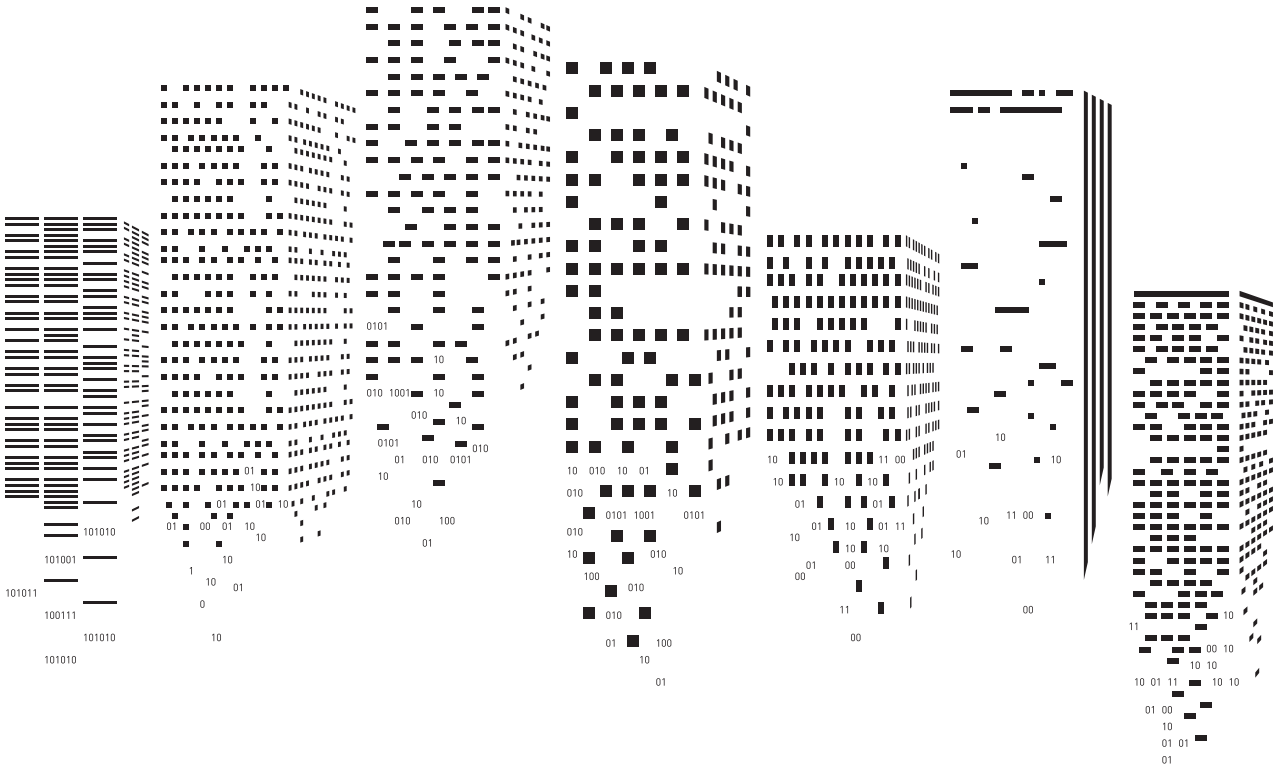


Follow the Data: Analyzing Breaches by Industry

Trend Micro Analysis of Privacy Rights Clearinghouse
2005–2015 Data Breach Records

Numaan Huq
Forward-Looking Threat Research (FTR) Team

A **TrendLabs**SM Research Paper

Contents

Our data source

The Privacy Rights Clearinghouse (PRC) is a nonprofit corporation based in California. PRC's mission is to engage, educate, and empower individuals to protect their privacy[1]. They do this by raising consumers' awareness of how technology affects personal privacy, and they empower consumers to take actions to control their personal information by providing practical tips on privacy protection. PRC responds to privacy-related complaints from consumers and where appropriate intercedes on the consumer's behalf/refers them to the proper organizations for further assistance. PRC documents consumers' complaints and questions about privacy in reports and makes them available to policy makers, industry representatives, consumer advocates, media, etc. PRC advocates consumers' privacy rights in local, state, and federal public policy proceedings.

PRC publishes the "Chronology of Data Breaches Security Breaches 2005–Present[1]," which is a collection of publicly disclosed data breach incident reports that occurred in the United States. The data is compiled from a variety of sources including: media, Attorney General's Office press releases, company press releases, privacy websites, etc.

All data breach incident reports in this paper have been collected from the PRC database for the period, January 2005–April 2015[2]. PRC's original "Organization Types" were expanded to include a wide range of industries in order to provide a fine-grained view into victim profiles. Each entry was analyzed to determine the record types compromised. The data collected was analyzed using tools that include KH Coder[3], MSBNx[4], and Explore Analytics[5].

Further analysis of publicly disclosed data breach incidents

This material supplements the paper, "Follow the Data: Dissecting Data Breaches and Debunking Myths," where we looked at the volume of incidents over a span of 10 years and formulated a Bayesian model in order to determine the probability that different breach scenarios would occur. We also took a closer look at where stolen data eventually ended up and what interested parties could do with it.

Here, we will take a closer look at the different breach methods, record-type combinations stolen, and industry cross-sections of the PRC data.

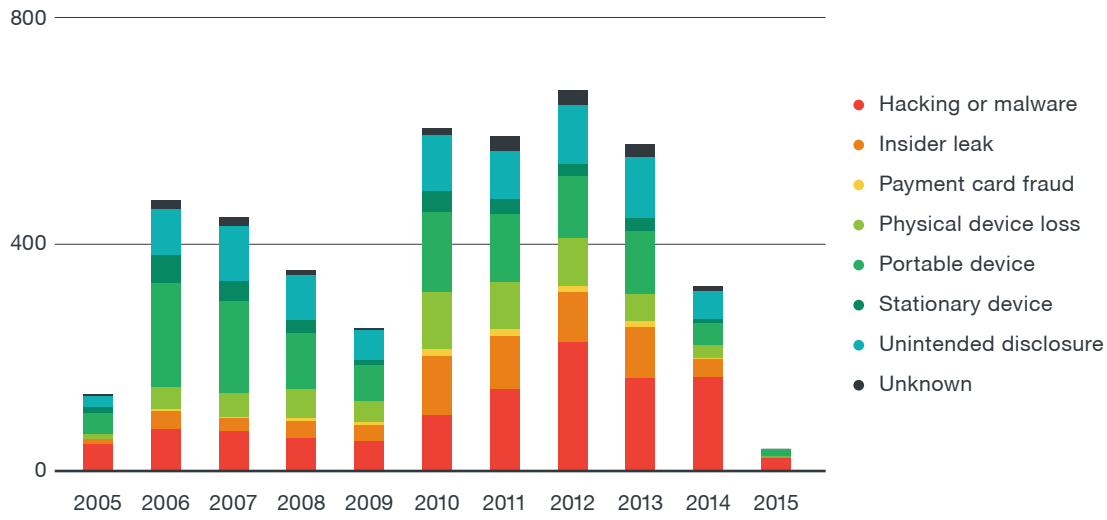


Figure 1: Breach methods observed from 2005 to April 2015

Figure 1 shows an interesting visualization that emerged when breach methods were plotted on a timescale.

- Portable device (USB keys, portable drives, laptops, etc.) loss or theft is a major contributing factor in the leakage of sensitive data. This remains a constant threat over the years.

- The number of hacking or malware attacks steadily rose from 2010 onward. As the Internet expands and new applications are introduced, businesses are steadily growing their online presence, leading to an increase in hacking or malware attacks against them.
- A big increase in the number of insider threats was seen from 2010 onward. There are two plausible explanations for this. First, insider threats have always been present and not properly reported. Second, insiders now find it monetarily lucrative to steal and sell sensitive data, and so, commit more crimes.

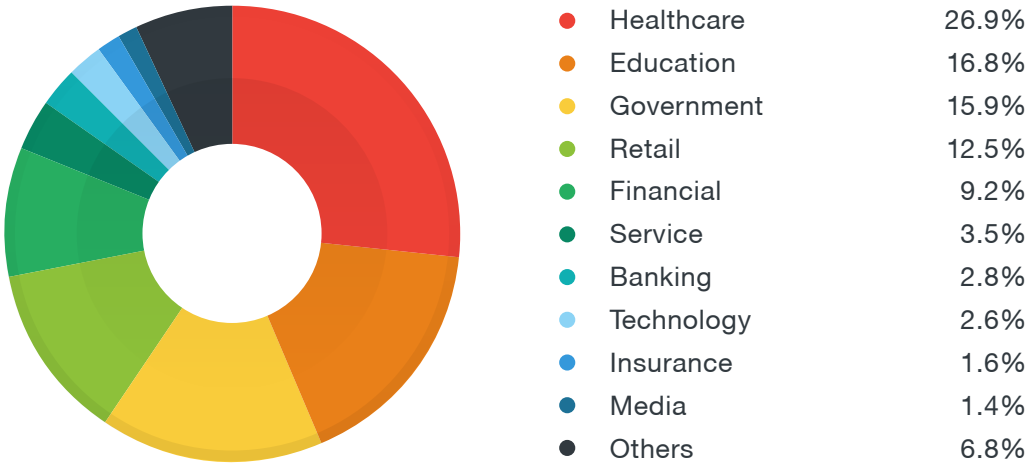


Figure 2: Industries affected by data breaches

Any business or organization that processes and/or stores sensitive data is a potential breach target.

