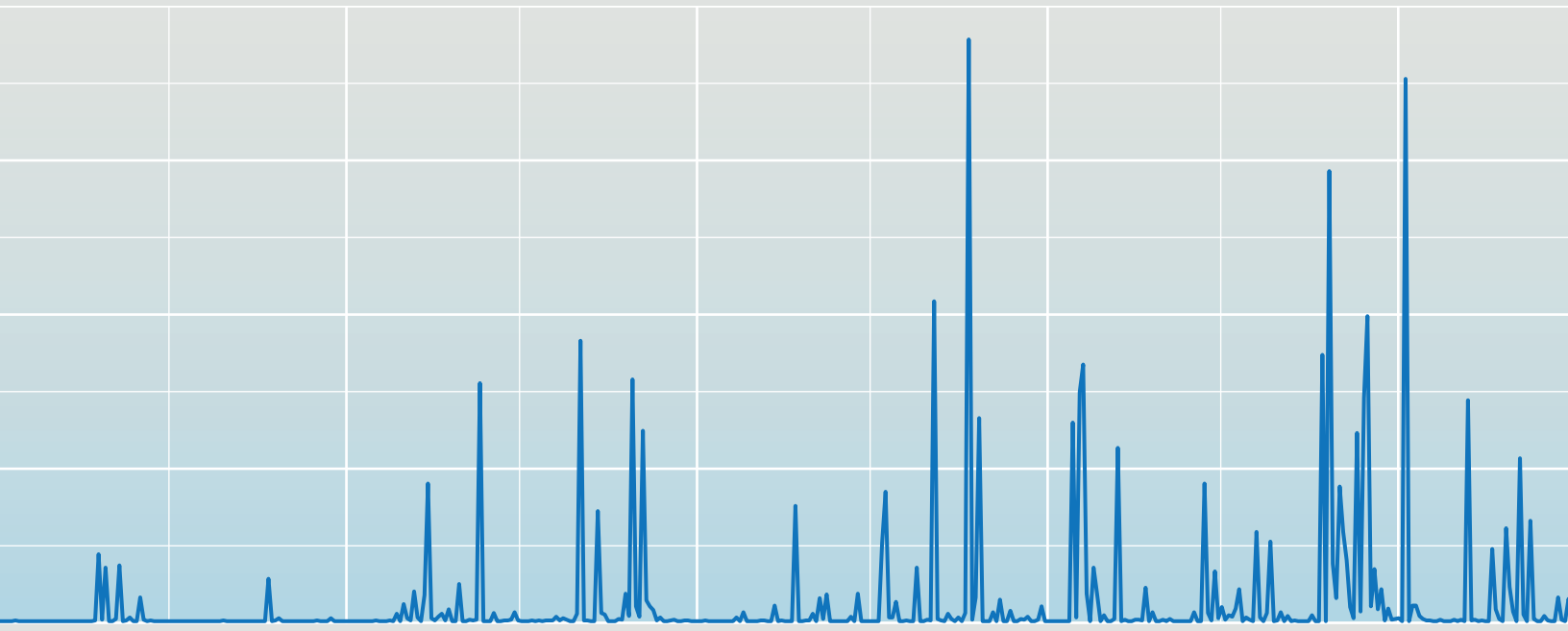
2015 Protected Health Information Data Breach Report

90% of industries have experienced a PHI breach.

verizon



2015 PHI Data Breach Report Contributors

As with the *Data Breach Investigations Report* (DBIR), we could not do this without the contributions from our partners, and we want to extend our thanks to them.

Contents

Lead author
Suzanne Widup

Authors
Gabriel Bassett
Dave Hylender[1]
Bob Rudis
Marc Spitler

1 Chief Snarkitect

Introduction

Welcome to the first *Verizon Protected Health Information Data Breach Report* (PHIDBR). We're the same team that has brought you the *Verizon Data Breach Investigations Report* (DBIR) since 2008, and we are excited to revisit some of that data and bring in some new incidents for this report.

The purpose of this study is to shed light on the problem of medical data loss—how it is disclosed, who is causing it and what can be done to combat it. This is a far-reaching problem that impacts not only organizations that are victims of these breaches, but also doctor-patient relationships. And it can have consequences that spread more broadly than just those directly affected by the incidents.

For the purposes of this study, protected health information (PHI) is defined as personally identifiable health information collected from an individual, and covered under one of the state, federal or international data breach disclosure laws. PHI may be collected or created by a healthcare provider, health plan, employer, healthcare clearinghouse or other entity. The main criteria is whether there is a reasonable basis to believe the information could be used to identify an individual. In the U.S., the disclosure of this type of information would trigger a duty to report the breach under the Health Insurance Portability and Accountability Act (HIPAA), the Health Information Technology for Economic and Clinical Health Act (HITECH) and one or more of the state laws.

PHI redux

Even those who help defend it are often unaware just what makes up PHI data. While you need to work with your legal staff on the specifics for each jurisdiction, the following data elements associated with an individual—alone or in combination—are interpreted as PHI by many laws:

- **Name**, **address** (including just postal code), **telephone** and **fax numbers**
- **E-mail addresses**
- **Medical insurance** or **Social Security/National Insurance numbers**
- Any **date** more granular than year
- Information about **beneficiaries**
- Other (financial or otherwise) **account numbers**, **license**, **vehicle** or **certificate numbers**
- (Medical or otherwise salient) **device** or **serial numbers**
- Any associated **Internet Protocol (IP) addresses** or **URL/URIs**
- All **biometric** data (i.e., finger, retinal or voice prints and/or DNA)
- Full-**facial photographic images** or images that have **unique identifying characteristics**
- **Medical records**

Methodology

In this report, we focus on PHI and the many ways it can be disclosed. Our dataset for this consists of 1,931 records taken from a combination of the DBIR and the Vocabulary for Event Recording and Incident Sharing (VERIS) Community Database (VCDB).[2] The oldest record is from 1994, but most incidents occurred between 2004 and 2014.

Instead of simply taking incidents where the industry came under the North American Industry Classification System (NAICS[3]) code for "healthcare" (62), we chose a more comprehensive approach to capture incidents that would indicate the most common ways PHI is disclosed. We selected records that met any of the following criteria:

- The industry was "healthcare."
- The data type lost was "medical records."
- The data subject/victim relationship was "patient."

In the DBIR, a data breach is "an incident that resulted in confirmed disclosure (not just exposure) to an unauthorized party."[4] Since the reporting laws that govern PHI—largely HIPAA and HITECH—do not have a requirement for confirmation of disclosure, this report expands the definition to include the at-risk category.[5] If the data is at risk and in someone else's hands, it triggers the requirement to report, so we have included these incidents. What is an example of the difference between at-risk data and confirmed compromises? Think of a laptop that is stolen—can you confirm whether the data has been accessed? Probably not. As you no longer have custody of the asset you cannot perform digital forensics on it. Is the data at risk? Absolutely. Especially if the device is only protected with a password, which is trivial to bypass.[6]

2 http://vcdb.org/
3 http://www.census.gov/eos/www/naics/
4 http://verizonenterprise.com/DBIR
5 http://veriscommunity.net/enums.html#section-attributes
6 In fact, you can do your own search for "bypass Windows password on laptop" or use this handy URL: http://bfy.tw/2UYL

Victims' vitals

(Demographics)

Our dataset includes incidents from 25 countries, with 90% of the top-level NAICS industry codes represented. There were over 392 million records disclosed that we know of—since 24% of these organizations did not provide a finite number of records involved, the total could be much higher.

There is a strong U.S. bias to the data (87% of incidents), since it includes U.S. Department of Health and Human Services (HHS) incidents, as well as a significant number of records from the U.S. Department of Veterans Affairs (VA), as reported to Congress. Since the VCDB dataset focuses on publicly disclosed breaches, the likelihood of a country appearing is related to the strength of its breach-reporting laws. As countries implement tougher reporting requirements and the makeup of our DBIR data contributors change, we may gain views into other countries where we currently have no representation.

The U.S. bias does not mean that this report isn't useful for organizations elsewhere in the world. Our data has consistently shown that adversaries' tactics are influenced by the data they are interested in, as well as the assets that process and store the data—not the country in which the data resides. Attack methods are not tied to latitude and longitude—human error, a major cause of breaches, is a global phenomenon too.

Our data has consistently shown that adversaries' tactics are influenced by the data they are interested in, as well as the assets that process and store the data—not the country in which the data resides.

