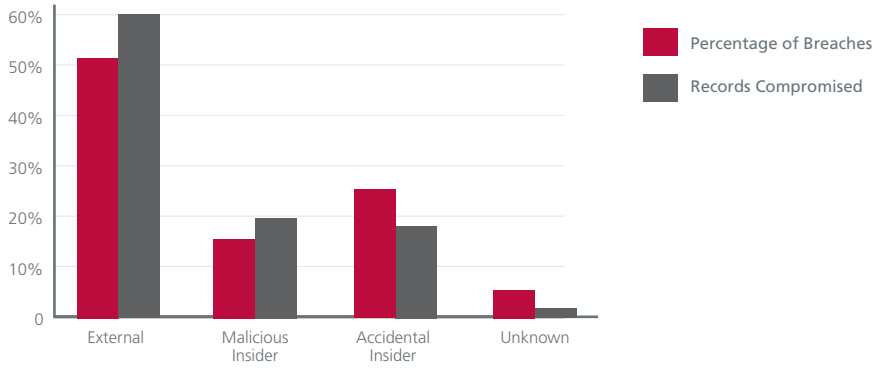

Figure 1. Percentage of breaches and records compromised.

External attacks accounted for about half the breaches at 52 percent of the incidents. These attacks were responsible for 60 percent of the records compromised. Malicious insiders accounted for 16 percent of the incidents with 20 percent of the records compromised.

The average numbers of records compromised per breach were:

Malicious insider	72,325
External	67,070
Accidental insider	40,046
Unknown	27,133

On average, each of the external attacks and the malicious insider attacks impacted more records than the accidental insiders or unknown incidents. While the total of insider incidents equated to 44 percent, the great majority were accidental. Because malicious insiders have trust and access and because their attacks are easier to perpetrate (for example, they can simply download files for which they have legitimate access onto their USB stick), it makes intuitive sense that while the number of incidents from external attackers was much higher along with the total records affected, the average number of records per incident was highest for malicious insiders.

Incidents by vertical

Of the 386 incidents analyzed, the number of breaches per vertical was highest for retail, universities, and healthcare providers, while the lowest number of breaches occurred within the US Federal Government, industry, and K-12 schools. Notice that the total breaches per vertical equal 415, not 386, which was defined as the sample set. This is because some organizations are listed across multiple verticals such as financial plus insurance and university plus healthcare.

After the Breach: Heartland Payment Systems

In 2010 Heartland reached a $60 million settlement with VISA to resolve all potential claims in respect to the 2009 breach. A year earlier, Heartland reached a similar settlement with Amex for $3.5 million.

(Source: http://mcaf.ee/gvxkh)

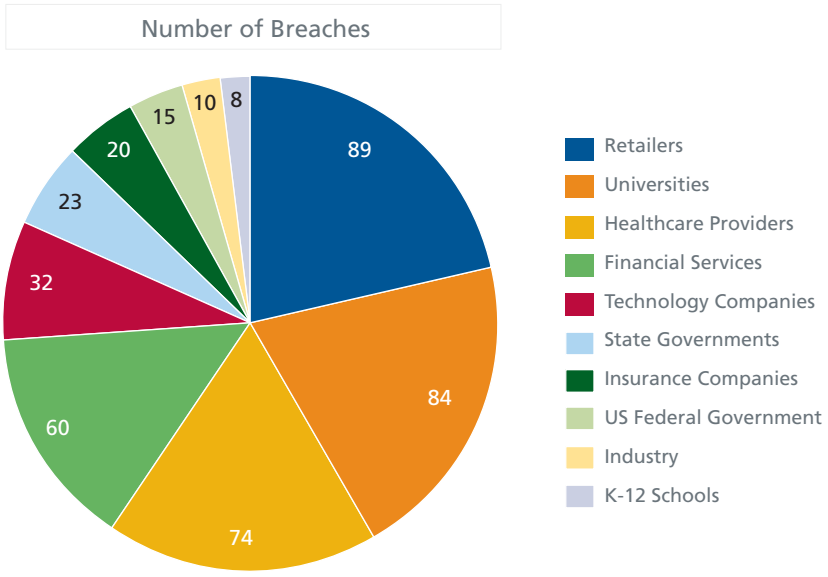

Figure 2. Number of incidents by vertical.

Retailers, universities, and healthcare providers topped the list while the US Federal Government, industry, and K-12 schools were at the bottom. It is important to note that not all verticals are subject to the same disclosure regulations. Many organizations within critical infrastructures, like electric power, for example, don't need to disclose many incidents. In the case of the federal government, specifics around breaches many not be shared with the public for any number of reasons such as national security.

Across all the verticals the average number of records affected per breach is 58,600. Note again that extreme cases were removed for more generalized statistical averages.

Incidents by data type

The three most significant data types compromised were names and/or addresses, Social Security numbers and equivalents, and credit card numbers. Some of the percentages in the graph will add up to more than 100 percent because an incident may affect more than one data type.