Figure 2 shows that the healthcare, education, government, retail, and financial industries were the most frequent victims of data breach crimes. The PRC data shows that these five industries accounted for 81.3% of the total number of disclosed data breach incidents. This is not really surprising, as businesses or organizations that belong to these industries are a treasure trove of valuable sensitive data and so would provide good returns on investment (ROIs) for the cybercriminals that compromise them.

All of the aforementioned industries are subject to specialized state and federal data breach disclosure laws. That's why we see them dominate incident reports. Other industries are also targeted but incident reports for them are few likely because of:

- Failure to identify data breaches
- Failure to properly assess data breaches
- Noncompliance with data breach notification laws

- Lack of knowledge of data breach notification laws
- Nondisclosure because of active investigation by authorities
- Breaches did not lead to the compromise of data that legally mandates public disclosure

Often, organizations are not even aware of ongoing data breaches and require external parties (banks, law enforcement agencies, customers, etc.) to inform them. Data breach news stories are regularly leaked to the media. Media practitioners also have ongoing investigations or learn about breaches from insiders. Large data breaches are typically followed by class-action lawsuits, which damage businesses or organizations and may even deter them from immediately disclosing incidents.

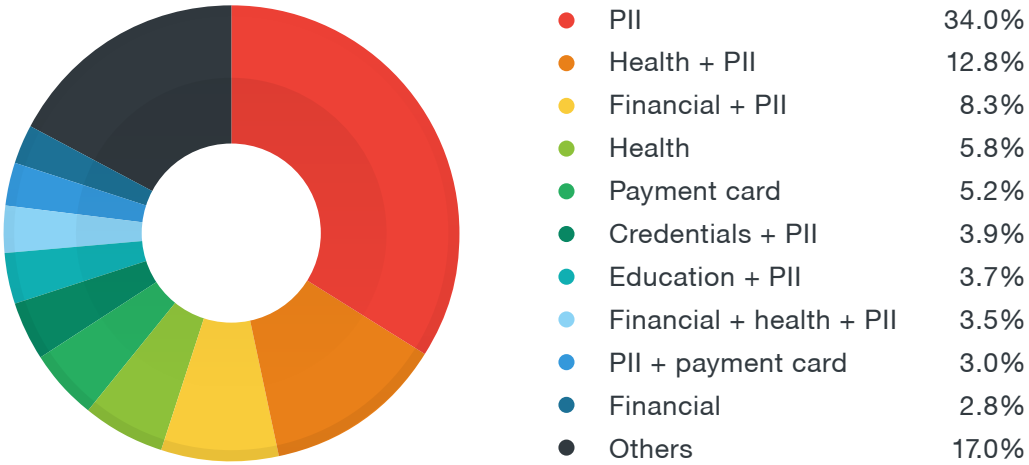


Figure 3: Record-type combinations compromised

Figure 3 shows that a wide range of data gets stolen from businesses (big and small) and individuals, including:

- **Personally identifiable information (PII):** Names, addresses, Social Security numbers, and other information.
- **Financial data:** Banking, insurance, billing, and other information.
- **Health data:** Hospitals and doctors' office records, medical insurance, and other information.
- **Education data:** School, college, university, or related records.
- **Payment card data:** Credit, debit, store-branded credit, and prepaid gift cards.
- **Credentials:** Log-in credentials to eBay, PayPal, Web-based emails, online banking sites, and others.
- **Others:** Intellectual property, intelligence about an organization, and other information.
- **Unknown:** In many cases, investigators failed to determine what was stolen.

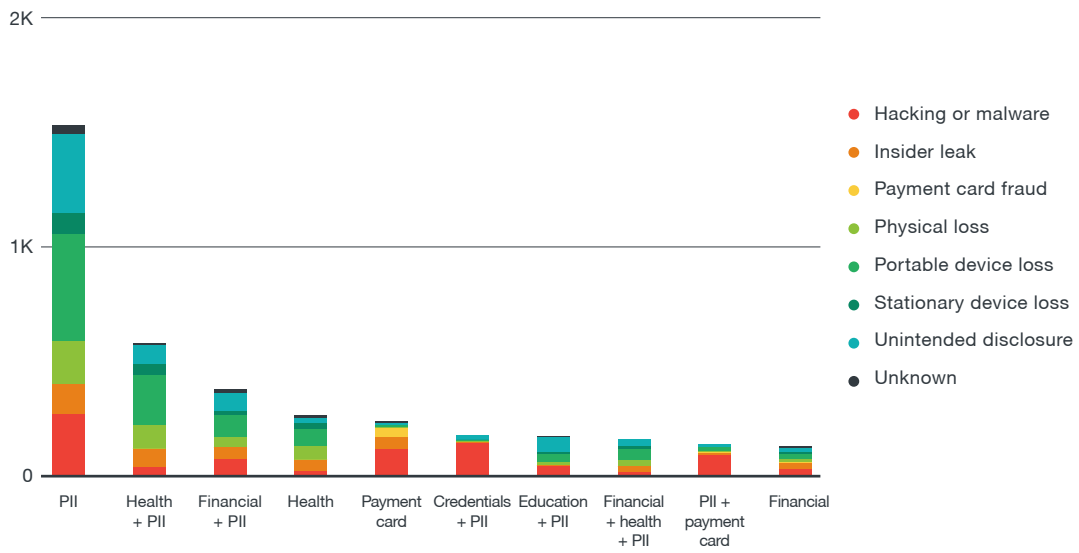


Figure 4: Top 10 record-type combinations compromised versus breach methods used

Figure 4 shows that the most popular record types stolen were PII; health, financial, education, and payment card data; and credentials. These are easy to monetize and thus make lucrative targets. Health and education data is mostly stolen because they contain PII. Data presented later will support this claim.

PII was most commonly compromised. The usual breach methods for PII theft were unintended disclosure, loss or theft, insiders, and hacking or malware. No breach method could be singled out as the major contributing factor for compromising PII; all available means were instead employed.

Company name	Records compromised	Disclosure date	Record types compromised
eBay	145M	21 May 2014	Credentials, PII
Heartland Payment Systems	130M	20 January 2009	Payment card data
Target	110M	13 December 2013	PII (70M), payment card data (40M)
The Home Depot	109M	2 September 2014	PII (53M), payment card data (56M)
Sony PlayStation Network	101.6M	27 April 2011	Credentials, financial data, PII, payment card data
TJ stores (TJX)	100M	17 January 2007	PII, payment card data
Anthem	80M	5 February 2015	Financial data, PII
JP Morgan Chase	76M	28 August 2014	Credentials, financial data, PII
Evernote	50M	3 March 2013	Credentials
Epsilon	50M	2 April 2011	PII
CardSystems	40M	16 June 2005	Payment card data
Adobe	38M	4 October 2013	Credentials, PII, payment card data
Steam (The Valve Corporation)	35M	10 November 2011	Credentials, PII, payment card data
RockYou	32M	15 December 2009	Credentials
LivingSocial	29M	26 April 2013	Credentials, PII
Zappos.com	24M	15 January 2012	Credentials, PII
WordPress	18M	14 April 2011	Credentials, other data
Countrywide Financial Corp.	17M	2 August 2008	Financial data, PII
DeviantArt	13M	17 December 2010	PII
Bank of New York Mellon	12.5M	26 March 2008	Financial data, PII

Figure 5: Top 20 publicly disclosed data breach incidents as of April 2015

The total number of records compromised is typically used to quantify the size and extent of a data breach incident. Figure 5 shows that the biggest incidents were not concentrated in the past five years but spread out across 2005 to April 2015.

The biggest payment card breach incident was not the one that involved Target or Home Depot but that which affected Heartland Payment Systems in 2009. Albert Gonzalez was charged by the authorities for this breach and sent to prison for 20 years[6]. The biggest credential leak involved eBay in 2014. Hackers compromised its employees' log-in credentials, allowing them access to the corporate network and to compromise the main databases that held all user passwords. The criminals also gained access to PII like names, email addresses, home addresses, phone numbers, and dates of birth[7]. The biggest data breach to date involved Anthem where 80 million records were compromised and up to 18.8 million people were affected. PII and employment information like income data was stolen[8].