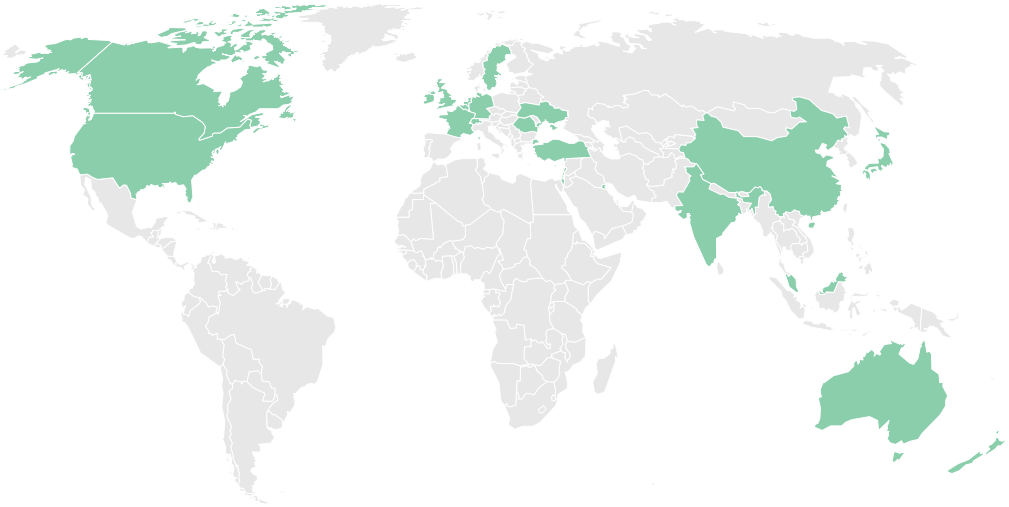


Figure 1.

Countries represented in this study

Unsurprisingly, healthcare is the top industry in this dataset—but remember, falling under NAICS 62 was just one of the criteria for being included in this dataset. It is interesting to see how many others are included as having met one of the other two criteria—either disclosing medical data or having the relationship with the data subject as "patient." Only two of the top-level industries have no incidents matching the criteria. This really illustrates the diversity of industries that have lost PHI.

Only two of the top-level industries have no incidents matching the criteria.

Industry (NAICS code)	Total	Small	Large
Agriculture (11)	1	1	-
Mining (21)	2	1	1
Utilities (22)	-	-	-
Construction (23)	1	1	-
Manufacturing (31–33)	11	2	5
Trade (42)	10	7	3
Retail (44–45)	43	15	17
Transportation (48–49)	8	2	4
Information (51)	2	-	-
Finance (52)	113	42	54
Real estate (53)	4	3	1
Professional (54)	35	18	-
Management (55)	-	-	-
Administrative (56)	24	14	5
Educational (61)	51	5	33
Healthcare (62)	1,403	573	339
Entertainment (71)	1	-	1
Accommodation (72)	3	1	1
Other services (81)	19	14	2
Public (92)	177	31	38
Unknown	23	-	-

Table 1.

Breaches by industry and organization size (where organization size is known)

Why are so many other industries having breaches that include PHI? If you think about it, it is surprising that more are not included. How many companies have employees? How many of those employees are involved in workers' compensation claims? These are likely to include health information, so that is one source where we'd expect to see this type of data collected.

What about companies that collect information for their wellness programs? Some of that data qualifies as PHI as well. Still other organizations obtain PHI as part of managing their employee health insurance programs. Whether they manage these programs directly (as self-insured entities), or they are getting information from

the partner that handles this type of benefit, these can be sources of PHI in organizations that are not covered by HIPAA. Even though an organization is not a HIPAA-covered entity, if PHI is disclosed, many of the existing laws will require notification of a breach to any potentially affected party.

PHI loss is not strongly correlated with organization size.

The fact that an organization is not in the healthcare industry or isn't a HIPAA-covered entity doesn't mean that it's not at risk of a PHI data breach.

Apart from employees, many organizations collect PHI as part of doing business with their customers. The insurance industry is a prime example, and one where we have seen some very large data disclosures recently. The fact that an organization is not in the healthcare industry or isn't a HIPAA-covered entity doesn't mean that it's not at risk of a PHI data breach.

While we show more incidents in smaller organizations, there were also many at organizations where we are unsure of the number of employees. So what does this tell us? Even after taking into account the incidents where the company size was not provided, the remaining incidents show that PHI loss is not strongly correlated with organization size.

Invasive organisms

(Actors and Actions across the dataset)

We'll spend most of this report breaking out the data into different points of view. Let's start with a quick look at the Actors and their Actions.

Actors
Figure 2 shows us information about the perpetrators—those who are actively causing these breaches. While they're not all going to be "Bad Guys Doing Bad Things," you can see that overall there are quite a number of External Actors doing dirty deeds.[7]

But the insider threat is alive and well, given the high number of Internal Actors. Keep in mind that many of the insider actions are accidental—not everyone here is malicious, as you'll see in the Actions section.

The top three actions related to PHI incidents are Physical, Error and Misuse.

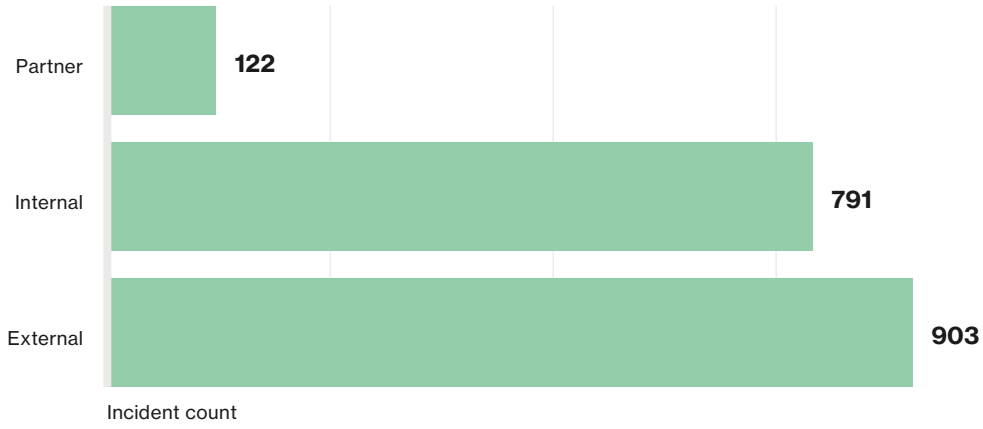


Figure 2.
PHI dataset overview—Actors

Actions
The top three Actions related to PHI incidents are Physical, Error and Misuse. Readers of the DBIR have seen the Physical Action manifest itself in several types of incidents, including the deployment of skimming devices on ATMs and gas pumps. In the PHI dataset, Physical Actions are primarily incidents of theft.

Threat Action categories are commonly associated with a specific type of Actor—although there can be overlap and even collusion. For example, hacking is most often an External Actor, but we have seen hacking by insiders too. Physical Actions are also mostly externally committed, but employees have been known to steal

7 Are they doing them dirt cheap?

equipment. Social Actions frequently include phishing by External Actors, but sometimes we see employees using their influence to get people to do something they should not.