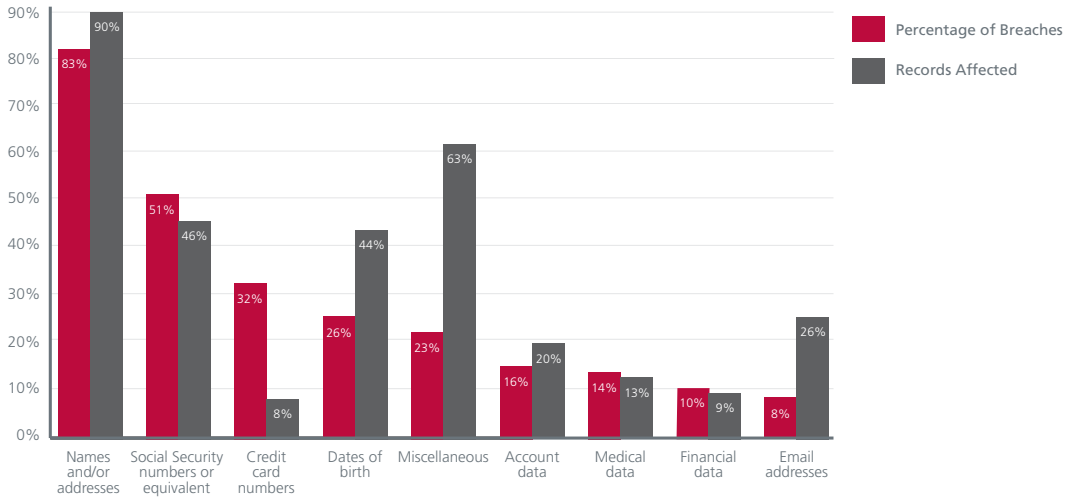

Figure 3. Incidents by data type—percentage of breaches and record types affected.

After the Breach: Honda

In 2011 Honda was hit with a class action lawsuit after a breach where attackers targeted the Honda Canada website and stole 283,000 records containing customer names, addresses, vehicle identification numbers, and related personal data. The suit charges that Honda did not inform customers of the breach in a reasonable time. This breach is similar to one that occurred at Honda America in December 2010 when accounts for 4.9 million customers of Honda and Acura were exposed.

(Source: http://mcaf.ee/n3et5)

The top three data types stack ranked by percentage of breaches was very intuitive and included:

- Names and/or addresses were involved in 83 percent of the breaches and 90 percent of the records affected
- Social Security numbers and equivalents were involved in 51 percent of the breaches and 46 percent of the records affected
- Credit card numbers were involved in 32 percent of the breaches and 8 percent of the records affected

The least common data types compromised included: medical information, financial information, and email addresses.

Data types by vertical

The following table illustrates verticals juxtaposed with the top three data types most common to their incidents.

Data Types Affected by Vertical

	1	2	3
Retailers	Credit card numbers	Names and/or addresses	Email addresses
Financial services	Names and/or addresses	Social Security numbers and equivalents	Account information
Technology companies	Names and/or addresses	Credit card numbers	Miscellaneous
US Federal Government	Names and/or addresses	Social Security numbers and equivalents	Miscellaneous
State Governments	Social Security numbers and equivalents	Names and/or addresses	Date of birth
Universities	Social Security numbers and equivalents	Names and/or addresses	Date of birth
K–12 schools	Names and/or addresses	Social Security numbers and equivalents	Date of birth
Healthcare providers	Names and/or addresses	Medical	Social Security numbers and equivalents
Insurance companies	Names and/or addresses	Social Security numbers and equivalents	Date of birth
Industry	Names and/or addresses	Social Security numbers and equivalents	Date of birth Account information Financial information

Interestingly enough, names and/or addresses were the most common data type compromised in eight of the 10 verticals, and in the other two instances, they were the second most common. Only retailers, leading with credit card numbers, and universities, leading with Social Security numbers or equivalents, deviated.

Conclusion

Nobody, consumers or organizations, wants to suffer though the damage that data loss causes. But it's clear that the "bad guys" are not backing off when it comes to stealing sensitive data. Scammers, fraudsters, malicious insiders, hackers, and the like are finding more creative ways to get this information. Ground zero for these attacks: data centers. Regardless of business vertical, most of the battles to protect sensitive data will occur in and around the data center throughout physical, virtual, and cloud environments.

We hope that by understanding some of the practical statistics we've provided around data loss, you'll be able to leverage this information to assist in your data center security efforts. Regardless of external or internal attacks, malicious or careless activity, having comprehensive security controls around servers, storage, content and networks will help mitigate the risk.

We are working diligently to help our customers achieve effective and efficient security controls for their data centers. We offer a wide range of solutions for servers, storage, content and networks as well as specialized solutions for virtualized environments and security SaaS offerings. To learn more about how McAfee and its partners can secure your data centers please visit: http://mcaf.ee/7c3za.

About McAfee

McAfee, a wholly owned subsidiary of Intel Corporation (NASDAQ:INTC), is the world's largest dedicated security technology company. McAfee delivers proactive and proven solutions and services that help secure systems, networks, and mobile devices around the world, allowing users to safely connect to the Internet, browse, and shop the web more securely. Backed by its unrivaled global threat intelligence, McAfee creates innovative products that empower home users, businesses, the public sector, and service providers by enabling them to prove compliance with regulations, protect data, prevent disruptions, identify vulnerabilities, and continuously monitor and improve their security. McAfee is relentlessly focused on constantly finding new ways to keep our customers safe. http://www.mcafee.com

McAfee
2821 Mission College Boulevard
Santa Clara, CA 95054
888 847 8766
www.mcafee.com

31600wp_data-loss-numbers_0811_fnl_ETMG