Massive-scale data breaches are not that common, as they entail careful planning and prolonged effort on the criminals' part in order to succeed. Criminals tend to get better ROIs by targeting smaller businesses and organizations, which have weaker defenses and so are easier to penetrate and compromise.

Analysis of top 5 industries affected by data breaches

The healthcare, education, government, retail, and financial industries are frequent data breach victims. We studied five data sets for each industry and looked at trending patterns.

- **Data breach incident disclosures:** In this data set, we looked at incident numbers broken down by year reported. Only a fraction of all of the incidents was actually reported. An increase in the number of reported incidents strongly indicates that the total volume of data breaches has increased and vice versa.
- **Data breach methods:** In this data set, we studied the breach methods used. Data breach incidents could be due to hacking or malware attacks; insider threats; payment card fraud; physical loss or theft of portable drives, laptops, office computers, files, and others; or unintended disclosures. In a small number of cases, the actual breach method remains unknown or undisclosed.
- **Top 3 data breach methods observed and trends from 2005 to 2014:** In this data set, we looked at the incident report numbers for the top 3 data breach methods broken down by year reported. A logarithmic trend line was calculated for each breach method.
- **Record types compromised:** Criminals tend to steal all kinds of available data. Examples include PII stored as part of health and education records. In some cases, financial data like billing, insurance, and other information is stored along with health records. As such, these get stolen together. In this data set, we looked at the record types and record-type combinations compromised.
- **Probability of compromising different record types for each breach method:** In this data set, we looked at the probability of different record types getting compromised for each breach method.

Data breaches in the healthcare industry

The number of data breach incidents increased from 2010 onward. The "Health Insurance Portability and Accountability Act (HIPAA)," which outlines patient data privacy requirements, has been around since 1996 and mandates that organizations in the healthcare industry disclose all data breaches. The upward trend in incident report numbers suggests that the healthcare industry has recently become a lucrative target.

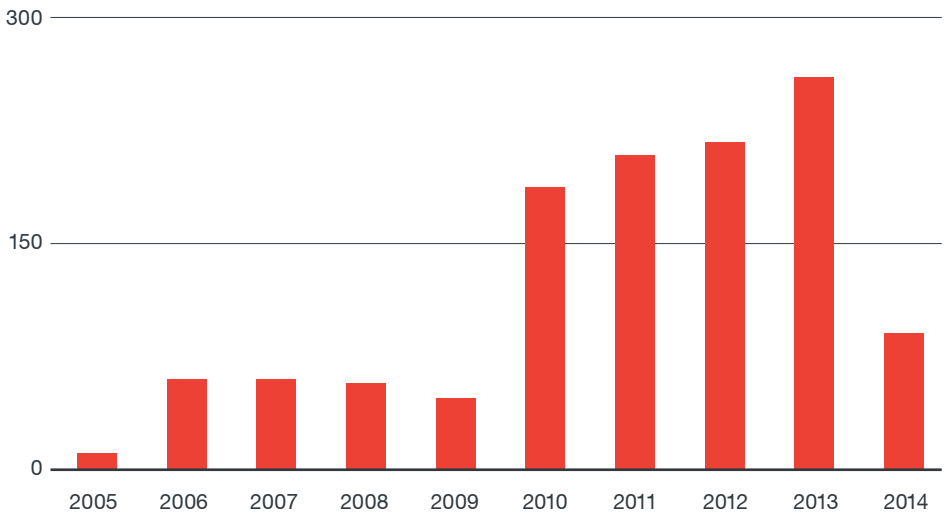


Figure 6: Healthcare data breach incident disclosures from 2005 to 2014

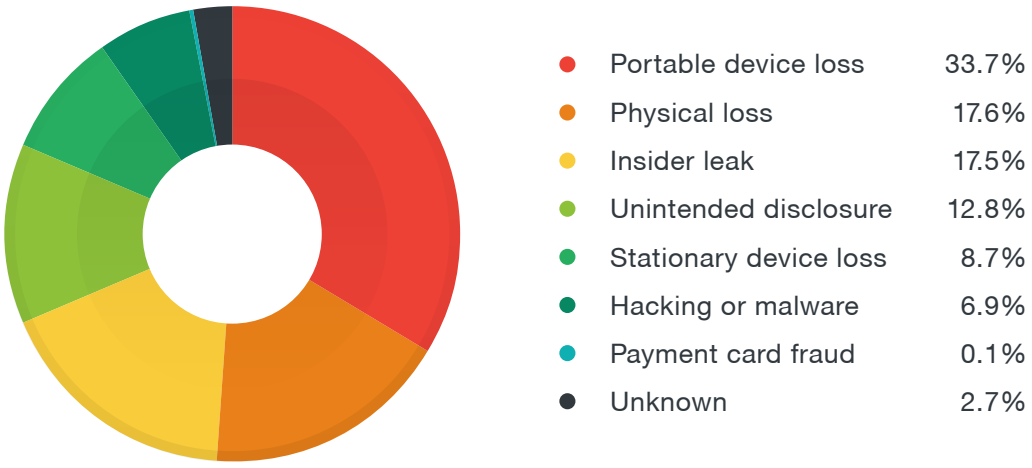


Figure 7: Breach methods observed in the healthcare industry

The loss or theft of portable devices, backup drives, files, laptops, office computers, and other devices accounted for almost two-thirds of all breaches. Surprisingly, hacking or malware attacks accounted for less than 10% of all of the breaches.

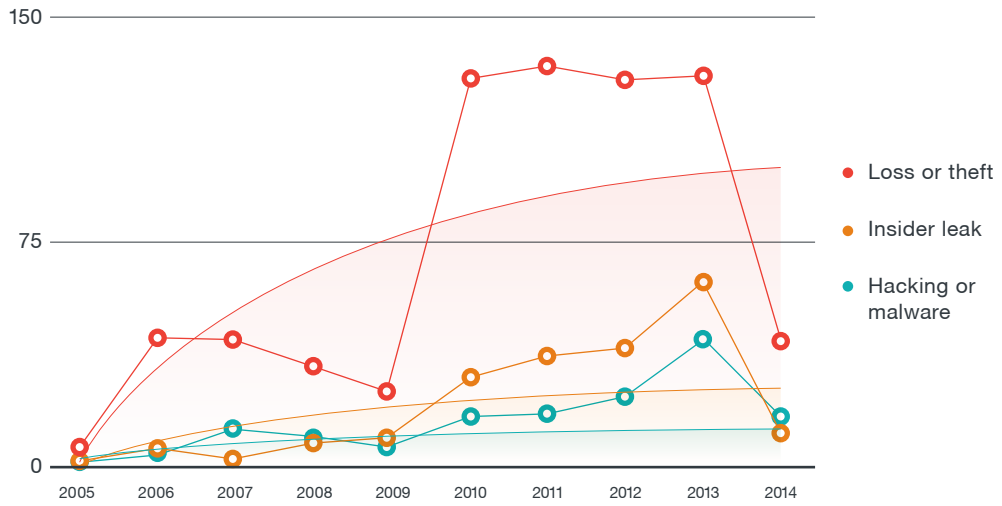


Figure 8: Top 3 breach methods observed in the healthcare industry from 2005 to 2014

The top 3 breach methods in the healthcare industry were loss or theft, insider leaks, and unintended disclosures. From the reported numbers, we observed a marked increase in loss or theft incidents from 2010 onward. A possible explanation for this increase, apart from more theft, was better incident reporting. The trend line for each breach method shows an upward trend with loss or theft growing more than both insider threats and unintended disclosures.

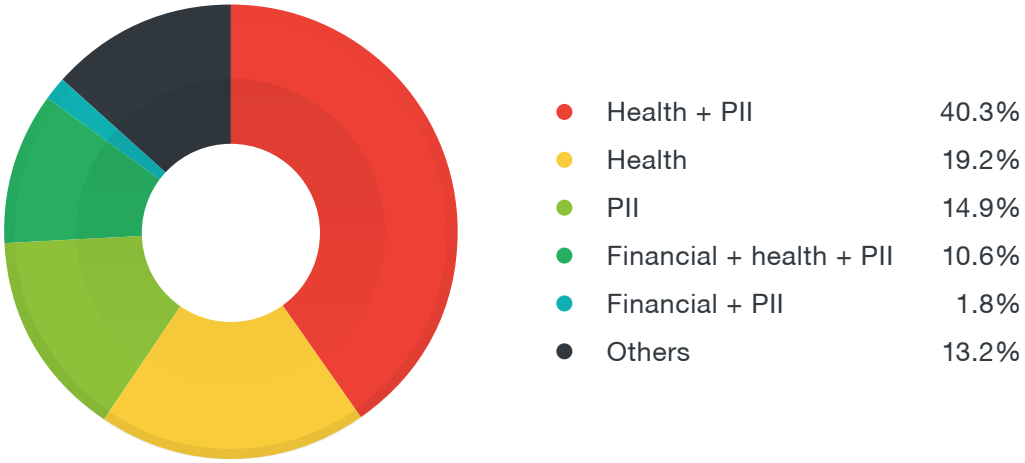


Figure 9: Record types compromised in the healthcare industry

The most popular record types and record-type combinations compromised were health and PII, PII, health data, and PII and financial and health data. In some cases, patient files contained billing and insurance information, leading to their theft.