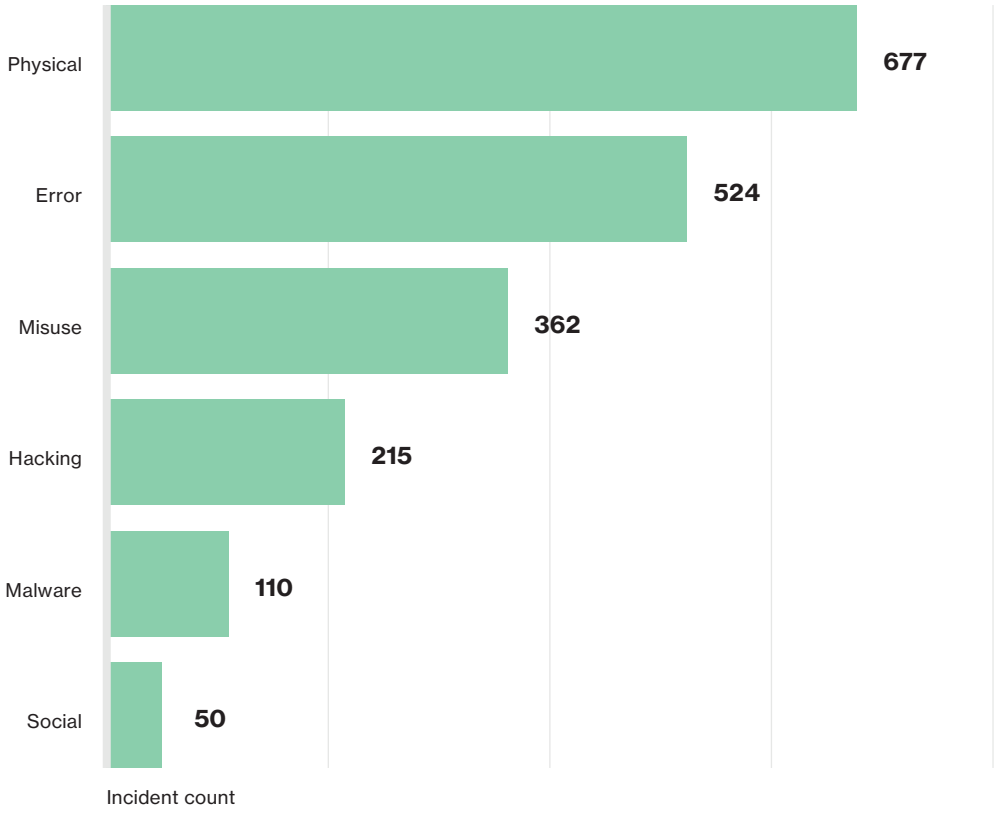


Figure 3.
PHI dataset overview – Actions

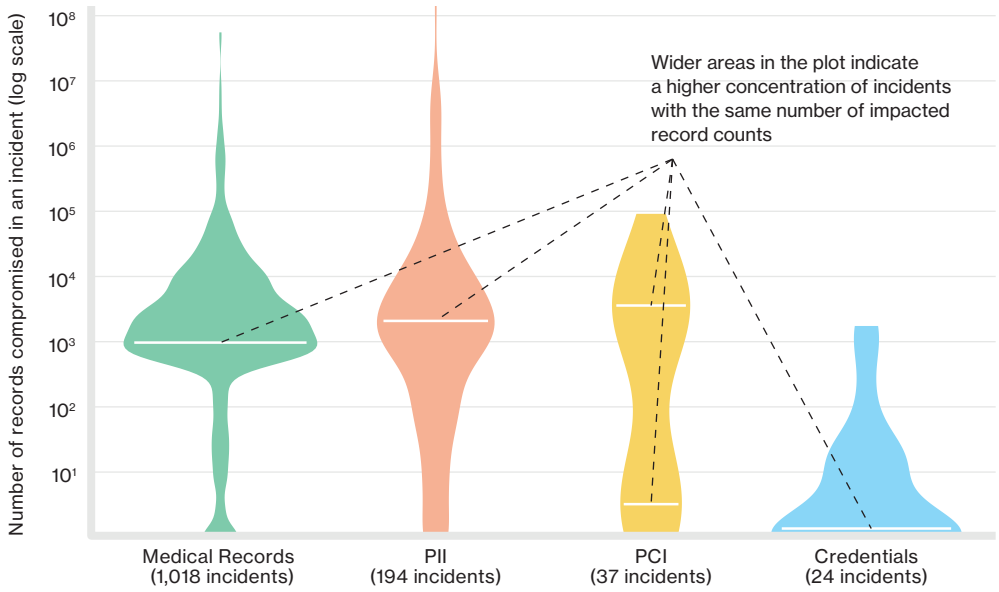
Data types

Earlier we presented a non-inclusive list of elements any one of which would, by law, define data as PHI. Many of these elements group together into broader data types. The following data types were disclosed as a result of the incidents that comprise this study:[8]

- **Medical records** – Likely what first comes to mind when one thinks of a PHI breach. Includes, but is not limited to, diagnosis information, lab results, treatment plans, etc.
- **Payment or payment card industry (PCI) information** – Credit card information
- **Personal or personally identifiable information (PII)** – Personal information (e.g., Social Security/National Insurance numbers, name, date of birth)
- **Credentials** – Unlike the preceding types, this information is not PHI in and of itself, but compromised credentials often are the gateway to the theft of data of other types.

In incidents where records contain more than one data type, we will use the more specific of the two when categorizing them. In the common scenario where a database record contains medical information as well as PII, the incident is classified as "medical records."

8 Other data types, such as banking information and sensitive internal data was also disclosed, but not with enough frequency to include in this section.



In the common scenario where a database record contains medical information as well as PII, the incident is classified as "medical records."

Figure 4.

Data types disclosed

Spending some quality time with Figure 4 reveals some interesting stories about the various data types. Starting with credentials, we can see from the bell-shaped thingy that high-volume credential breaches are the exception and single credential losses are far more prevalent. We can infer that in most cases credentials are not being harvested in bulk from databases, but stolen one at a time via keyloggers, phishing campaigns or even guesswork. Credentials are a gateway data variety that provide access to other targeted data such as PCI information or PII.

Medical record and PII loss is far likelier to be described in the thousands or greater. Acknowledging that there are plenty of Misuse incidents involving small disclosures of these data types, medical records and PII lend themselves to larger breaches. This is due to the nature of these records. They are often compromised from databases and other assets that store data in bulk.

PCI data shows a wider distribution across records per incident. We see significant areas of concentration in record losses totaling under 100, as well as between 1,000 and 10,000. This shows, regardless of the amount of payment card transactions an organization processes, security practices around protection of the point-of-sale (POS) environment is a necessity.

Patients losing patience

(Data subjects and relationships)

Think about the last time you talked to your doctor. You were likely in a state of undress, feeling vulnerable and discussing symptoms that you would not want disclosed to the world at large. You placed your trust in this healthcare professional (and their organization) to safeguard your privacy. Now, think how you would feel if this sensitive information was released. Would it make you feel less inclined to share your medical information? If so, you wouldn't be alone.

Recent studies have found that people are withholding information—sometimes critical information—from their healthcare providers because they are concerned that there could be a confidentiality breach of their records. This is not only a potential issue for the treatment of a specific patient; there are potential public health implications. An unwillingness to fully disclose information could delay a diagnosis of a communicable disease. This is especially true if the disease has an attached stigma.[9, 10]

Recent studies have found that people are withholding information—sometimes critical information—from their healthcare providers because they are concerned that there could be a confidentiality breach of their records.

This problem illustrates why it is so difficult to measure the true impact of breaches. What many organizations fail to remember is that the data they collect is about the relationship they have with those data subjects. As reports of medical record losses continue to pile up, the trust between medical providers and their patients is being eroded. The implications of this may be wider than practitioners anticipate.

An unwillingness to fully disclose information could delay a diagnosis of a communicable disease.

9 http://www.ncbi.nlm.nih.gov/pubmed/23975624
10 http://www.ncbi.nlm.nih.gov/pubmed/25059953

Your test results are in

(Viewing the data from different perspectives)

We wanted to see what impact perspective had on our results, to make sure we covered as many angles as possible. In the Methodology section, we talked about the criteria for incidents to be included in this report. A quick refresher: If the data subject was "patient," the data disclosed was "medical records" or the industry was "healthcare," the incident was included in this study. This means that we have incidents that may have different characteristics depending on how you look at them.

If the data subject was "patient," the data disclosed was "medical records" or the industry was "healthcare," the incident was included in this study.

To get a better idea of how incidents with each of these criteria differ, we separated them into three distinct points of view:

- The patient perspective – based on the data subject's relationship with the organization that disclosed the data
- The medical records perspective – based on the type of data lost (if it was medical records, even if other types of data were also lost)
- The healthcare industry perspective – based on the classification of the victim organization (NAICS 62)

In looking at each of these perspectives, keep in mind that they offer a view into the data based on their discrete, separately applied criteria. Some incidents will belong to more than one perspective (for example, if a medical record is disclosed and the relationship to the organization is "patient," that incident will be included in both perspectives).

When looking at this data we found some common elements that were present regardless of which of these three criteria resulted in the inclusion of the incident.

Commonalities

In all three of the perspectives we present in this section, the top three Actions remain Physical, Error and Misuse. Only the number of incidents change – the order or ranking does not. Figure 5 shows the three perspectives' Actions side-by-side (although the data is separately selected based on the criteria of each perspective). For efficiency, they are presented together to make comparison easier.

The Physical Action is typically something being stolen, as stated in the Action section. If the device was lost, it would be classified as an Error Action. Thieves are not only targeting electronic devices; many theft incidents involved paper documents or even X-ray films. Both of these are especially concerning, as you can't (easily) encrypt dead trees (or radiographic film). The instances of X-ray film theft have been thought largely to be for the recycle value rather than the data, but that doesn't excuse the organization from having to report a breach.[11]

In our section "The Nefarious Nine," we discuss how these actions are combined to describe the data in the DBIR.

