2008 DATA BREACH INVESTIGATIONS REPORT

Four Years of Forensic Research. More than 500 Cases. One Comprehensive Report



2008 Data Breach Investigations Report

Authors

Wade H. Baker C. David Hylender J. Andrew Valentine

Contributors

Peter Tippett, M.D., Ph.D.

A. Bryan Sartin

Stan S. Kang

Christopher Novak

Members of the RISK Team

A study conducted by the Verizon Business RISK Team

TABLE OF CONTENTS

Executive Summary
Introduction
Verizon Business Investigative Response
Methodology6
A Primer on Cybercrime
Results and Analysis
Demographics
Sources of Data Breaches
Breach Size and Source
Threat Categories
Attack Difficulty
Targeted vs. Opportunistic Attacks 18
Common Attack Pathways19
Information Repositories and Channels20
Types of Data Compromised
Time Span of Data Breach Events
Data Breach Discovery Methods22
Anti-Forensics
Unknown Unknowns
Conclusions and Recommendations



2008 Data Breach Investigations Report

A study conducted by the Verizon Business RISK Team

Executive Summary

Data breaches. You've gleaned all you can from the headlines; now you have access to information directly from the investigator's casebook. The 2008 Data Breach Investigations Report draws from over 500 forensic engagements handled by the Verizon Business Investigative Response team over a four-year period. Tens of thousands of data points weave together the stories and statistics from compromise victims around the world. What valuable insights can your organization learn from them? Here is a sample of findings discussed in the report:

Who is behind data breaches?						
 73% resulted from external sources 18% were caused by insiders 39% implicated business partners 30% involved multiple parties 	In a finding that may be surprising to some, most data breaches investigated were caused by external sources. Breaches attributed to insiders, though fewer in number, were much larger than those caused by outsiders when they did occur. As a reminder of risks inherent to the extended enterprise, business partners were behind well over a third of breaches, a number that rose five-fold over the time period of the study.					

How do breaches occur?

Most breaches resulted from a combination of events rather than a single action. Some form of error often directly or indirectly contributed to a compromise. In terms of deliberate action against information systems, hacking and malcode proved to be the attack method of choice among cybercriminals. Intrusion attempts targeted the application layer more than the operating system and less than a quarter of attacks exploited vulnerabilities. Ninety percent of known vulnerabilities exploited by these attacks had patches available for at least six months prior to the breach.

62% were attributed to a significant error
59% resulted from hacking and intrusions
31% incorporated malicious code
22% exploited a vulnerability
15% were due to physical threats

What commonalities exist?

- 66% involved data the victim did not know was on the system
- 75° of breaches were not discovered by the victim
- 83% of attacks were not highly difficult
- 85% of breaches were the result of opportunistic attacks
- 87% were considered avoidable through reasonable controls

Nine of 10 breaches involved some type of "unknown unknown," the most common of which was data that was not known to be on the compromised system. Most breaches go undetected for guite a while and are discovered by a third party rather than the victim organization. Attacks tend to be of low to moderate difficulty and largely opportunistic in nature rather than targeted. Due, in part, to these reasons, investigators concluded that nearly all breaches would likely have been prevented if basic security controls had been in place at the time of attack.

Where should mitigation efforts be focused?

Given the opportunistic nature and difficulty (or rather lack thereof) of attacks leading to data breaches, organizations are wise to focus on ensuring essential controls are met across the organization and throughout the extended enterprise. This includes following through on security policies so they are actually implemented and ensuring that a basic set of controls is consistently met across the organization. Accomplishing these goals will make it much more likely that attackers will pass over your organization in favor of more lowhanging fruit.

In the modern organization, data is everywhere and keeping track of it is an extremely complex challenge. The fundamental principle, however, is quite simple—if you don't know where data is, you certainly can't protect it. Based on the hundreds of breaches investigated, efforts to locate, catalogue, track, and assess the risk of data stored in and flowing through information assets are highly beneficial in reducing the likelihood of data compromise.

Though crucial, data protection efforts cannot stop with discovery. Once critical data repositories and flows are identified, they must be monitored. Rather than seeking information overload, organizations should strategically identify what systems should be monitored and what events are alertable. Steps should then be taken to ensure alerts are noticed and acted upon when they do happen.







Introduction

"That's how I feel about the skeletons in my laboratory. These have tales to tell us, even though they are dead. It is up to me, the forensic anthropologist, to catch their mute cries and whispers, and to interpret them for the living." -William R. Maples, Dead Men Do Tell Tales In a real sense, this report is a "post-mortem" examination of hundreds of data breach victims. Just as the forensic scientist seeks clues to the sequence of events surrounding a crime, the cause of death, and the identity of suspects, the Verizon Business Investigative Response team is focused on examining evidence of computer crime. Common to investigations in both the physical and cyber worlds is a dependence on cold, hard facts. Building a credible case often requires the collection of heaps of data.

To that end, the 2008 Verizon Business Data Breach Investigations Report integrates a vast amount of factual evidence from forensic investigations over the last four years. The study is unique in that it offers an objective, first-hand view of data breaches directly from the casebooks of our Investigative Response team. Tens of thousands of data points weave together the stories and statistics from compromise victims around the world. We have attempted to interpret their tales and it is our hope that your organization will learn from these findings and thereby avoid their end.

Verizon Business Investigative Response

Security breaches and the compromise of sensitive information are a very real concern for organizations worldwide. When such incidents are discovered, response is critical. The damage must be contained quickly, customer data protected, the root cause found, and an accurate record of events and losses produced for authorities. Furthermore, the investigation process must collect this evidence without adversely affecting the integrity of the information assets involved in the crime.

The Verizon Business Investigative Response team has a wealth of experience and expertise, handling over 500 security breach and data compromise engagements between 2004 and 2007. This includes roughly one-third of all publicly disclosed data breaches in 2005 and a quarter of those in both 2006 and 2007.¹ This caseload represents a large proportion of total known compromised records during this time frame as well as three of the five largest data breaches ever reported.

During such investigations, the team regularly interacts with governmental agencies and law enforcement personnel from around the world to transition case evidence and set the stage for prosecution. In addition to security breach and data compromise cases, the Investigative Response team provides services such as litigation support, e-discovery, expert witness testimony, chain-of-custody, mock-incident training, and incident response program development.