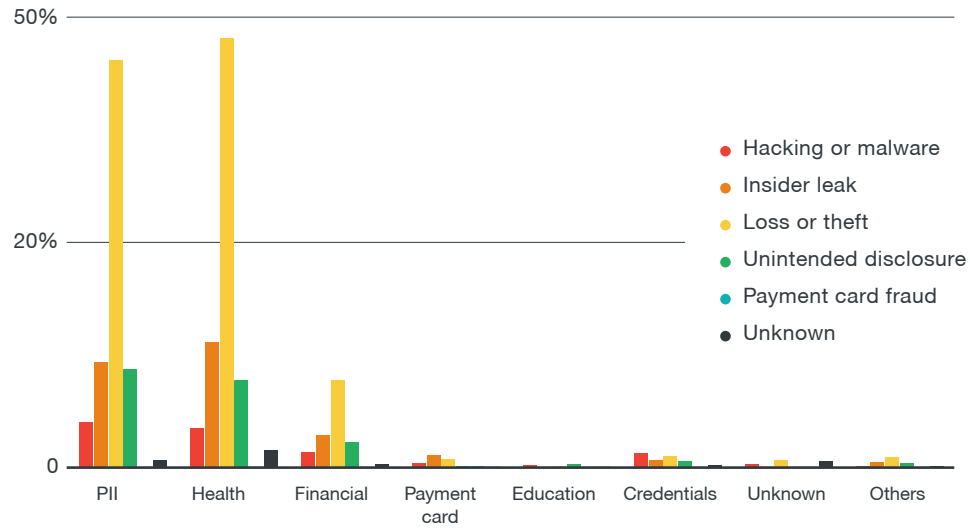


Figure 10: Probability of compromising different record types per breach method in the healthcare industry

PII and health data had the highest probability of getting compromised. There is a small chance that financial data like billing, insurance, and other information will also get compromised. The most probable breach method for stealing PII and health data is loss or theft.

Data breaches in the government sector

Data breaches follow a pattern that starts with a big increase in incident numbers reported in one year, followed by several years of decline. Every time there is an increase in data breach incidents, new policies, protocols, and procedures are most likely implemented, driving the incident numbers down.

Figure 11: Government sector data breach incident disclosures from 2005 to 2014

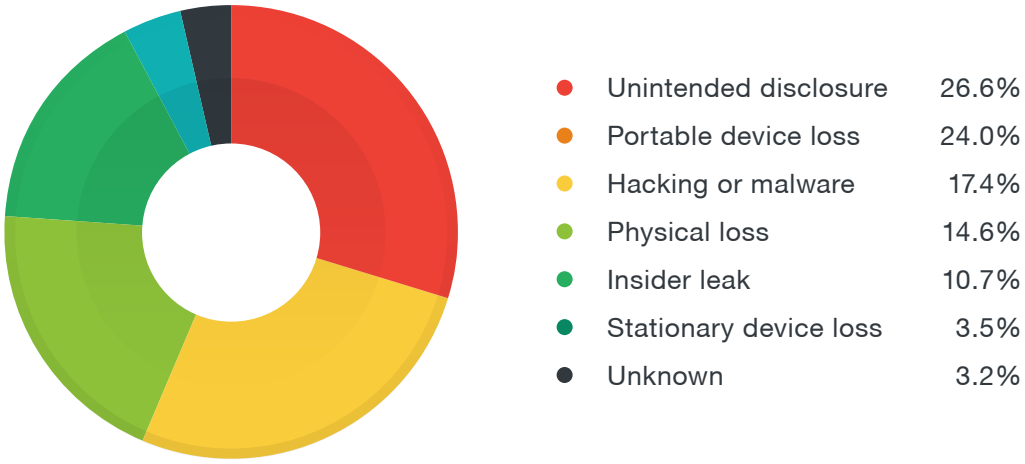


Figure 12: Breach methods observed in the government sector

Across government organizations, loss or theft of portable devices, backup drives, and others was the biggest contributing factor in data breaches. Unintended disclosure of sensitive data through mistakes or negligence is another major problem.

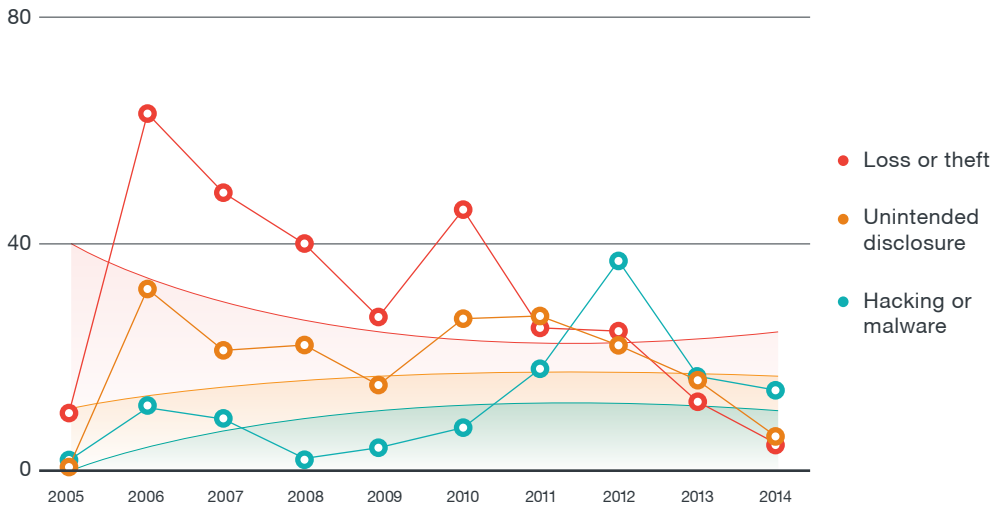


Figure 13: Top 3 breach methods observed in the government sector from 2005 to 2014

The top 3 breach methods in the government sector were loss or theft, unintended disclosures, and hacking or malware attacks. Based on the reported incident numbers, there was a gradual decline in loss or theft incidents. This suggests the presence of strong policies, protocols, and procedures that help reduce such incidents. The trend line for unintended disclosures shows that this threat remained fairly consistent over the years whereas hacking or malware attacks increased.

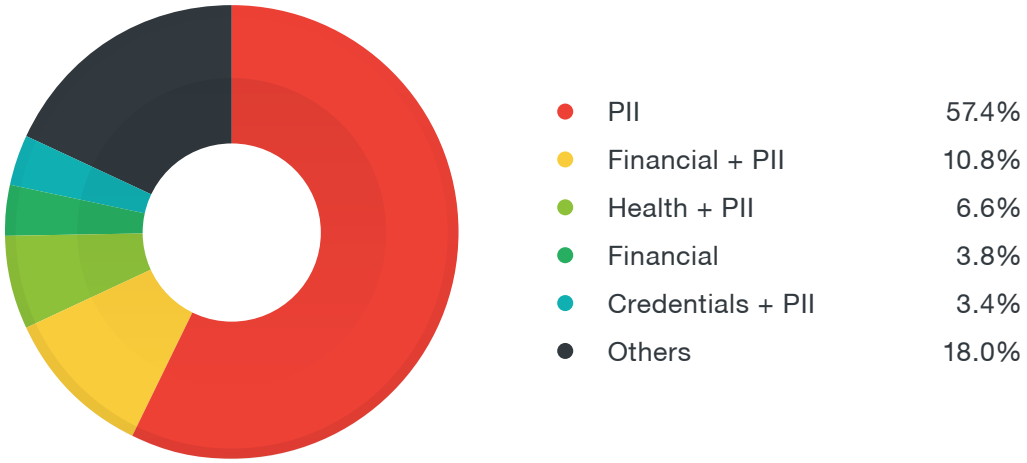


Figure 14: Record types compromised in the government sector

PII theft dominated the data breach incidents in the government sector.