11 http://www.databreaches.net/raleigh-clinic-says-x-rays-were-stolen-may-have-included-patient-information

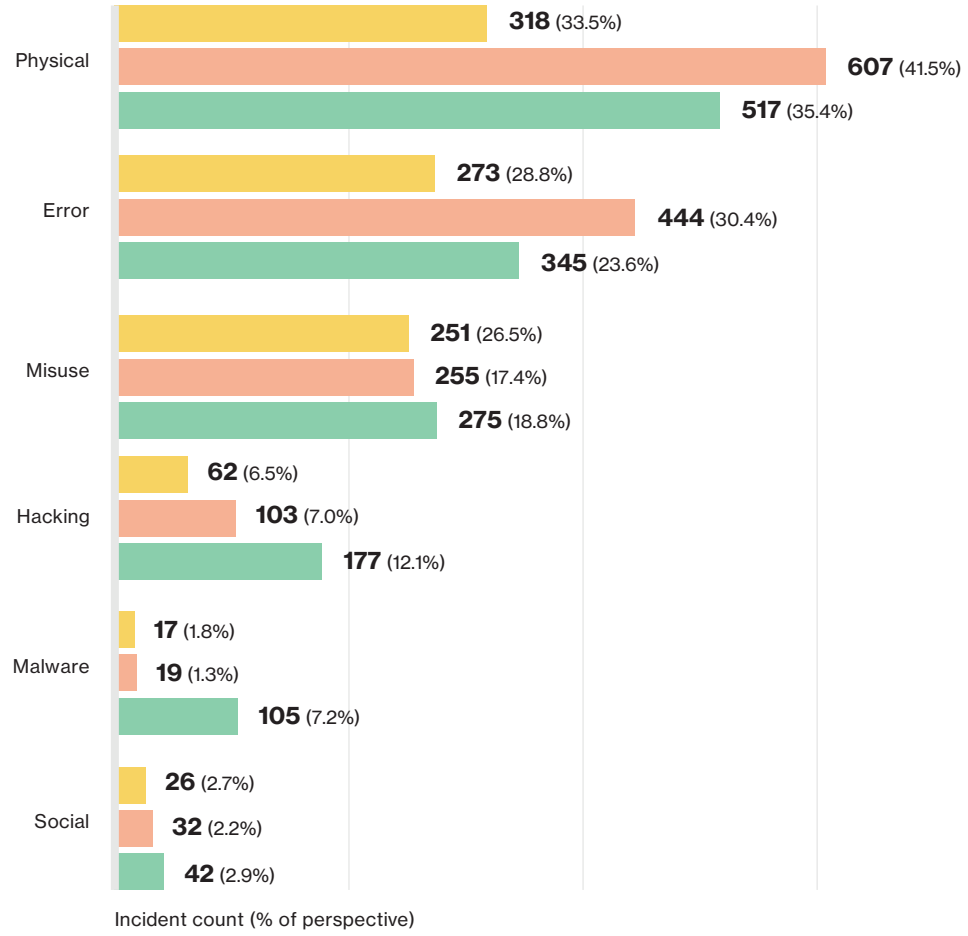


Figure 5.
Actions by perspective

Even when medical records are taken with malicious intent, it is frequently the associated PII that is targeted and used to commit various types of financial crime, including tax fraud and identity theft.

The patient perspective

We identified 970 incidents where the relationship between the data subject and the organization is defined as "patient"—the fewest number of incidents of the three perspectives. There were over 296 million records disclosed in this view (75% of incidents included a figure for number of records affected), and although the relationship is "patient," the data type lost is not always medical records. Sometimes, PII is breached that does not include medical information.

For those still wondering why PCI data is represented in this study, it is due to attacks on the POS systems that are processing co-payments in clinics everywhere. Payment card data is attractive to financially motivated actors, and POS compromises are not just a retail or accommodation industry problem. Even when medical records are taken with malicious intent, it is frequently the associated PII that is targeted and used to commit various types of financial crime, including tax fraud and identity theft. The medical information may not be what motivated the Threat Actor, but it provided a means to the end.

Now, when thinking of this relationship of "patient," you would expect all of the records to be disclosed from companies in the healthcare sector, amirite? Actually, 17 industries were represented, although healthcare had 70% of the incidents. The public sector (which is where the VA hospitals land) had 10% of the incidents, and the finance and insurance sector had 6%.

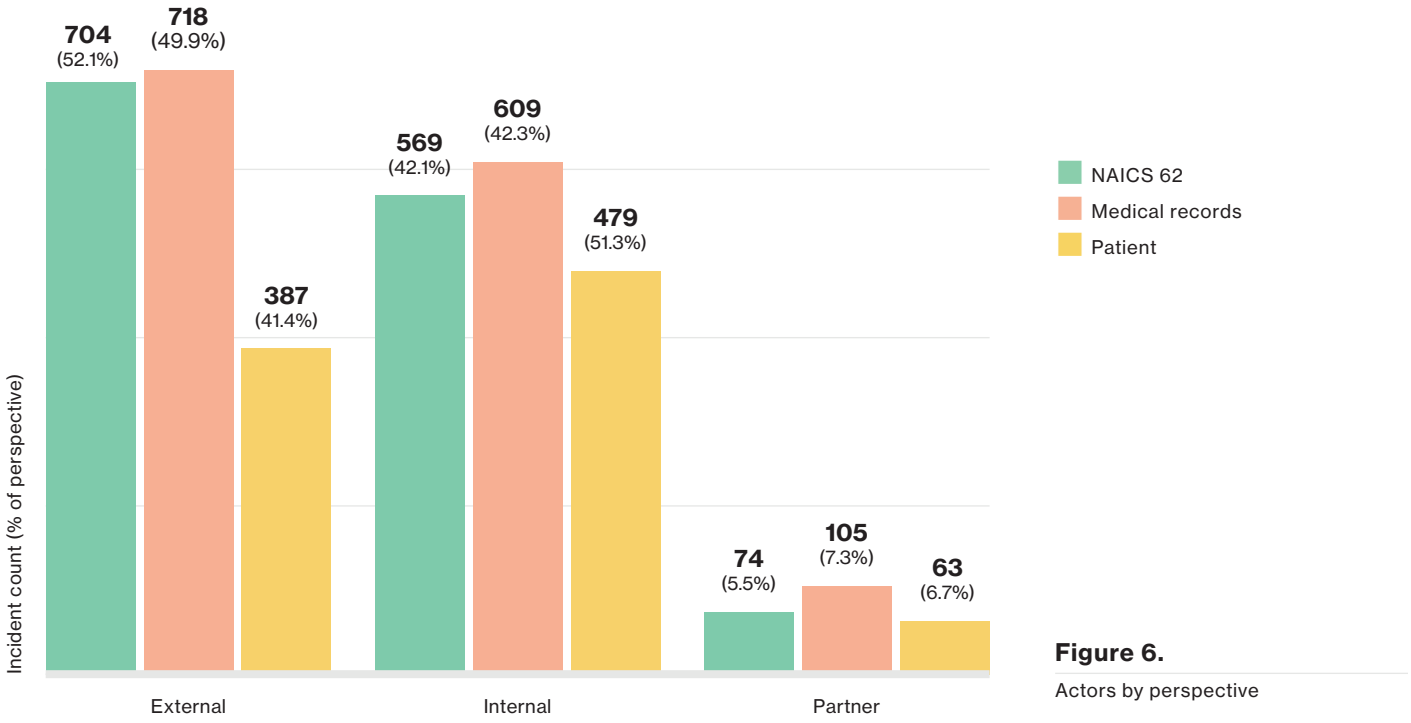


Figure 6.
Actors by perspective

The patient perspective (Figure 6) is the only perspective of the data where the insider incidents outnumber the External Actors. The number of Actors (if you add up the columns) exceeds the number of incidents—this is because VERIS allows for the presence of more than one Actor (and even Actions, Assets or Attributes) in an incident. Take the example of an External Actor recruiting an Internal Actor to use their access to steal patient data. Clearly, there are multiple Actors involved, and multiple actions—bribery by the External Actor and Misuse by the employee. This is why, in a number of these cases, you will find that the total number of incidents does not match the number of Actors or Actions.

The patient perspective (Figure 6) is the only perspective of the data where the insider incidents outnumber the External Actors.

The medical records perspective

When the record type disclosed is actually medical data, there were 1,523 incidents, with over 217 million known records disclosed (with the number of records reported for 79% of incidents). That's more incidents than the previous perspective, but fewer records disclosed. In this viewpoint, 18 industries are represented, with the healthcare sector accounting for the majority of the incidents (Figure 7). This is the only perspective that has representation of all of the industries that are present in the study. More simply stated, every industry—with the exception of utilities (NAICS 22) and management (NAICS 55)—lost medical data.

It was surprising to see that the healthcare industry was responsible for the smallest number of known records disclosed, compared to the other two perspectives and the overall PHI report dataset.

The healthcare industry perspective

The third lens we used to look at this dataset is when the top-level NAICS code is 62 (healthcare proper). In this viewpoint, there were 1,403 incidents and over 95 million records disclosed (with 76% reporting). It was surprising to see that the healthcare industry was responsible for the smallest number of known records disclosed, compared to the other two perspectives, and the overall PHI report dataset. This is largely due to some of the sizeable hacking breaches from other industries, but recent medical record breaches have shown that the healthcare industry has become a target for attackers.