The expansive statistical data set generated through these activities offers an interesting glimpse into the trends surrounding computer crime and data compromise.

¹ Percentage of cases per year based upon comparison of Verizon Business caseload to information obtained from http://www.idtheftcenter.com/.

Methodology

As one might imagine, forensic investigations of security breaches are a potential goldmine of data and insight. This is especially true when one considers that the security industry has long suffered from a dearth of quality data. As such, Investigative Response is a key contributor to the Verizon Business seven types of risk intelligence, helping to inform clients and improve security products and services.

The analysis of data breach trends has been an important function of the Investigative Response team for years. Statistics were collected and used to fuel further inquiries into various topics of interest. Over time, it became evident that a more extensive and systematic process was needed to tap the full potential of Investigative Response as an unparalleled source for quality information security data. At considerable investment in time and resources, an initiative was begun in 2007 to identify a comprehensive set of metrics to record during each data compromise investigation. When completed, several hundred data points were parameterized and a process for collecting them was created and adopted as standard operating procedure on all new cases. At the close of each case, the principal investigator systematically records the metrics along with any other relevant details and then adds this information, after removing client identifiers, to a centralized repository.

As this report covers cases between 2004 and 2007, an alternate method was necessary to compile statistics on historical cases. Two primary methods were employed to collect the data presented in this report. Case files and notes, being the most objective source of information, were the preferred method and were referenced if within retention limits. Even when original reports were available, interviews with case investigators provided a wealth of supplemental data and insight for this study and were absolutely crucial when the former sources were unavailable.

The result of these efforts is the creation of an information repository unlike any other in the world. It includes tens of thousands of data points across 500+ investigations. About one-fourth of publicly disclosed data breaches are contained as well as many never reported. More than 230 million records compromised over a four-year period are represented. Vital details on all aspects of attacks from "probe to p0wn" are included. Furthermore, it contains first-hand information on actual security breaches rather than on network activity, attack signatures, vulnerabilities, public disclosures, and media interpretation that form the basis of most publications in the field. While many reports in the security industry rely on surveys as the primary data collection instrument, this data set is inherently more objective.

In addition to a description of how the statistics presented in this report were collected, another topic warrants attention in a section dedicated to methodology. Though challenges such as sampling techniques, response rates, and self-selection are not relevant to the research method used in this study, it cannot be concluded that the findings are therefore unbiased. Perhaps most obvious is that the data set is dependent upon cases which Verizon Business was engaged to investigate. Readers familiar with publicly available statistics on data loss will quickly recognize differences between these sources and the results presented in this report. This has much to do with caseload. For instance, it is simply more likely that an organization will desire a forensic examination following a network intrusion than a lost laptop. Similarly, the evolution of disclosure and notification laws influences an organization's decision to pursue investigation. That said, there is a wealth of information here and no shortage of valid and clear takeaways. As with any study, the reader will ultimately decide which findings are applicable within their organization.

Finally, it is important to note that Verizon Business is committed to maintaining the privacy and anonymity of Investigative Response clients. Once the investigator records and submits case metrics, this information is sanitized and the client's name is removed from the records. The central repository of case data contains no information that would enable one to ascertain a client's identity. Furthermore, the statistics within this report are always presented in aggregate; individual records are never the focus of analysis.

A Primer on Cybercrime

Crucial to the interpretation of the findings presented in this study is an understanding of the forces that drive cybercrime and the market systems in which it takes place.

Easy money is a motivation that is very powerful to anyone and especially so to the criminal. Data theft is not the only way to achieve this end, but it is one of the easiest, safest, and most lucrative. Criminals could, and do, steal wallets and purses to obtain information necessary to commit identity fraud, access bank accounts, and acquire cash, but the yield is low and the risk is high. Conversely, obtaining the same information on thousands of individuals, often without them even knowing it, is a much wiser course of action. By gaining access to online information systems, the cybercriminal operates with several distinct advantages:

- Higher yield—Vulnerable systems hold information on tens of thousands of victims.
- *Less target resistance*—When breached, systems tend not to fight back and many do not keep a record of what happened.
- Low target sensitivity—It often takes system owners weeks or even months to discover a breach. This allows the criminal to harvest information over a longer period of time.
- Easier escape—When the jig is up, it is significantly easier for the cybercriminal to run and disappear.

The potential value of engaging in cybercrime would not be realized if a market for stolen data did not exist. The social network that is the by-product of the information black market enables players in the criminal underground (hackers, fraudsters, and organized crime groups) to collaborate with one another to find vulnerable systems, compromise data, and commit fraud. Additionally, this market has made the incentives available to a broader population and has allowed individuals and smaller groups to participate in any phase of the data compromise life cycle they choose.

This combination of powerful motivation and an accessible market has enabled the business of cybercrime to grow quickly and rapidly. Prior to the market's existence, the hacker may not have had the social network to sell stolen data and the fraudster may have been limited in the volume of data available to them. A marketplace for compromised data facilitates networking among likeminded criminals, lowers barriers to entry, and enables individuals or groups to make money through cybercrime. Ultimately, it allows the pilfered zeros and ones to be converted into cash and material goods.

Results and Analysis

Demographics

The breaches investigated as part of this study represent a broad spectrum of industries. Supporting the age-old maxims that criminals follow the money trail as well as the path of least resistance, the retail and food and beverage industries account for more than half of all cases. Financial services, though certainly keepers of great monetary assets, are also typically well protected in comparison to other industries; they account for 14 percent of breaches. It's the reason convenience store robberies have always outnumbered bank robberies—risk vs. reward. The calculation is no different when the "cash" is digital. Technology services, which include software firms, data warehousing companies, telecommunication providers, etc., is the only other industry with over 10 percent of breaches. Given the many reports of data breaches involving educational institutions in recent years, it may be surprising to some that these account for a relatively small proportion of our data set. This is certainly more indicative of caseload than any real or important global trend.