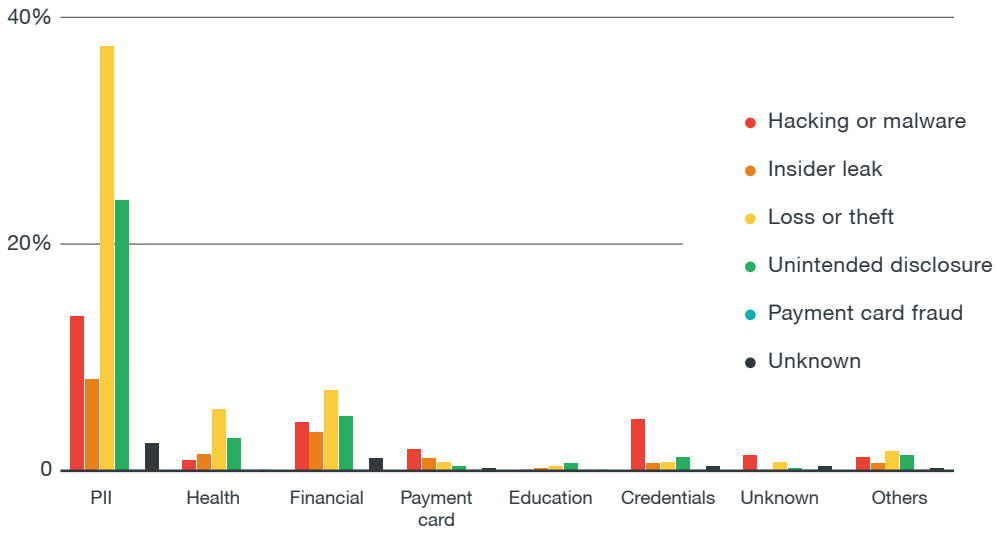


Figure 15: Probability of compromising different record types per breach method in the government sector

Among breaches that targeted government organizations, PII had the highest probability of getting compromised. There is a small chance that financial and health data will also be compromised. The most probable breach methods for stealing PII were loss or theft, unintended disclosures, hacking or malware attacks, and insider threats.

Data breaches in the retail industry

Data breaches have become commonplace in the retail industry after the development of PoS RAM scrapers sometime between 2007 and 2008[9,10] The upward trend observed in the incident report numbers reflected this growing threat.

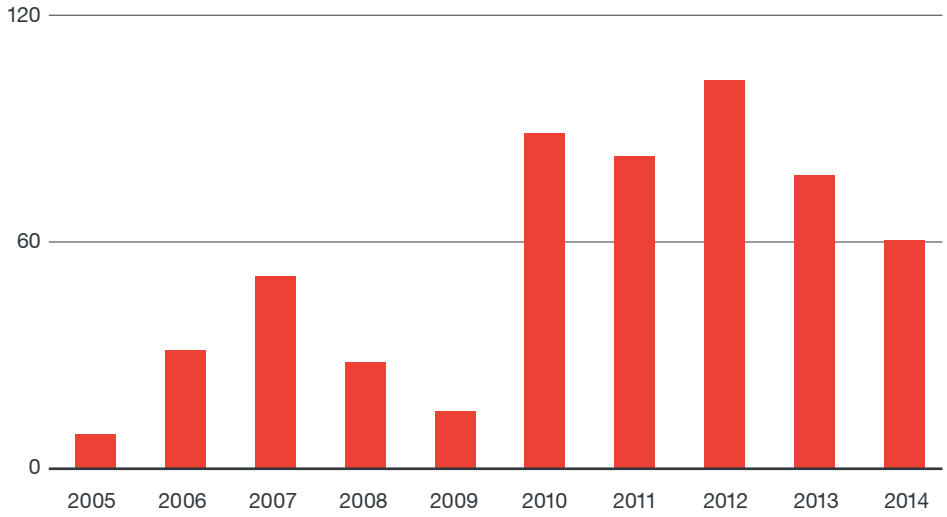


Figure 16: Retail industry data breach incident disclosures from 2005 to 2014

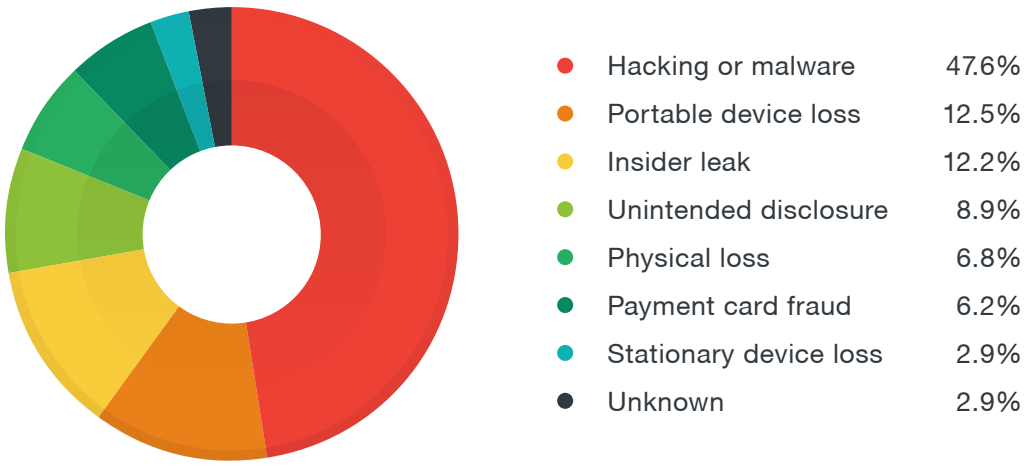


Figure 17: Breach methods observed in the retail industry

Hacking or malware attacks were common in the retail industry. PoS RAM scrapers were used to collect payment card data while a variety of infiltration techniques were employed to gain initial entry into and laterally move across target networks in order to compromise PoS servers. Insider threats mostly refer to employees who use skimming devices to steal credit card data.

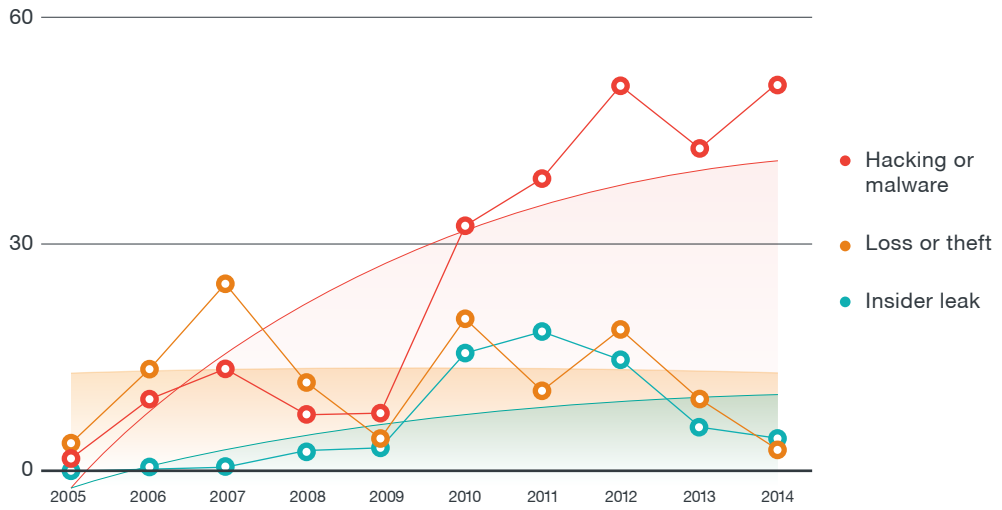


Figure 18: Top 3 breach methods observed in the retail industry from 2005 to 2014

The top 3 breach methods used in the retail industry were hacking or malware attacks, loss or theft, and insider threats. The reported incident numbers show a big increase in hacking or malware attacks, which could be attributed to PoS RAM scrapers. Insider threats showed an upward trend while the trend line for loss or theft remained consistent over the years.