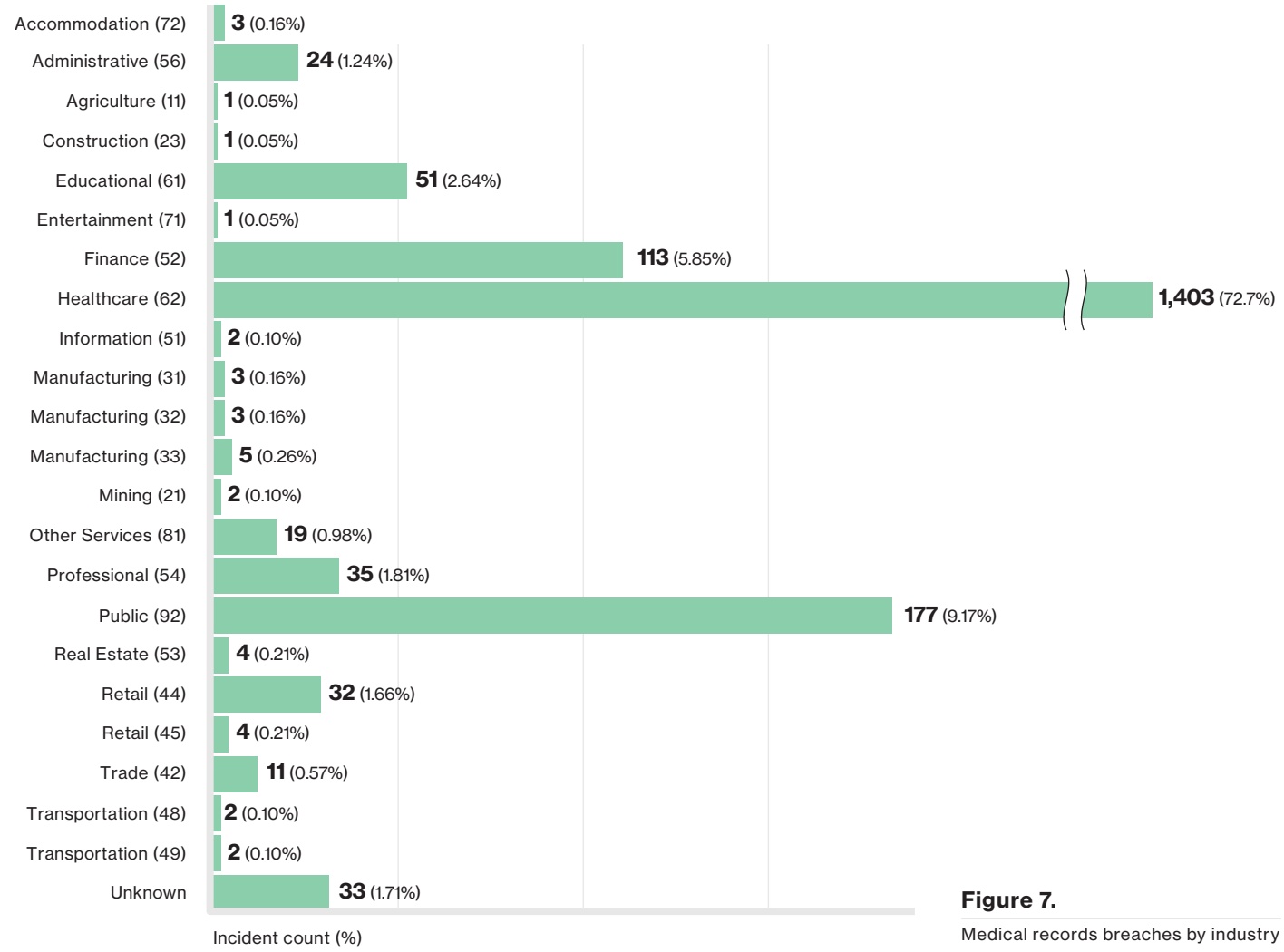


Figure 7.
Medical records breaches by industry

As with other perspectives, Physical, Error and Misuse make the podium, but hacking makes a strong fourth place—a more prominent showing than in the other two viewpoints. As stated earlier, criminals follow the money, and there are several areas where healthcare organizations collect payment, (e.g., co-pays, cafeterias, gift shops). Healthcare organizations are subject to opportunistic attacks on their POS, just like any other business accepting payment cards.

Nosing around the NAICS
In some of the data, we get more granular-level NAICS codes. This means we can look into the healthcare NAICS (those beginning 62) and see how the industry is divided—it spans a wide array of organizational types. We were interested to see if there were differences in the incidents for, say, a hospital (NAICS 622) and a nursing home (NAICS 623). Are the people attacking your local doctor's office (NAICS 621) the same people who are causing incidents in a social assistance facility (NAICS 624)?

First, let's look at the Actions for these incidents. If you are not familiar already with VERIS,[12] it is the framework we use to classify data—both from the cases we investigate and from partners. It is very good at taking a narrative, such as a forensic case report or a news article, and putting it into a standard format. This means that we know we are making apples-to-apples comparisons, regardless of the format the data initially came to us in.

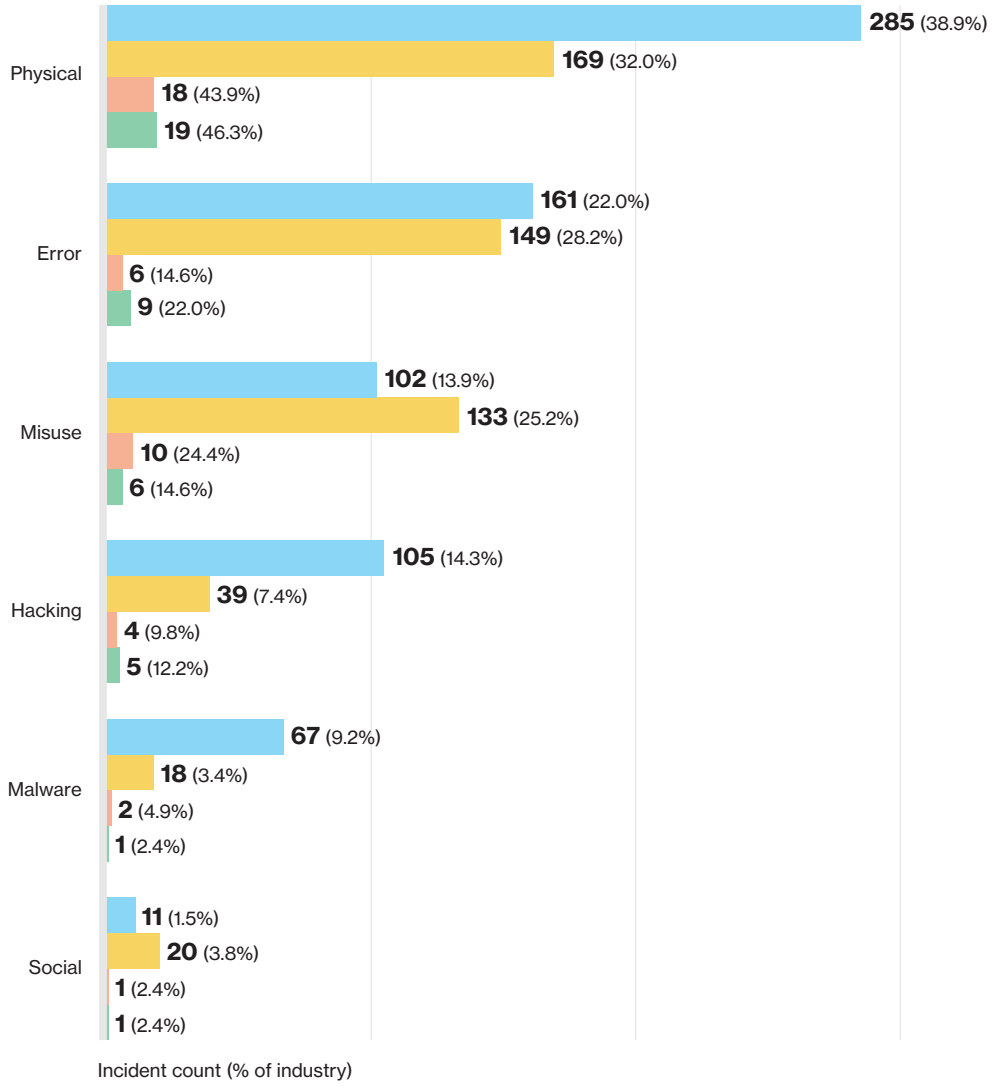
12 VERIS stands for the Vocabulary for Event Recording and Incident Sharing, and we give it away for free. Check out http://www.veriscommunity.net for more information.