Figure 1. Industries Represented

Information is everywhere. It should come as no surprise then that data breaches are without question a worldwide phenomenon. The Verizon Business Investigative Response team handled a marked increase in the number of forensic engagements outside North America during the time frame of this study. Though related to caseload, this fact is surely reflective of a broader trend. As the world becomes more interconnected through information technologies, as enterprises aggressively seek global partnerships, and as the laws governing the handling and disclosure of such incidents mature, it is likely that this trend will continue.

In addition to numerous industries and countries, companies of all sizes are included within the data set that forms the basis of this report. No obvious patterns emerge from the chart below. Data thieves appear perfectly willing to victimize smaller "mom and pop" operations as well as larger enterprises. Where there is information that can easily be converted to cash, that is where the criminals will invariably go.



Figure 2. Number of Employees

Sources of Data Breaches

During the course of an investigation it is critical that pertinent facts surrounding a data compromise be determined as soon as is feasible. One of the more important determinations involves uncovering the source (or sources) of the breach. For many, the phrase "data breach" carries the connotation of criminal intent on the part of some external entity. This is not always the case; security incidents result from deliberate and unintentional actions as well as malicious and non-malicious actors both within and outside the organization.

Although nearly endless sub-categorizations are possible, at a high level information security incidents originate from one or a combination of the following threat sources:

- **External**—Intuitively, external threats originate from sources outside the organization. Examples include hackers, organized crime groups, and government entities but also environmental events such as typhoons and earthquakes. Typically, no trust or privilege is implied for external entities.
- Internal—Internal threat sources are those originating from within the organization. This encompasses human
 assets—company executives, employees, and interns as well as other assets such as physical facilities and
 information systems. Most insiders are trusted to a certain degree and some, IT administrators in particular, have
 high levels of access and privilege.
- Partner—Partners include any third party sharing a business relationship with the organization. This value chain
 of partners, vendors, suppliers, contractors, and customers is known as the extended enterprise. Information is
 the lifeblood of the extended enterprise and it flows far beyond the boundaries of any single organization. For
 this reason, some level of trust and privilege is usually implied between business partners.

Now the obvious question: Who was responsible for the data breaches investigated by Verizon Business between 2004 and 2007? The following chart holds the answer.



Figure 3. Sources of Data Breaches

Our findings indicate that data compromises are considerably more likely to result from external attacks than from any other source. Nearly three out of four cases yielded evidence pointing outside the victim organization. In keeping with other studies revealing risks inherent to the extended enterprise,² business partners were involved in 39 percent of the data breaches handled by our investigators. Internal sources accounted for the fewest number of incidents (18 percent), trailing those of external origin by a ratio of four to one.

The relative infrequency of data breaches attributed to insiders may be surprising to some. It is widely believed and commonly reported that insider incidents outnumber those caused by other sources. While certainly true for the broad range of security incidents, our caseload showed otherwise for incidents resulting in data compromise. This finding, of course, should be considered in light of the fact that insiders are adept at keeping their activities secret.

For others, the real surprise may be that the ratio of external to internal is so slim. In days long past when mainframes ruled the computing world, internal threats were the predominant concern. Ever since outsiders joined the network, however, external attacks (not incidents) have vastly outnumbered those from insiders. The fact that the rate of external and internal compromise is even remotely similar speaks to the higher success rate of insider attacks. These threats are exceedingly difficult to control and, as will be shown later in this section, their consequences far greater.

Simple arithmetic will yield the realization that percentages provided in Figure 3 exceed 100 percent. This is not an oversight; more than a quarter of cases in the study involved multiple sources. Though this sometimes indicated collusion, more commonly one party was an unsuspecting participant to the crime. In a scenario witnessed repeatedly, a remote vendor's credentials were compromised, allowing an external attacker to gain high levels of access to the victim's systems. During an investigation that did reveal definite evidence of collusion, an internal IT administrator was solicited by an external group to essentially open a door that would enable them to access corporate systems.



10

2 Risky Business: Information Risk in the Extended Enterprise, a Verizon Business whitepaper

During the time frame of this study, interesting changes occurred with respect to sources of data breaches. While the percentage of insider events remained relatively stable, breaches of external or partner origins changed significantly, as depicted in Figure 4. The important trend is not so much that external incidents trended downward but that breaches involving business partners increased five-fold between 2004 and 2007. The decline in the percentage of breaches from external sources is simply the by-product of the rise of compromises from partners. This finding is certainly reflective of parallel trends within the extended enterprise emphasizing information sharing, systems integration, and collaboration among business partners.

Breach Size and Source

Having discussed the percentage of incidents originating from outsiders, insiders, and partners, a very intriguing question arises concerning differences among the groups in relation to the size of data breaches. To this point, findings based on our historical caseload prove very revealing.



Figure 5. Median Number of Records Compromised

The median size (as measured in the number of compromised records) for an insider breach exceeded that of an outsider by more than 10 to one. Likewise, incidents involving partners tend to be substantially larger than those caused by external sources. This supports the principle that privileged parties are able to do more damage to the organization than outsiders. Though the number of records is not an equivalent measure of the overall impact from a data breach, it is certainly an indicator. Thus, a "back of the napkin" calculation of risk (likelihood x impact) finds that partners represent the greatest risk for data compromise, followed closely by insiders.

Source	Likelihood	Impact (# of Records)		Risk (Pseudo)
External	73 percent	30,000	=	21,900
Internal	18 percent	375,000	=	67,500
Partner	39 percent	187,500	=	73,125

External Breach Sources

The process of determining the specific source of an external attack is rife with difficulties. The authenticity of the source IP address, the primary means of making this determination, is often questionable. This is especially true for cases resulting from nefarious activity as those responsible are prone to cover their tracks. Many IP addresses discovered during an investigation are either spoofed or tie back to anonymous "zombie" machines or botnets. Furthermore, a crime scene devoid of any network and system logs, a key resource for computer forensics, is a disturbingly common occurrence.

That said, methods do exist for corroborating IP addresses and this information is by no means worthless evidence. Commonalities between cases, correlative fraud patterns, cooperation with law enforcement agencies, and collaboration with Verizon Business NetIntel and ICSA Labs all lend credibility to related findings. Based on investigative evidence and other supplemental information, the geographic distribution of external data breach sources is provided in Figure 6.