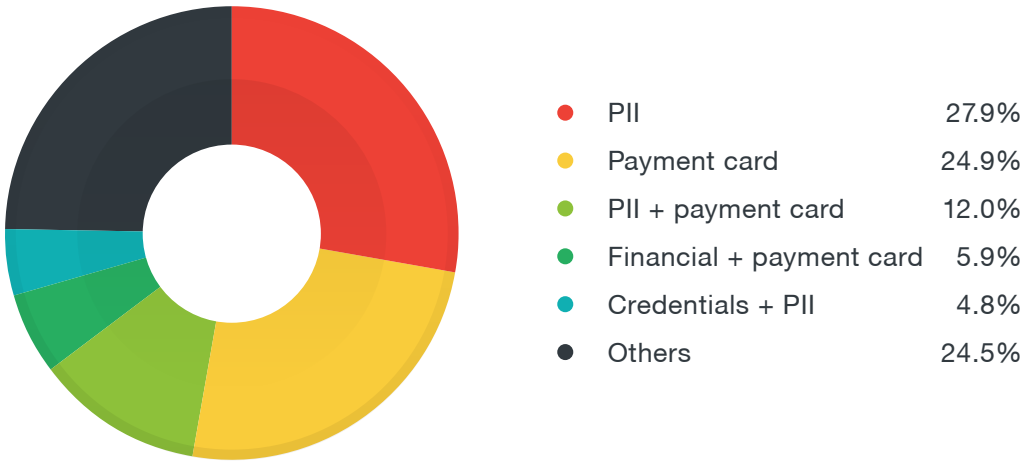


Figure 19: Record types compromised in the retail industry

The retail industry is a hotbed for the theft of payment card data, financial information, and PII and payment card data.

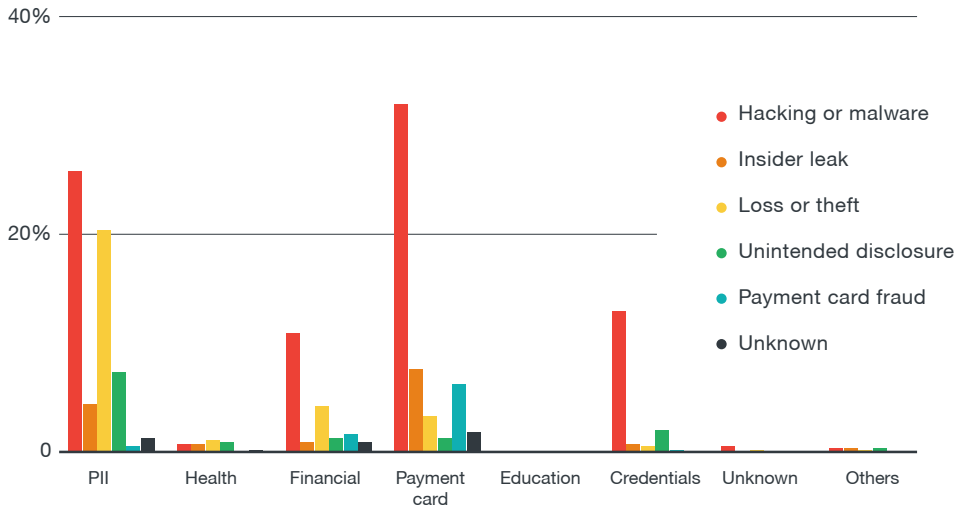


Figure 20: Probability of compromising different record types per breach method in the retail industry

Payment card data and PII had the highest probability of getting compromised. The most probable method for stealing payment card data was hacking or malware attacks. For PII, the most probable breach methods were hacking or malware attacks and loss or theft.

Data breaches in the financial industry

Similar to the pattern observed in the government sector, the financial industry posted a big increase in incident numbers reported in one year, followed by several years of decline. It is probable that every time the number of incidents increases, new policies, protocols, and procedures were implemented, which drove the trend down.

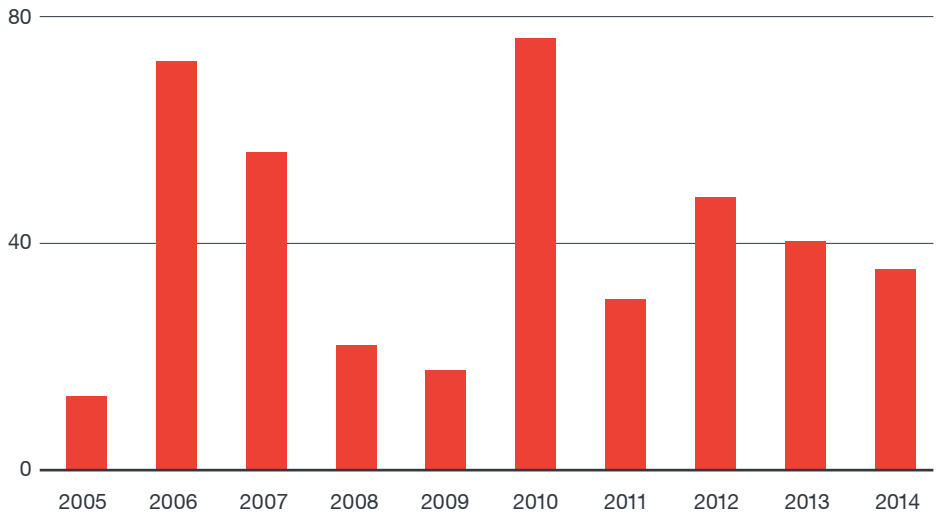


Figure 21: Financial industry data breach incident disclosures from 2005 to 2014

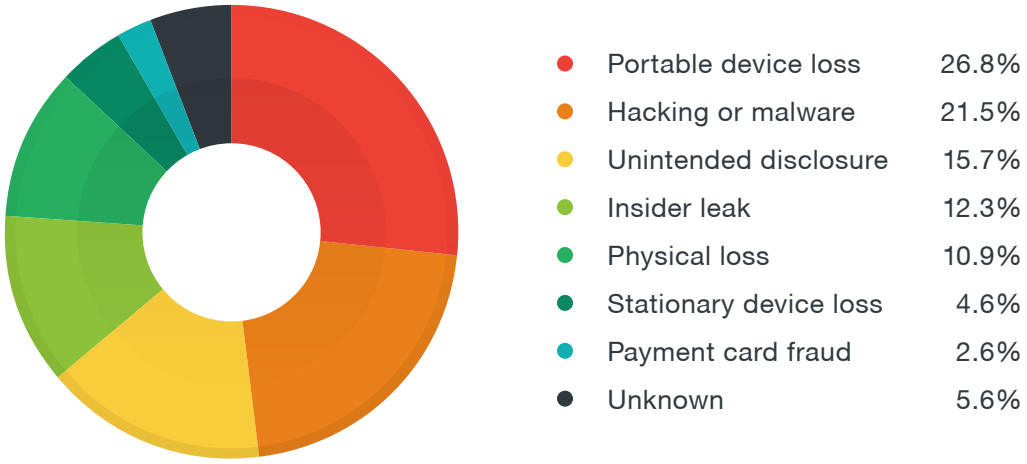


Figure 22: Breach methods observed in the financial industry

There was a fairly even distribution of incidents involving loss or theft, hacking or malware, insider threats, and unintended disclosures in the financial industry. Perpetrators employ any means to compromise financial data instead of relying on one or two proven breach methods.

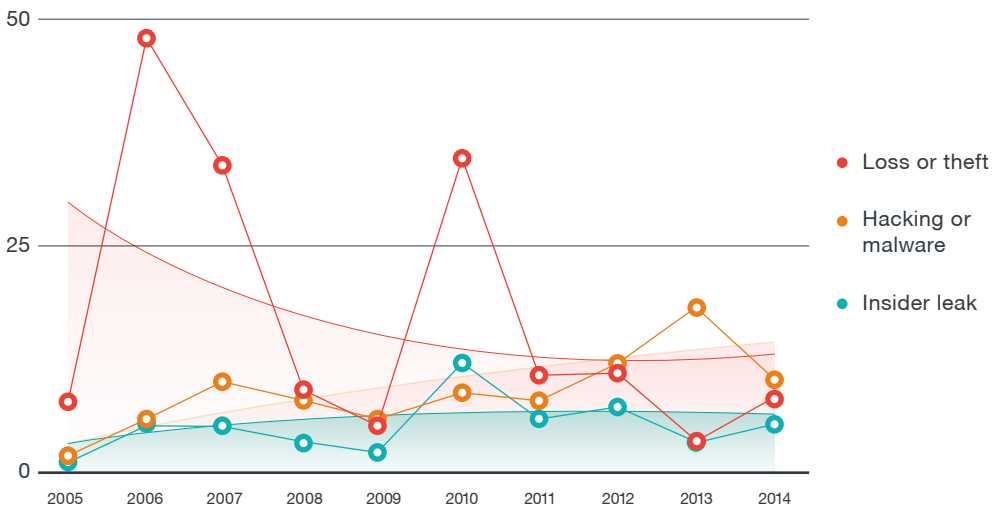


Figure 23: Top 3 breach methods observed in the financial industry from 2005 to 2014

The top 3 breach methods used in the financial industry were loss or theft, hacking or malware attacks, and insider threats. The reported incident numbers show a decline in loss or theft, which suggests that strong policies, protocols, and procedures may have been put in place, helping reduce such incidents. Hacking or malware attacks and insider threats increased.