Classifying incidents using VERIS
The incident classification section of the VERIS framework translates the incident narrative of "who did what to what (or whom) with what result" into a form more suitable for trending and analysis. To accomplish this, VERIS employs the A4 Threat Model developed by Verizon's RISK Team. In the A4 model, a security incident is viewed as a series of events that adversely affects the information assets of an organization. Every event is comprised of the following elements (the four A's):

- **Actor**–Whose actions affected the asset
- **Action**–What actions affected the asset
- **Asset**–Which assets were affected
- **Attribute**–How the asset was affected

It is our position that the four A's represent the minimum information necessary to adequately describe any incident or threat scenario. Furthermore, this structure provides an optimal framework within which to measure frequency, associate controls, link impact and many other concepts required for risk management.

We'll first look at the Actions to see what kinds of incidents organizations in each of the three-digit NAICS codes experienced.



The healthcare industry is divided into four separate three-digit NAICS codes:

- 621 – Ambulatory healthcare services (includes all types of doctor and dentist offices)
- 622 – Hospitals
- 623 – Nursing and residential care facilities
- 624 – Social assistance

Figure 8.
Actions by three-digit NAICS code

Clearly, some of the sectors in this industry experienced far more incidents than others. NAICS 621 and 622 combined accounted for 94% of the incidents between them. NAICS 621 surpassed half of the total on its own. That said, looking at NAICS 621, you see that the problem with Physical Actions (primarily theft of physical assets) far surpasses other Actions. For NAICS 622, the top three actions are much closer to each other.

To answer the second question about the people perpetrating these breaches, we can look at the Actors for each sector.

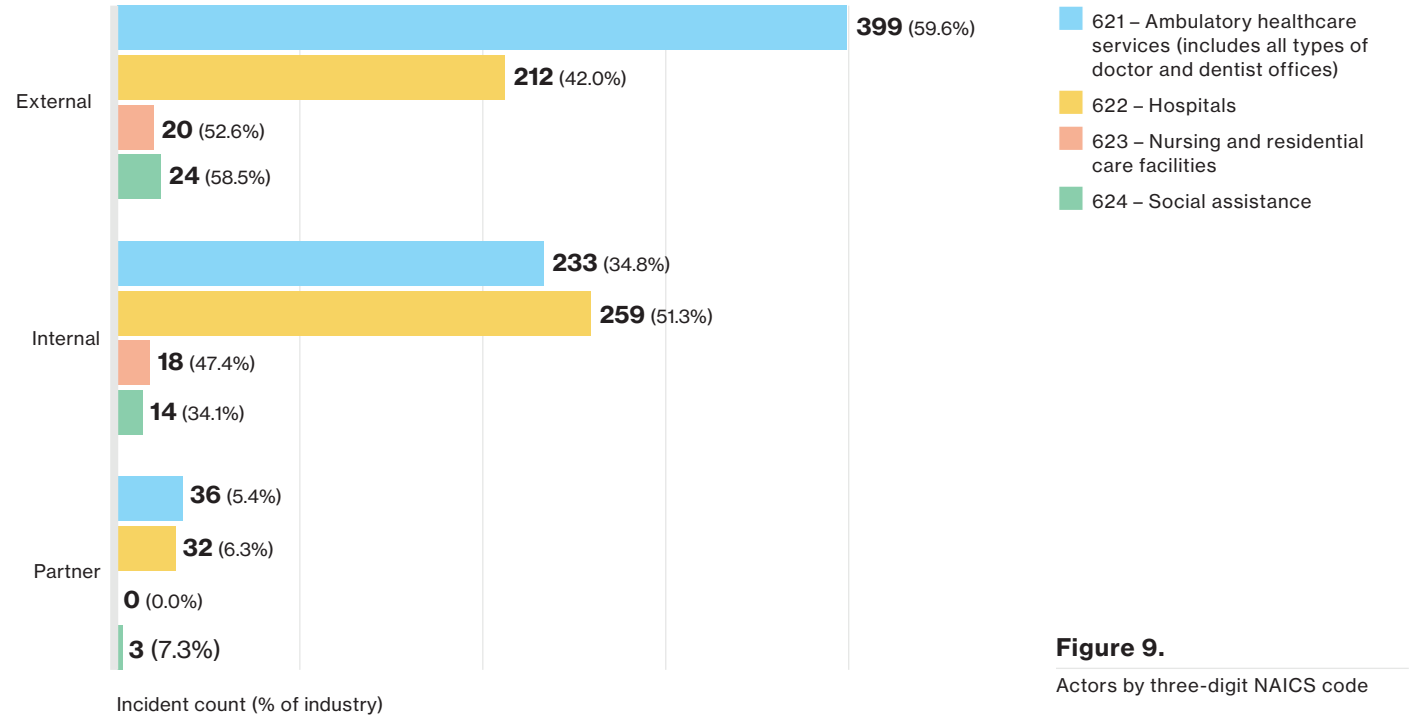


Figure 9.

Actors by three-digit NAICS code

For our ambulatory care providers, External Actors are having a field day. Given the preponderance of the Physical Attack vector in Figure 8, it stands to reason that there will be many of these thefts committed by people who are external to the organization. Hacking activity is also typically perpetrated by External Actors, and NAICS 621 shows this too.

Look at the Actors for the hospitals (622). While Physical Actions are most likely external parties, Error and Misuse are most frequently Internal Actors. Hospitals also have more Internal Actors committing breaches. Looking back at Figure 5, they have close to parity between the Physical, Error and Misuse Actions.

“Error” is the second most common action for NAICS 622. Since most of these are unintentional actions by employees, we wanted to take a closer look at the types of errors occurring. In Figure 10, we see the most common errors for each sector. These errors are prevalent across not just the PHI dataset, but also the DBIR. We’ll discuss them further when we talk about the incident classification patterns in the next section.

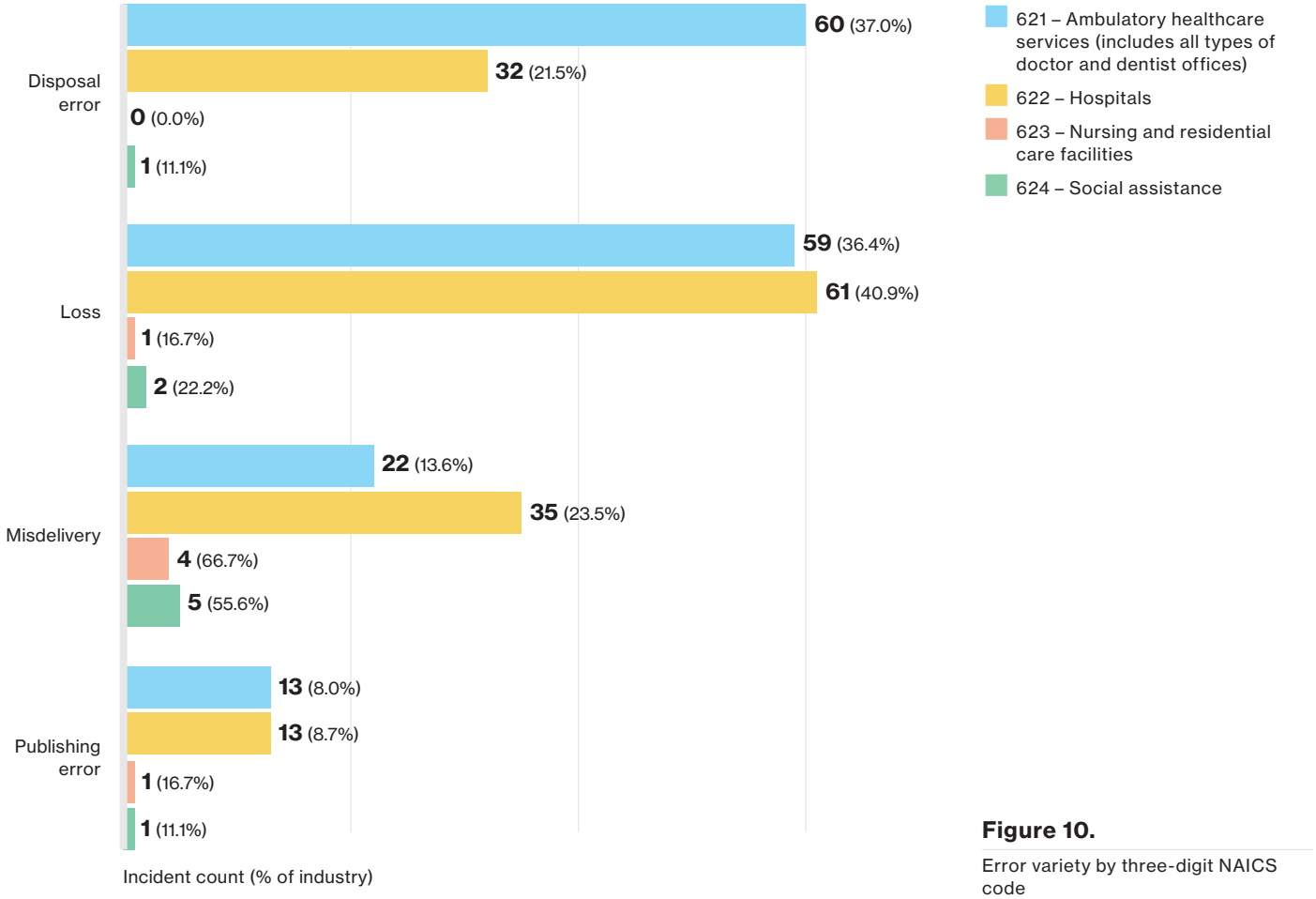
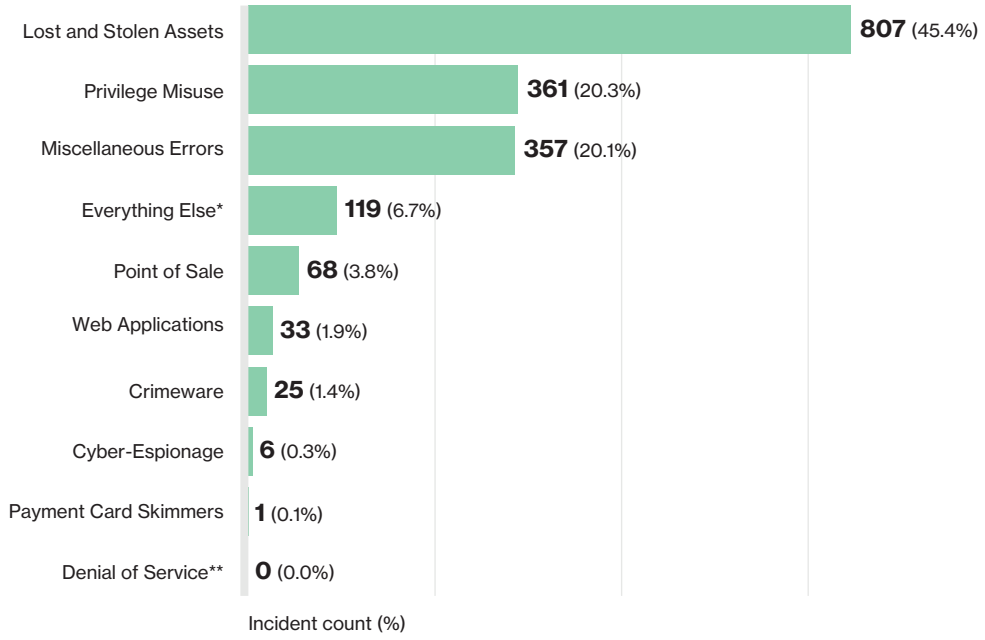