Figure 6. Location of Attacking IP(s)

Whether or not the figure represents the "real" geographic source of attack, our caseload reveals very clear links between regions and specific types of attacks. For example, attacks coming from Asia (China and Vietnam in particular) often involve application exploits leading to data compromise while defacements frequently originate from the Middle East. IP addresses from Eastern Europe and Russia are commonly associated with the compromise of point of sale (POS) systems.

In addition to region-specific trends, some overarching themes evolved over the time frame of the study. The growth in both size and activity within online criminal groups is certainly foremost among them. Within our caseload, the percentage of breaches tied to known organized crime doubled each year. As affiliate business networks offer an increasing array of services to other criminal groups, hackers, spammers, and fraudsters, the geography-centric attack trends discussed above are showing signs of dissolution. It is very likely that such developments will continue to shape illicit activity on the Internet in the near future.

Internal Breach Sources

While geography provides an interesting breakdown of external sources, a closer examination of insiders is better accomplished by the individual's role within the organization. Several broad classifications of internal sources are presented in Figure 7, along with the percentage of data breaches attributed to each.



As one might suspect, IT administrators were responsible for more data compromises than any other insider role. The privileges entrusted to this group provide them a much larger opportunity to abuse corporate information systems. Many will note with interest the rather small difference between breaches caused by other employees and IT administrators. These findings are a reminder that high levels of access are not necessary in order to compromise data. Though a few cases involved corporate executives and espionage, these were few and far between.

Partner Breach Sources

The breakdown of partner-related sources in Figure 8 illustrates the pervasiveness of the scenario described earlier in this study. Partner-side information assets and connections were compromised and used by an external entity to attack the victim's systems in 57 percent of breaches involving a business partner. Though not a willing accomplice, the partner's lax security practices—often outside the victim's control—undeniably allow such attacks to take place. Exacerbating this situation, the victim organization frequently lacks measures to provide accountability for partner-facing systems. This contributed to the 21 percent of breaches in which partner involvement was evident but specific persons were not identified.



Figure 8. Breakdown of Partner Sources

Figure 8 also reminds us that not all data breaches within the extended enterprise are unintentional. Of all partner compromises, 16 percent were attributed to the deliberate malicious actions of remote IT administrators. Similar to privileged insiders, these individuals are granted access to internal systems and are prone to misuse it. The remaining 6 percent were traced to other remote partner employees and partners located at or visiting facilities owned by the victim organization.

Threat Categories

A glance at recent headlines is enough to illustrate the plethora of threats facing enterprise data. From network intrusion to laptop theft to administrative errors, sensitive data continues to be compromised from unwilling and often unwitting enterprises all over the world. For organizations trying to avoid such incidents, obvious questions arise: How do breaches occur? What threats are most common? How prevalent are customized attacks? Answering these questions and many others is the purpose of this section.

Most data breaches result from a series of distinct yet related events. Between footprinting and compromise, a myriad of possibilities exist. Though very specific threat and attack details are recorded during an investigation, all possibilities fall within the seven broad threat categories listed in Figure 9. Among those investigated, most incidents resulted from multiple "intra-category" events (i.e., utilized several types of hacking) and many encompassed several threat categories. Figure 9 records the prevalence of each category as a significant contributing factor to data breaches investigated by Verizon Business between 2004 and 2007.



Figure 9. Threat Categories

Error

Loosely defined, error is a contributing factor in nearly all data breaches. Poor decisions, misconfigurations, omissions, non-compliance, process breakdowns, and the like undoubtedly occur somewhere in the chain of events leading to the incident. Because error is so incredibly prevalent, this loose definition tends to lose its meaning in the greater picture. For this reason, only those errors which directly or significantly contributed to the compromise were considered by investigators. A glance at Figure 9 likely engenders questions as to why the "Error" bar is split into solid and shaded regions. The answer is quite simple; the smaller solid region pertains to errors that directly led to the data compromise while the shaded region signifies errors that significantly contributed to it in some way. Several classifications of error are given in Figure 10 along with their relative distribution among cases.



Figure 10. Breakdown of Error

Evident from the graph, significant omissions contribute to a huge number of data breaches. This often entailed standard security procedures or configurations that were believed to have been implemented but in actuality were not. It should be noted that in situations where the investigator felt a control measure could have been implemented to mitigate a breach, it was not considered to be error.

Misconfiguration was apparent in 15 percent of cases, usually manifested in the form of erroneous system, device, network, and software settings. Though accidental disclosure, user blunders, and technical glitches occur frequently, they are only a portion of errors leading to data compromise. Because so many hacking scenarios exploit the configuration (or lack thereof) of systems, these two categories share a kind of symbiotic relationship.

Hacking

In terms of deliberate action against information systems, hacking leads to more data breaches than any other category by a margin of almost two to one. Hacking is relatively free from the constraints that limit other methods (i.e., physical proximity, human interaction, system privileges), a fact making it a favored technique among the cyber underworld. Additionally, many tools are available to help automate and accelerate the attack process. Figure 11 provides a breakdown of the various types of hacking observed by the Investigative Response team.



Figure 11. Breakdown of Hacking

Attacks targeting applications, software, and services were by far the most common technique, representing 39 percent of all hacking activity leading to data compromise. This follows a trend in recent years of attacks moving up the stack. Far from passé, operating system, platform, and server-level attacks accounted for a sizable portion of breaches. Eighteen percent of hacks exploited a specific known vulnerability while 5 percent exploited unknown vulnerabilities for which a patch was not available at the time of the attack. Evidence of re-entry via backdoors, which enable prolonged access to and control of compromised systems, was found in 15 percent of hacking-related breaches. The attractiveness of this to criminals desiring large quantities of information is obvious.



Figure 12. Patch Availability at Time of Breach

For the overwhelming majority of attacks exploiting known vulnerabilities, the patch had been available for months prior to the breach. This is clearly illustrated in Figure 12. Also worthy of mention is that no breaches were caused by exploits of vulnerabilities patched within a month or less of the attack. This strongly suggests that a patch deployment strategy focusing on coverage and consistency is far more effective at preventing data breaches than "fire drills" attempting to patch particular systems as soon as patches are released.