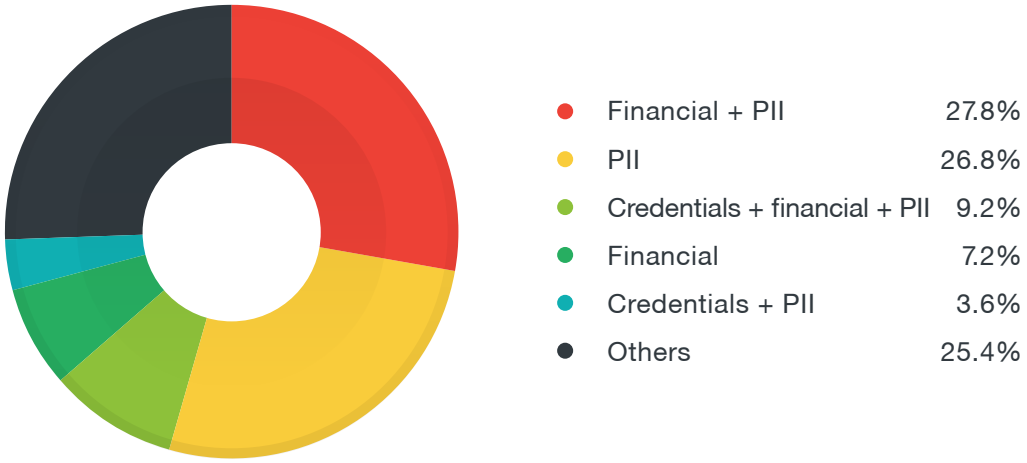


Figure 24: Record types compromised in the financial industry

The most popular record types and record-type combinations compromised were financial data and PII; PII; credentials, financial data, and PII; credentials and PII; and financial data. The credentials stolen usually included online banking log-in details.

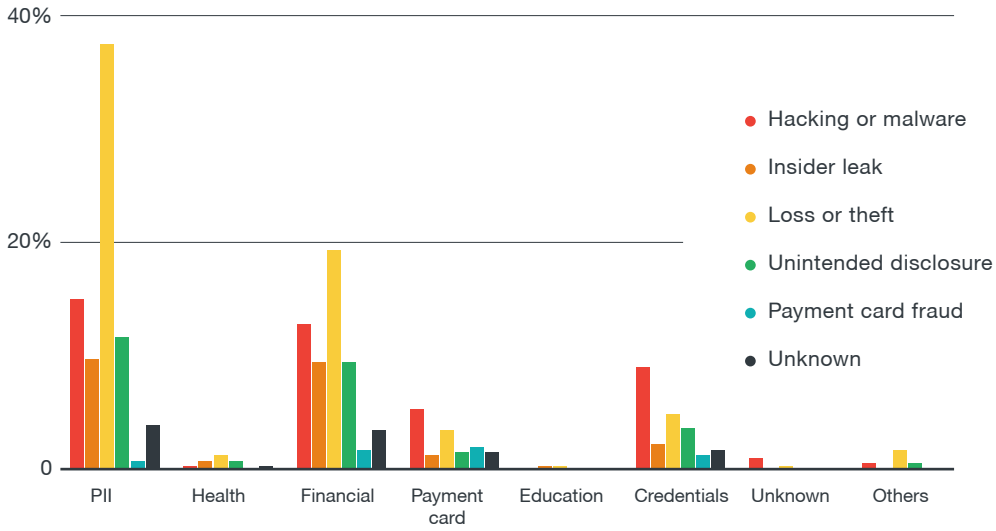


Figure 25: Probability of compromising different record types per breach method in the financial industry

PII and financial data had the highest probability of getting compromised. The most likely data breach methods were loss or theft and hacking or malware attacks. There is a small chance that credentials and payment card data will also get compromised.

Data breaches in the education sector

The data breach incident numbers in the education sector have been declining over the years. This could possibly be due to a shift in focus to more lucrative targets like the healthcare and retail industries.

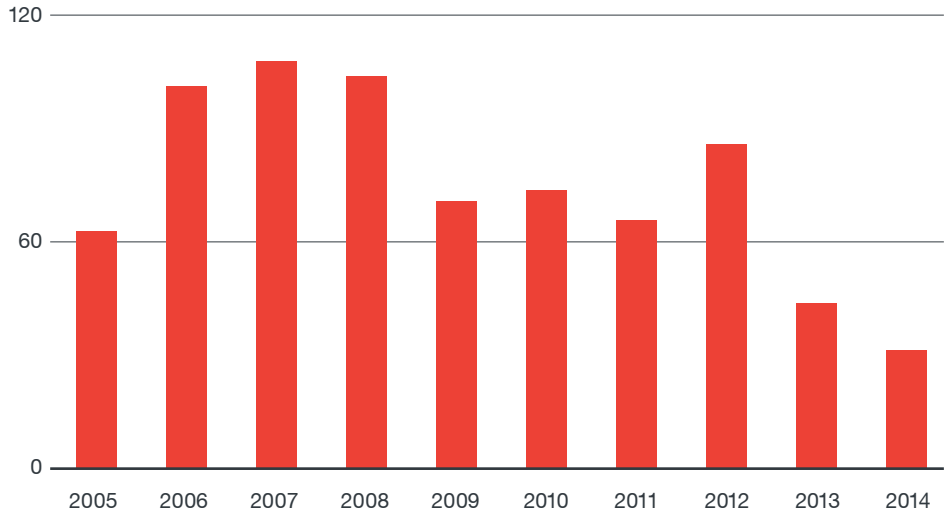


Figure 26: Education sector data breach incident disclosures from 2005 to 2014

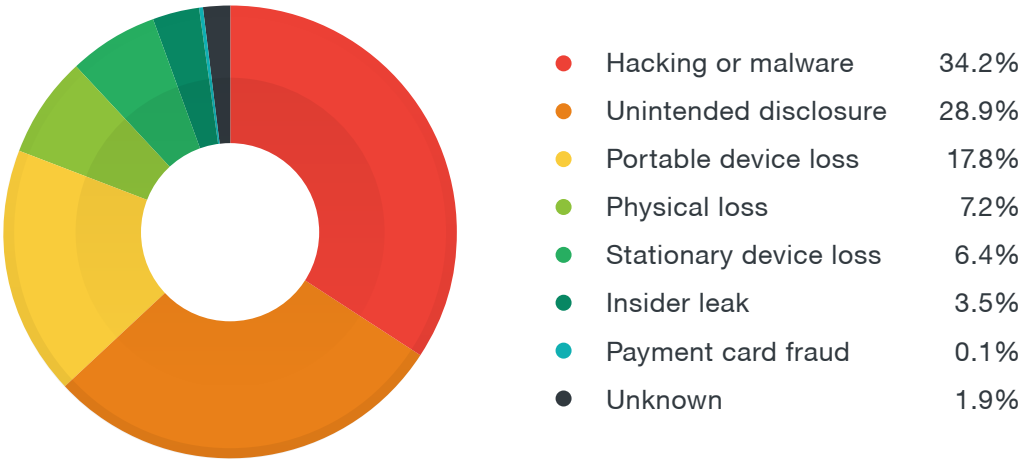


Figure 27: Breach methods observed in the education sector

Hacking or malware attacks, unintended disclosures, and loss or theft accounted for 94.5% of all of the breaches observed in the education sector.

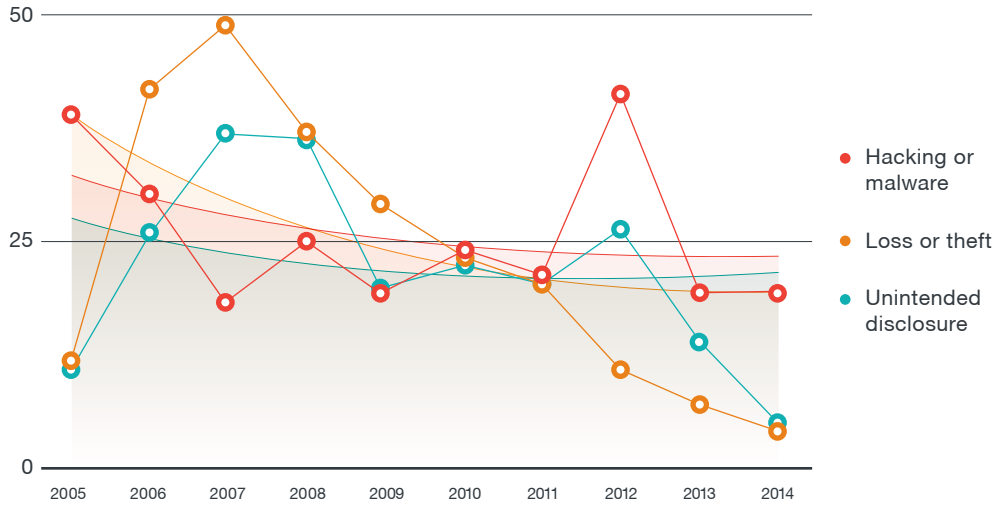


Figure 28: Top 3 breach methods observed in the education sector from 2005 to 2014

The top 3 breach methods seen were hacking or malware attacks, loss or theft, and unintended disclosures. The trend lines for all three methods show a downward pattern.