Figure 10.

Error variety by three-digit NAICS code

The “Nefarious Nine”

If you’ve been following the DBIR, you know that we can describe 96% of that data (in the overall dataset) using just nine incident patterns. In this PHI dataset, we find that we can describe 93% of the data with these same patterns. Just three of the patterns describe 85% of the incidents.



Even if organizations only encrypt a subset of their portable assets, it will reduce the overall risk of a breach on those assets that are not directly used for patient care.

Figure 11.

The “Nefarious Nine” patterns

* “Everything Else” is a catch-all category not included in the "Nefarious Nine" proper.

** Denial-of-service incidents without data disclosure are not included.

First among the patterns is Lost and Stolen Assets. It is frustrating to see this category return year after year because it’s one of the more easily solved problems. Encryption (particularly of portable devices) offers a figurative “get out of jail free” card since the data remains secure despite the loss of control over the asset.[13] In the vast majority of cases, this means the incident does not trigger a duty to report under most breach laws. However, in healthcare there is legitimate concern for any control that increases time to access data in an emergency situation. That said, not all of these assets had a function where emergency access is likely to be needed. We’ve seen researchers’ laptops holding significant numbers of records being lost or stolen that would not endanger patients if the asset had security controls. If there are areas where you can affect change in the risk profile,

13 Assuming the encryption passphrase is not written on a sticky note and lost along with the asset – don’t laugh; we’ve seen that.

they should be explored. Even if organizations only encrypt a subset of their portable assets, it will reduce the overall risk of a breach on those assets that are not directly used for patient care.

Privilege Misuse—when people who have legitimate access to the networks and systems of an organization use that access to do "bad things"—is driven by a variety of motivations. A common scenario is the "snooping employee," and the most obvious case is curious staff members looking at medical records of a celebrity or dignitary. External entities also employ strong financial incentives to recruit company employees to gain access to the rich array of information found in medical records. A good way to deter those who would be swayed to this life of crime is to educate them that people often get arrested for such actions.[14] Sanitized results of audits that catch people abusing their access are also useful to include in your awareness program.

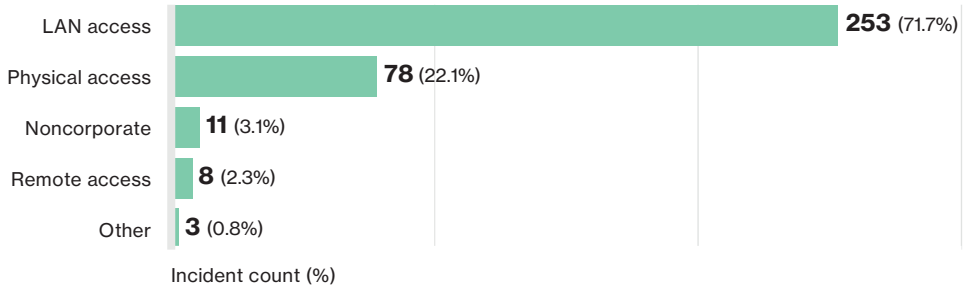


Figure 12.
Detail on the Misuse vector

Errors can be difficult for an organization to combat, and usually boil down to the need for checks along the way in processes that handle PHI. The most common errors are:

- **Loss**—Loss or misplacement of an asset.
- **Misdelivery**—Whether it is documents in the mail or electronic information in e-mail, it amounts to people getting data they weren't supposed to.
- **Disposal errors**—Primarily paper documents, but also electronic devices containing sensitive information.
- **Publishing errors**—When private information gets posted to an Internet-facing system and then is indexed by search engines.

For lost or even stolen devices, it is critical to ensure the organization has an easy way for people to report these incidents quickly. The sooner you know, the faster you can react to the breach.

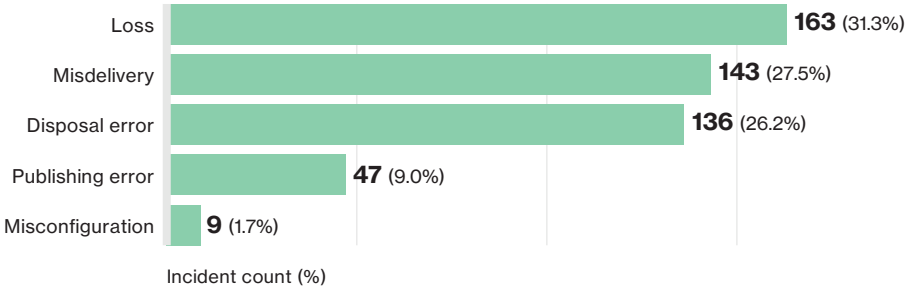


Figure 13.
Error Action top five

14 Don't believe us? Take a look for yourself: https://github.com/vz-risk/VCDB/issues?utf8=%E2%9C%93&q=label%3ABreach+PHI

Electronic misdelivery is more difficult to combat than paper documents. It is far too easy to accidentally address an e-mail to the wrong person—or even an entire e-mail distribution list. If the organization has data-loss prevention controls in place that can hold e-mails containing sensitive information from going out of the company while the "legit-ness"[15] is verified, there is at least a hope of catching these kinds of errors before they become an actual breach.

Most paper misdelivery incidents were from mass mailings where the envelope addresses and contents got out of sync, and nobody checked samples before sending it all off. This is a common problem in the public sector. It's difficult to hold a candle to the government when it comes to sending out large amounts of paper.

Disposal errors are sometimes the work of the third parties contracted to handle the disposal of paper and electronics not living up to their contracts. When you draft your legal requirements for the partnership, make sure to include penalties that are commensurate with the severity of a breach.

Web Applications attacks have also seen a steady growth over the years, whereas the Point of Sale attack appeared to not be an issue in 2013, but has since enjoyed a mini-renaissance.

The nine patterns over time

Looking at the trend of the nine patterns over time (Figure 14) shows how the mix has changed. The Internal Actor (Privilege Misuse) has been a constant companion for organizations in this dataset, but you can see how it really took off in 2013. Web Applications attacks have also seen a steady growth over the years, whereas the Point of Sale attack appeared to not be an issue in 2013, but has since enjoyed a mini-renaissance.