Malcode

Malicious code, or malcode, contributed to the success of nearly one-third of data breaches under investigation. Even more often, it was found on compromised systems but its role in the breach was not confirmed. In this sense, it was an indicator of the general security health of the system rather than an accessory to the crime. In years past, most malcode was delivered in the form of self-replicating e-mail and network worms. The objective of malcode creators was massive and rapid propagation. More recent trends emphasize stealth and smaller, more directed distribution. The modus operandi of the cyber underground has without question shifted away from "hacking for fame" to "hacking for fortune," and malcode mirrors this paradigm. Figure 13 illustrates the extent of this shift, especially with respect to data compromise.





Far more common than any other delivery method was malcode pushed to a compromised system by a remote attacker. The goal of this action, from the criminal's perspective, centers on capture and control. These programs either capture information to be harvested later, capture and then send information to a remote entity, or enable the attacker to access and control the system. Among malcode observed during data breach investigations, the ratio of these functions was roughly equal and often seen in combination.

Another noticeable trend is an increase in customized malcode, which was found in 25 percent of cases. Much of the time, this involved a simple repacking or slight modification of existing code in order to avoid detection by anti-virus scanners. However, in some instances the actual functionality was customized specifically for the victim's systems.

Misuse

Misuse refers to the use of organizational resources and/or privileges for any other purpose than for what or how they were intended. For this reason, the category is particular to insiders and partners, as they are trusted by the organization. It is also very difficult to control. There are two broad classifications of misuse: malicious and non-malicious. Malicious forms include abusing access privileges to steal information or sabotage systems, while the installation of personal software and surfing questionable sites are examples of non-malicious misuse. Among the hundreds of cases investigated, malicious misuse of access or privilege was a factor in 19 percent of data breaches. Though non-malicious misuse contributed to relatively few incidents (3 percent), it is a reminder that such activities can and do damage the company in question.

Physical

Those familiar with statistics on publicly disclosed data breaches may be surprised that this category of threat ranks low on our list. This is clearly attributable to case load. The nature of many physical events precludes the need for any investigation. Moreover, many disclosures related to physical incidents are not actually data compromises. Information on a lost laptop is considered "data at risk" and must be disclosed whether or not the data actually fell into the hands of criminals or was used for fraudulent purposes.



Figure 14. Breakdown of Physical Threats

Figure 14 presents findings from data breach investigations regarding the frequency of physical threats. Theft from victim-controlled premises was the most common physical means of data compromise (39 percent), followed by system access (via keyboard or console) or tampering (27 percent). Wiretapping and sniffing was attributable to 16 percent of physical breaches. Lost information assets and observation (i.e., "shoulder surfing") were each a factor in approximately 6 percent of incidents in this category. The theft of resources from an external location and assault were relatively rare occurrences.

Deceit

This category refers to any deliberate misrepresentation and deceit using both technical and non-technical means. Examples of deceit encountered during data breach investigations include phishing scams (5 percent) and spoofing and masquerading (4 percent). Social engineering, the poster child of this category, was relatively rare, and was used by criminals in only 2 percent of cases. Though often a very effective tactic, social engineering typically requires significant planning and effort to execute properly. As discussed in other sections of this report, criminals are unlikely to employ difficult tactics when easier avenues of compromise are available to them.

Environmental

Events of this category are a much greater threat to system availability than the confidentiality of information. That said, there were a few instances among our caseload in which environmental causes were a contributing factor to a data breach. In one case, a storm caused a power outage which led to a system reboot. Consequently, the system lost all security settings and was soon compromised as a result. Thus, business continuity procedures are in some rare instances also helpful in preventing data breaches.

Attack Difficulty

Though some movie plots would have us believe otherwise, cyber attacks in the real world rarely involve *Mission Impossible*-like scenarios. Quite the opposite, in fact, as the figure below demonstrates. Although rating attack difficulty admittedly involves some level of subjectivity on the part of the investigator, it is a profitable exercise nonetheless. During each data breach investigation, the attack in question was analyzed and given one of the following difficulty levels:

- None—No special skills or resources were used. The average user could have done it.
- Low-Low-level skills and/or resources were used. Automated tools and script kiddies.
- Moderate—The attack employed skilled techniques, minor customization, and/or significant resources.
- High—Advanced skills, significant customization, and/or extensive resources were used.



Figure 15. Attack Difficulty

Given enough time, resources and inclination, criminals can breach virtually any single organization they choose. They cannot breach all organizations. The math is simple: Money is the motivator. More compromises lead to more money. Compromises are maximized when effort is minimized. Unless the value of the information to the criminal is inordinately high, it is not optimal for him to expend his limited resources on a hardened target while a softer one is available. The goal, then, is to implement security measures such that it costs the criminal more to compromise your organization than other available targets. Figure 15 reminds us that those softer targets are not yet in short supply.

Targeted vs. Opportunistic Attacks

Standard convention in the security industry classifies types of attacks into two broad categories: opportunistic and targeted. Due to significant grey area in this distinction, we find it useful to separate opportunistic attacks into two subgroups. The definitions are provided below:

- Opportunistic (Random)—Attacker(s) identified the victim while searching randomly or widely for weaknesses (i.e., scanning large address spaces) then exploited the weakness.
- **Opportunistic (Directed)**—Although the victim was specifically selected, it was because they were known to have a particular weakness the attacker(s) could exploit.
- Targeted—The victim was first chosen as the target and then the attacker(s) figured a way to exploit them.

The mere mention of the phrase "targeted attack" is enough to generate concern among organizations the world over. As alluded to in the previous section, an organization singled out by an attacker with sufficient resources will find it difficult to mount an adequate defense. Significant investments are made toward initiatives focused on mitigating targeted attacks but are such expenditures warranted? Perhaps a better question is what is the likelihood that your organization will be targeted?



Figure 16. Targeted vs. Opportunistic Attacks

Based on data collected by our Investigative Response team, it is not as likely as one might think to be the victim of a targeted attack. In only 15 percent of cases did it appear that the organization was compromised by a truly targeted attack. Not surprisingly, the financial industry suffered a higher rate of targeted attacks. Another observation was that these attacks often utilized different methods than opportunistic attacks. For example, social engineering was used in several of the more prominent examples of this type. On the other hand, random opportunistic attacks, which accounted for 39 percent of data breaches investigated, lend themselves to less sophisticated and more automated methods.