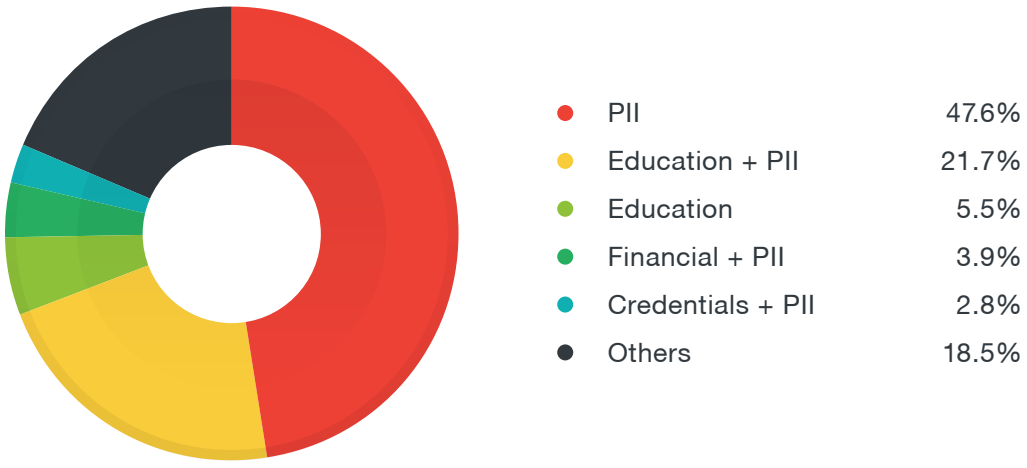


Figure 29: Record types compromised in the education sector

The most popular record types and record-type combinations compromised were PII as well as education and PII. Financial data was also compromised in some incidents.

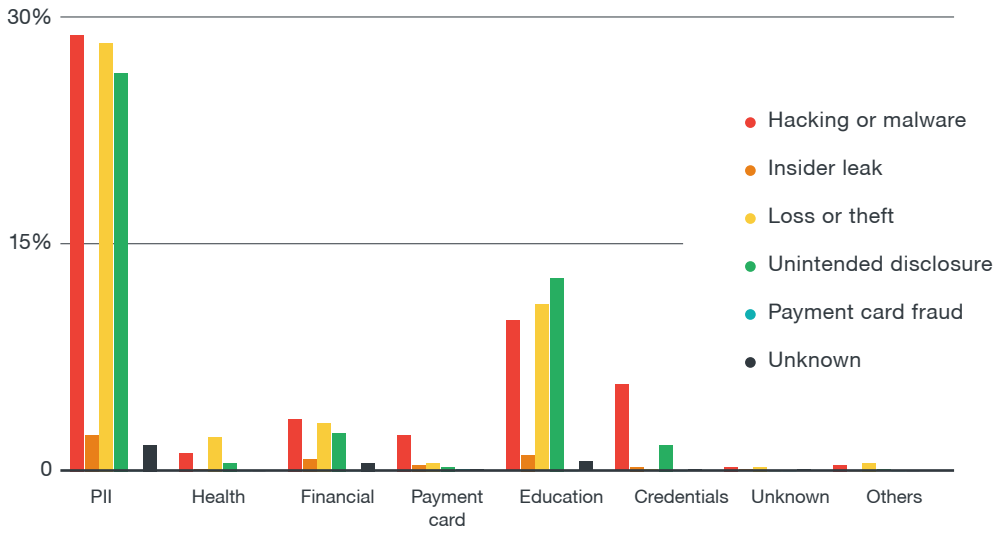


Figure 30: Probability of compromising different record types per breach method in the education sector

Education and PII posted the highest probability of getting compromised via methods such as hacking or malware attacks, loss or theft, and unintended disclosures.

References

1. Privacy Rights Clearinghouse. (2015). *About the Privacy Rights Clearinghouse*. Last accessed on 3 July 2015, https://www.privacyrights.org/content/about-privacy-rights-clearinghouse.

2. Privacy Rights Clearinghouse. (2015). "Chronology of Data Breaches Security Breaches 2005–Present." Last accessed on 23 June 2015, https://www.privacyrights.org/data-breach.

3. KH Coder [Computer Software]. (2015). Retrieved from http://khc.sourceforge.net/en/.

4. Microsoft Bayesian Network Editor [Computer Software]. (2010). Retrieved from http://research.microsoft.com/en-us/um/redmond/groups/adapt/msbnx/.

5. Explore Analytics [Online Software]. (2015). Retrieved from https://www.exploreanalytics.com/.

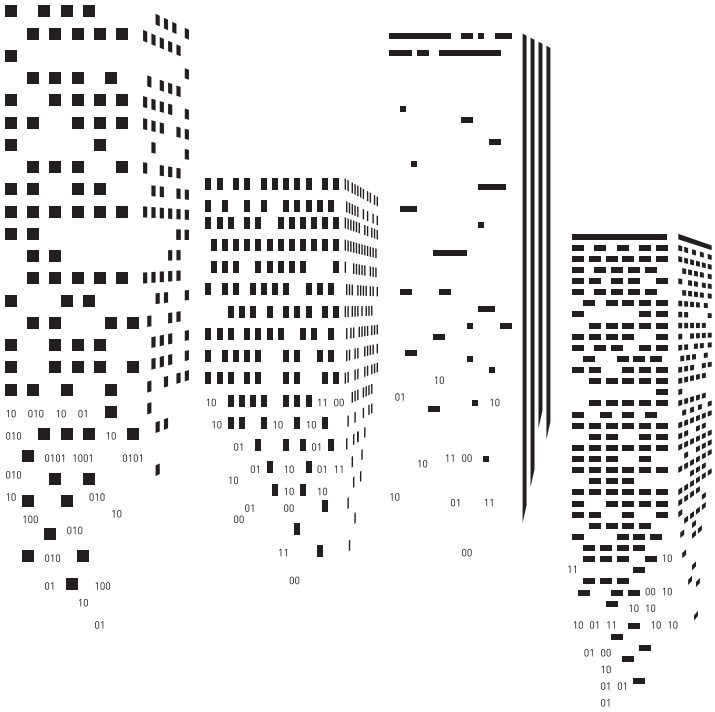
6. Kim Zetter. (29 December 2009). *Wired*. "Albert Gonzalez Pleads Guilty in Heartland, 7-11 Breaches—Updated." Last accessed on 5 July 2015, http://www.wired.com/2009/12/heartland-guilty-plea/.

7. Gordon Kelly. (21 May 2014). *Forbes*. "eBay Suffers Massive Security Breach, All Users Must Change Their Passwords." Last accessed on 5 July 2015, http://www.forbes.com/sites/gordonkelly/2014/05/21/ebay-suffers-massive-security-breach-all-users-must-their-change-passwords/.

8. Caroline Humer. (24 February 2015). *Reuters.* "Anthem Says Hack May Affect More Than 8.8 Million Other BCBS Members." Last accessed on 5 July 2015, http://www.reuters.com/article/2015/02/25/us-anthem-cybersecurity-idUSKBN0LS2CS20150225.

9. Numaan Huq. (September 2014). *Trend Micro Security Intelligence.* "PoS RAM Scraper Malware: Past, Present, and Future." Last accessed on 6 July 2015, http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-pos-ram-scraper-malware.pdf.

10. Numaan Huq. (March 2015). *Trend Micro Security Intelligence.* "Defending Against PoS RAM Scrapers: Current and Next-Generation Technologies." Last accessed on 6 July 2015, http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-defending-against-pos-ram-scrapers.pdf.



Created by:

TrendLabs

The Global Technical Support and R&D Center of **TREND MICRO**

TREND MICRO™

Trend Micro Incorporated, a global cloud security leader, creates a world safe for exchanging digital information with its Internet content security and threat management solutions for businesses and consumers. A pioneer in server security with over 20 years experience, we deliver topranked client, server, and cloud-based security that fits our customers' and partners' needs; stops new threats faster; and protects data in physical, virtualized, and cloud environments. Powered by the Trend Micro™ Smart Protection Network™ infrastructure, our industry-leading cloud-computing security technology, products and services stop threats where they emerge, on the Internet, and are supported by 1,000+ threat intelligence experts around the globe. For additional information, visit **www.trendmicro.com**.

TREND MICRO™

Securing Your Journey to the Cloud