Through the years we have encountered a great number of cases (46 percent) involving attacks that are neither fully opportunistic nor truly targeted according to the definitions above. These are clearly of opportunistic nature (the target was chosen because of a weakness) but they are not random and actually appear targeted in many instances. In a very common example of this type, an attacker exploits Software X at Brand A Stores and then targets Brand B Stores after learning that it also runs Software X. This is very common among the retail and food and beverage industries.

Common Attack Pathways

In addition to the threat categories discussed in the preceding section, it is useful to examine the pathways exploited by data thieves as they conduct their nefarious activities. In this context, the pathway refers to the interface through which an attacker gains access to corporate systems. Though by no means an exhaustive list, some of the more commonly observed attack pathways are shown in Figure 17 along with the percentage of cases in which they were exploited.



Figure 17. Attack Pathways

In over 40 percent of the breaches investigated during this study, an attacker gained unauthorized access to the victim via one of the many types of remote access and control software. On many occasions, an account which was intended for use by vendors in order to remotely administer systems was compromised by an external entity. These vendor accounts were then used to illegitimately access enterprise information assets. This scenario is particularly problematic due to the fact that, from the victim's perspective, the attacker appears to be an authorized third party. In many of these cases, the remote access account is configured with default settings, making the attacker's job all too easy.

Given recent trends, it is not surprising that web applications are near the top of the list. Unlike most information assets which have limited visibility outside the organization, web applications are by design accessible to the world at large. Through attacks like SQL injection (a commonly observed attack), criminals often exploit this exposure to their own ends. It is worth noting that other Internet-facing systems were a factor in almost a quarter of breaches. Though not always the case, these systems are regularly connected to the Internet without the knowledge of the victim. In another manifestation of this, the attacker gains entry through an existing backdoor.

Despite the large amount of media attention given to the supposed weakness of wireless networks, this vector was exploited considerably less than others presented in Figure 17. When wireless infrastructure was the means of entry, it was due to poor configuration and weak encryption rather than a successful attack against an adequately secured WLAN. As physical attacks require physical access, the relationship between this pathway and the physical threat category is self explanatory.

Information Repositories and Channels

Having discussed the methods and vectors utilized by attackers to gain access to corporate resources, a logical next step is to examine which types of information assets are actually compromised. Obviously, enumerating all possible types of repositories of information and the pathways through which it travels is a daunting task and difficult to present. Therefore, to provide a simple yet effective presentation of our findings, we have divided the world of information assets into four broad classes shown in Figure 18.



Figure 18. Compromised Assets

The type of asset compromised most frequently is without doubt online data. Compromises to online data repositories were seen in more cases than all other asset classes combined by a ratio of nearly five to one. Offline data, networks, and end-user devices were all closely grouped. An alternative method of analyzing these results is to examine the number of records of sensitive data compromised for each asset. This view is given below although the figure shows the same conclusion.



Figure 19. Compromised Assets (Percent of Records)

This fact may be surprising to some given the frequent public reports of massive amounts of data at risk from lost or stolen laptops, back-up tapes, and other media. To that point, it is noteworthy that the average number of records compromised per incident was higher when offline data repositories were involved than with online data. As with all results presented in this report, this is a by-product of our caseload and should be considered in that light.

Types of Data Compromised

The cases included in this study encompass an astounding 230 million compromised records, a large portion of publicly disclosed records³ breached during the four-year time frame of the study. The average number of records per breach was approximately 1.2 million. The median, however, is much lower at 45,000, indicating a skew in the dataset toward a few very large breaches. Even so, over 15 percent of cases involved more than 1 million records.

One of the most critical components of the investigative process is the determination of what types of data were compromised during the breach. A victim organization's remediation strategy is determined largely by the type of data compromised. For instance, certain types require public disclosure and/or notification of the individuals involved. Some types of data are highly regulated, while still others require monitoring for fraudulent activities. In addition to the victim organization, potentially millions of individuals could be affected by the compromise.



Figure 20. Compromised Data Types

As is evident from Figure 20 above, some type of cardholder data was compromised in 84 percent of cases. Once again, this correlates to the financial motivation of the criminals. Related findings support this statement, as fraudulent use of stolen information was detected following 79 percent of breaches. Additionally, 32 percent of cases involved one of the many types of personally identifiable information (PII). This is likely attributable to the usefulness of this type of data for committing fraud and other criminal activities.

Nonsensitive data was compromised in 16 percent of cases, but this is most likely the by-product of a breach in which other, more sensitive information was targeted. Authentication credentials (15 percent) are desired by attackers because they allow the prospect of increased privileges and access for subsequent illicit activities. The compromise of intellectual property and corporate financial data were relatively rare, likely due to the difficulty of quickly and easily converting this type of information into cash.

Time Span of Data Breach Events

As might be imagined, the time span of events leading up to and following a data breach varies greatly depending on a multitude of factors. Some attacks unfold rapidly, compromising systems within a matter of minutes. Others take months or even years of planning and execution. Though any number of events can occur during this time, it is helpful and straightforward to separate an incident into three major phases: point of entry to compromise, compromise to discovery, and discovery to mitigation. From Figure 21, the variation among these phases is immediately evident and very telling.

³ Based on publicly disclosed data breach information obtained from http://www.idtheftcenter.com/.



Figure 21. Data Breaches: A Time Span of Events

In comparison to the other categories, the length of time between the attacker's initial entry into the corporate network and the compromise of information is relatively short. During this phase, intruders typically explore the network and systems until finding their desired plunder. To an attacker unfamiliar with the territory, this can be a time-intensive activity. Surprisingly, our findings reveal this was accomplished within minutes or hours in just under half of cases investigated.

Such a short period of time required to locate and compromise information often indicates that the attacker had prior knowledge of the victim's systems. In many cases exhibiting a short duration in this phase, no evidence of network mapping or exploration was found. This scenario was very common among breaches occurring within a short time frame and linked by some shared platform or application. In such instances, the attacker essentially engaged in a "crime spree," using the same techniques on multiple victims using similar technologies in rapid sequence. As long as criminals do not have to work very hard or very long to achieve their objectives, the "ROI" for such attacks will remain high from their perspective.

Of course, not all attacks followed this format. Roughly 25 percent took a week or more to compromise data after breaching the perimeter. In these scenarios, intruders often left malcode on systems to capture information over a period of time and then returned later to retrieve it.

In sharp contrast, it takes much longer for organizations to discover a compromise. Months or even years transpired before this realization dawned on the majority of those in our caseload. We find this statistic to be astounding. What factors contribute to this discouraging state of affairs? To this question, we offer two suppositions. Firstly, and perhaps most obviously, criminals do not want to be discovered. They have great financial incentive to retain access to corporate systems for as long as possible and will go to great lengths to ensure their activities remain under the radar. Secondly, and perhaps most importantly, organizations are simply not watching. Most breaches are discovered by a third party rather than the victim, a fact that will be discussed in greater detail in the following section.

Once data compromise is finally discovered, results show that organizations are rather slow to respond. Containment often takes weeks or months and is rarely accomplished within hours of discovery. Rather than apathy, we believe the main reason for this is that victims do not know how to respond. Many organizations—even those with full-time security resources—either have no incident response plan, or have never vetted it against real-world incident scenarios.

Data Breach Discovery Methods

The protracted length of time during which a breach goes unnoticed by the victim begs the question of how organizations finally become aware of their circumstances. There is no shortage of technologies, processes, or services available to alert customers of such events. In cases handled during the four-year period of this study, investigators made it a point to ascertain how the compromise was discovered. These observations are represented among the categories depicted in Figure 22. The percentage of cases for which each method contributed to the discovery of the breach is also included.





By a substantial margin, the most common way in which organizations became aware of data breaches was through notification by a third party. Often, this involved the third party detecting suspicious activity or fraudulent use of compromised data that was later traced back to the victim. Interestingly, the organization's own employees were second on the list, catching 12 percent of breaches during the course of their daily work activities. All other methods fell well below the 10 percent mark.

Perhaps the most notable statistic in Figure 22 is the 4 percent of incidents that were detected through event monitoring and other forms of analytic technologies. Intuitively, these controls should detect a large proportion of data compromise events, yet our findings strongly contradict this position. Are these technologies not deployed? Are they inherently ineffective? Has the evolution of cyber attacks rendered these measures obsolete? We answer in the negative on all accounts. These are not new technologies and adoption rates have been high for some time. ICSA Labs, an independent division of Verizon Business, has tested many of these devices over the years and certified their effectiveness. Furthermore, most information security guidelines contain provisions for log monitoring, routine audits, and incident response procedures.



Figure 23. Discovery Methods—Simplified

The fact of the matter is that though most organizations have the technologies, people, and know-how required to detect and respond to data compromise events, they seldom do so. In 82 percent of cases, our investigators noted that the victim possessed the ability to discover the breach had they had they been more diligent in monitoring and analyzing event-related information available to them at the time of the incident. The breakdown is in the process. What these organizations seem to lack is a fully proceduralized regimen for collecting, analyzing, and reporting on anomalous log activity.

This point is clearly illustrated in Figure 22, which provides an alternate view of the same data presented in Figure 23. As shown, the organization was "active" (referring to measures taken that are specifically designed for detection) in the discovery process in only 7 percent of cases. Eighteen percent of breaches are discovered "passively" (i.e., due to unusual behavior exhibited by a compromised system), while third parties alerted the victim in the remaining 75 percent.

Anti-Forensics

The term "anti-forensics" is used to describe any and all actions taken by an unauthorized intruder to conceal evidence of their actions and make ensuing investigations difficult. Although anti-forensics often involves sophisticated software and techniques, it can also take the form of simple hacks and workarounds that mask an intruder's digital footprint. Securely deleting critical log files such that they can not be easily recovered, for example, would be a considered an anti-forensic technique.

Unfortunately for investigators, many anti-forensic tools are readily available and operationally intuitive. What's more, these tools are becoming ever smarter. Some newer proof-of-concepts directly attack the very tools used by investigators to examine evidence. In practice, however, anti-forensic techniques are not perfect; intruders often remove some traces of their actions but leave investigators plenty of evidence to examine.

Collecting and analyzing statistical data surrounding the use of anti-forensic techniques presents an intrinsic challenge. That is to say, the use of truly effective anti-forensic measures should ostensibly leave no trace that they were used at all. With that in mind, the Investigative Response team discovered signs pointing to the use of anti-forensics in 39 percent of cases. Given recent activity, this will be a trend to watch over the next few years.

Unknown Unknowns

Throughout hundreds of investigations over the last four years, one theme emerges as perhaps the most consistent and widespread trend of our entire caseload. Nine out of 10 data breaches involved one of the following:

- A system unknown to the organization (or business group affected)
- A system storing *data* that the organization did not know existed on that system
- A system that had unknown network *connections* or accessibility
- A system that had unknown accounts or *privileges*

We refer to these recurring situations as "unknown unknowns" and they appear to be the Achilles heel in the data protection efforts of every organization—regardless of industry, size, location, or overall security posture. For this reason, investigators make a special point of determining whether any of these scenarios contributed to a data compromise incident. The percentage of cases in which each of these unknown unknowns was present is shown below in Figure 24.



Figure 24. Unknown Unknowns

Two-thirds of the breaches in the study involved data that the organization did not know was present on the system. We believe this to be due to the common practice of establishing security requirements for a system commensurate with the sensitivity of the information stored within it. While certainly logical, this approach fails when the organization is unaware that sensitive data exists on the system. Less stringent controls are prescribed for the system, leaving the data inadequately protected. As information is propagated and replicated throughout the organization, it invariably makes its way to places it was not intended to be. Criminals, ever vigilant for easy prey, often exploit such circumstances.

Due largely to integration within the extended enterprise, unknown network connections were a factor in 27 percent of breaches while unknown privileges contributed to 10 percent. Business needs often require that partner-facing connections and accounts be provisioned quickly. Unfortunately, proper management and eventual deprovisioning of these assets is overlooked in many cases. Though not as common as other unknowns, 7 percent of breaches did involve an asset the victim did not even know was under control of their business group. Organizational silos, poor governance, unclear ownership, and poor communication exacerbate these issues.

Conclusions and Recommendations

Perhaps the most significant statistic coming out of this historical analysis is that, in 87 percent of cases, investigators concluded that the breach could have been avoided if reasonable security controls had been in place at the time of the incident.

This simple statistic calls for a fundamental shift in data protection and incident response mentality. Traditionally, organizations have aligned their focus on building security controls around the network perimeter, and in many cases, have turned a blind eye toward data within the network. While a strong network perimeter is important, it cannot be the only or even the main layer of protection around sensitive information assets. Information itself— wherever it flows—must be the focus of security efforts and this cannot be achieved under this paradigm. The following recommendations provide a starting point.

Align process with policy—In 59 percent of data breaches, the organization had security policies and procedures established for the system but these were not enacted through actual processes. Stated differently, victims knew what they needed to do, fully intended to do it, but did not follow through. For this reason, controls focused on accountability and ensuring that policies are carried out can be extremely effective in mitigating the risk of data compromise. Checks and rechecks are certainly not a novel recommendation and they lack the panache of new gizmos but our findings attest to their value within the security program.

Achieve "essential" then worry about "excellent"—Eighty-three percent of breaches were caused by attacks not considered to be highly difficult. Eighty-five percent were opportunistic. These statistics are important because they remind us that criminals prefer to exploit weaknesses rather than strengths. In most situations, they will look for an easy opportunity and, finding none, will move on. Many of the victims in this study worked hard to achieve very high levels of security in numerous areas but neglected even minimal control of others. Can you guess which door the criminals chose? Identifying a set of essential controls and ensuring their implementation across the organization without exception and then moving on to more advanced controls where needed is a superior strategy against real-world attacks.

Secure business partner connections—Partners, whether intentionally or unintentionally, contributed to 39 percent of data breaches in the study. A large proportion of these would likely have been avoided through the implementation of basic partner-facing security measures. Additionally, partner assessments against a set of essential controls, contracts that clearly delineate responsibilities and liabilities, improved provisioning, management and deprovisioning of partner connections and accounts, and adherence to the principle of least privilege are all viewed as beneficial in managing partner-related risk based on these findings.

Create a data retention plan—Sixty-six percent of breaches involved data that the victim did not know was on the system. Clearly, knowing what information is present within the organization, its purpose within the business model, where it flows, and where it resides is foundational to its protection. The purpose of an official data retention plan is to provide very specific policies and procedures regarding an organization's management of sensitive data. Organizations should identify and quantify the types of data retained during business activities and then work to categorize data based on risk and liability. In doing so, they should determine what data absolutely cannot suffer compromise and prioritize accordingly. Where not necessitated by valid business need, a strong effort should be made to minimize the retention and replication of data. The creation of a data retention plan should force an organization to discover unknown information, where it lives, who touches it, and what controls are in place to protect it.

Control data with transaction zones—Once an organization has created a strategy for data retention, the next step is to define an approach to securing that data. In so doing, the creation of specific "transaction zones" should be considered. Transaction zones serve as the foundation for IT security which enables organizations to establish granular controls as well as additional layers of accountability (logging). On this platform, organizations can deploy

measures such as two-factor authentication or one-time passwords for contractors, etc. Events out of compliance with data control standards are prime candidates for alerts which can be acted upon by the organization. These noncompliance alerts may allow the organization to identify and react to events taking place between the point of entry and compromise.

Monitor event logs—Evidence of events leading up to 82 percent of data breaches was available to the organization prior to actual compromise. Regardless of the particular type of event monitoring in use, the result was the same: information regarding the attack was neither noticed nor acted upon. Processes that ensure the timely, efficient, and effective monitoring of and response to network events are critical to the goal of protecting data. Such procedures are not new—but they are needed.

Create an incident response plan—If and when a breach is suspected to have occurred, the victim organization must be ready to respond. An effective incident response plan helps ensure that a breach can be stopped prior to a data compromise, and that evidence is collected in such a manner that enables the business to pursue prosecution when necessary. The incident response plan should also address the organization's freeze points—the circumstances which exceed local resources' knowledge and capabilities. A proper incident response plan also details established relationships with law enforcement, third-party counsel, and investigative support. As victim organizations may be required to inform the impacted customer about the breach and data compromise situation, policies and procedures detailing that process should be included in the incident response plan as well.

Increase awareness—Twelve percent of data breaches were discovered by employees of the victim organization. This may not seem like much, but it is significantly more than any other means of internal discovery observed during investigations. Why not improve on a good thing? By implementing a required awareness program, an organization can effectively educate employees about the risks of data compromise, their role in preventing it, and how to respond when incidents do occur. If delivered effectively, and with proper incentives, this training can provide a blanket of basic knowledge across the organization on issues pertinent to data protection.

Engage in mock incident testing—In order to operate efficiently, organizations should undergo routine training in the area of incident response. Attendance at this training should be required as mandatory by policy and cover response strategies, threat identification, threat classification, process definition, proper evidence handling, and mock scenarios. Mock scenario training addresses several key facets of the incident response process and is designed to specifically articulate the step-by-step procedural elements presented within documentation. These training scenarios should provide a complete walkthrough of the incident response and investigative process and specific "discussion points" that represent key learning opportunities.

These recommendations are in no way a comprehensive strategy for enterprise data protection and certainly cannot guarantee against compromise. Rather, they are derived from experience gained through hundreds of data breach investigations and are intended to highlight specific problem areas common to many organizations within our caseload. Used properly, these measures can help guide organizations toward improving their existing controls, adding accountability structures, and defining proper processes to guide them through containment and remediation in an efficient manner. The threat landscape is always changing and organizations must continually adapt their data protection strategies accordingly.

www.verizonbusiness.com

© 2008 Verizon. All Rights Reserved. WP13028

Subject to Terms of Use available at http://securityblog.verizonbusiness.com/disclaimer-2/.

The Verizon and Verizon Business names and logos and all other names, logos, and slogans identifying Verizon's products and services are trademarks and service marks or registered trademarks and service marks of Verizon Trademark Services LLC or its affiliates in the United States and/or other countries. All other trademarks and service marks are the property of their respective owners.