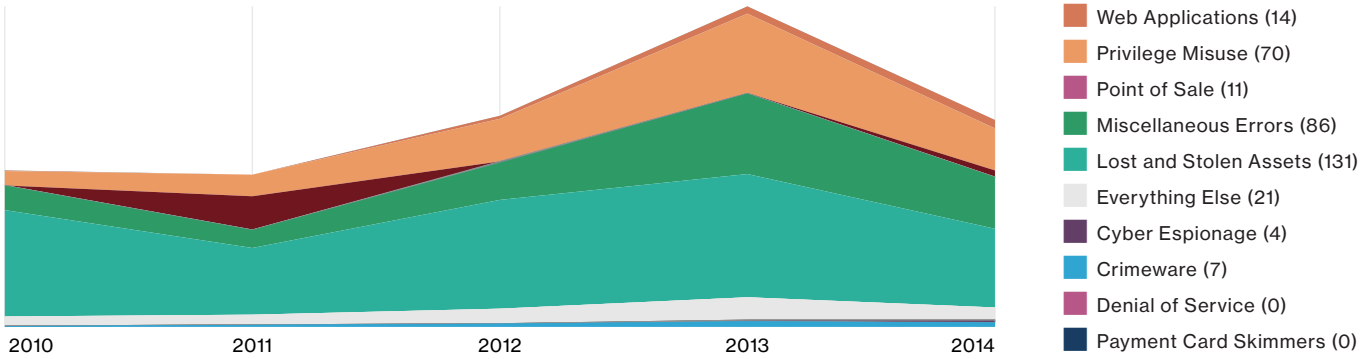


Figure 14.
Number of breaches by pattern over time

You can see the dominance of the Lost and Stolen Assets pattern through the years for this dataset. There hasn't been much progress on mitigating that risk, and it has seen a steady growth. The Miscellaneous Errors pattern also saw a jump in 2013, but has dropped some since then. However, we don't think people are going to stop making mistakes any time soon, so it is a safe bet that pattern will be in it for the long haul.

15 Yes, that IS a technical term.

Threat patterns under the microscope

(Attack graphs)

Attack graphs can seem a bit daunting if you've not seen them before. To the beginner, they may even seem like a big knot of kill chains[16] devoid of meaning. Fear not! We will explain the unexplainable and perhaps even help you learn to love them. If nothing else, you'll no longer look at attack graphs and think, "What a mess!"

First, gaze upon the attack graph for the PHI dataset.[17]

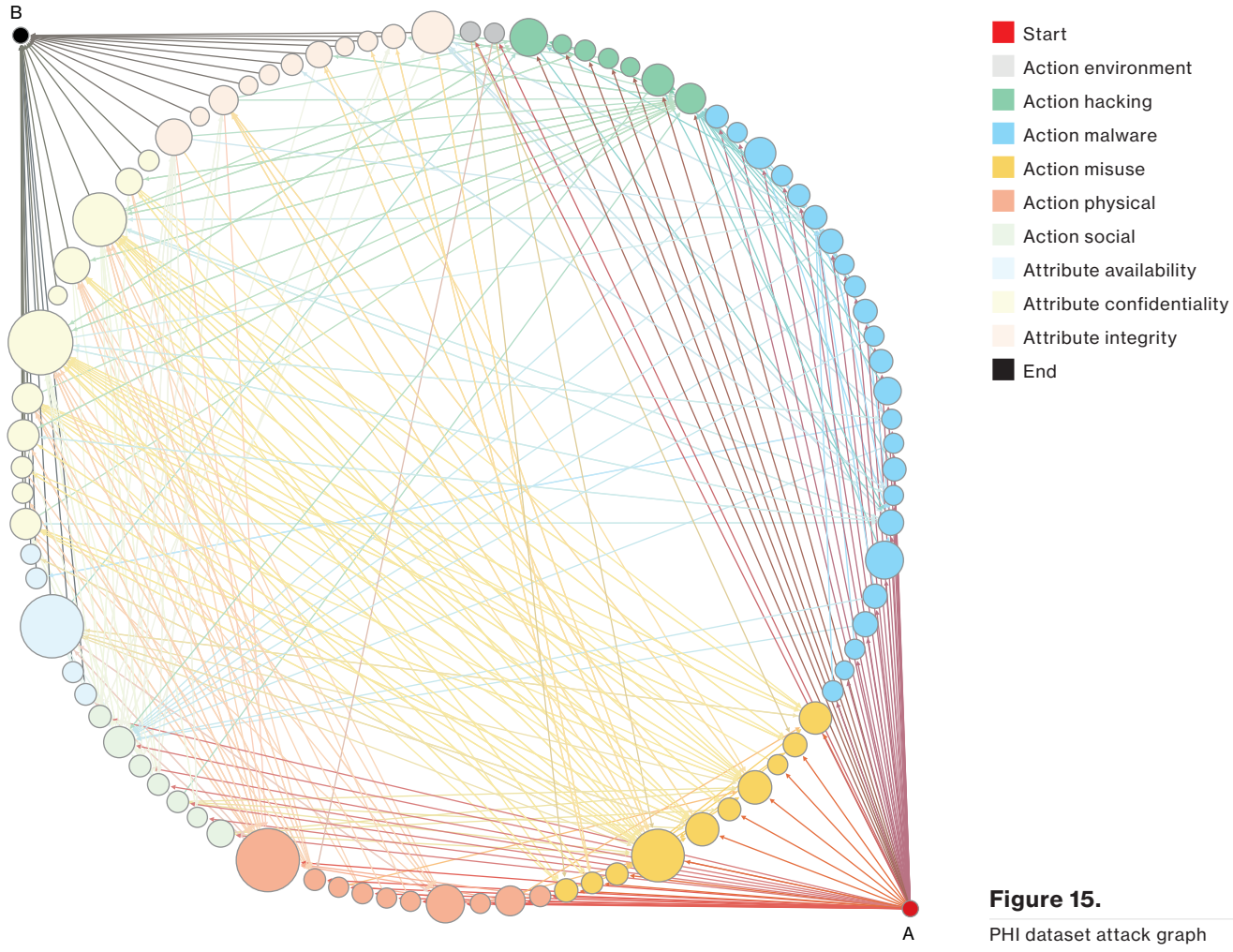


Figure 15.
PHI dataset attack graph

16 Or a plate of spaghetti. Who knew breach research can also make you hungry? Seriously, back to kill chains: http://www.lockheedmartin.com/us/what-we-do/information-technology/cybersecurity/tradecraft/cyber-kill-chain.html

17 Does it remind you of that nail-and-string art from the 1970s? No, me neither—I'm not old enough to remember that.

Attack graphs are really just illustrating how the perps are getting from point A to point B (sometimes having to take detours through points C, D and E if the direct route doesn't work for them). What are points A and B? Point A is where an Actor starts, and point B is their ultimate target—out the other end of the graph, clutching your data in his claw-like hands. Are you with us so far? Okay, then looking at Figure 15, that is just the collection of paths between the points taken.

The idea is that if you make it more difficult for the attacker to get to their ultimate goal, they'll move along to an easier target.

Attack graphs are important for defenders. We all know incidents aren't just a single point in time, but if you just think about incidents as a chain of events, you might miss the fact that attacks are more like a waltz around the dance floor[18] than they are a straight line. You have to mitigate *all* paths an attacker can take—not just the straight path from point A to point B. The idea is that if you make it more difficult for the attacker to get to their ultimate goal, they'll move along to an easier target.

The paths in the attack graph are risks. They show the action taken and the target that was successfully compromised. Listed in Table 2, you can see that the top 10 easiest-starting actions/ending compromises for attackers were:

1	Theft leading to loss of physical assets
2	Theft leading to breached medical records
3	Privilege abuse leading to breached medical records
4	Theft leading to breached personal information
5	Privilege abuse leading to breached personal information
6	Disabled physical controls leading to loss of physical assets
7	Disabled physical controls leading to breached medical records
8	Knowledge abuse leading to breached medical records
9	Phishing leading to altered behavior[19]
10	Data mishandling[20] leading to breached medical records

Table 2.
Most likely risks

Now that you know what the most common paths are, you need to tailor your mitigations to make it harder for the attacker to successfully traverse your organization—throw some roadblocks in their way! No organization is completely secure, but you want to put up as many obstacles for the attacker to overcome as you can within your existing resources. The biggest challenge is that you need to stop every way an attacker can get from that first action to their final goal. That's why information security is hard. Defenders must stop every avenue of attack; attackers only have to find one path that still lets them compromise their target.

The biggest challenge is that you need to stop every way an attacker can get from that first action to their final goal.

Prioritized mitigations

We all want to get the most bang for our buck. The less we spend on mitigations, the more we get to spend on holiday parties and quadcopters for the office. The attack graph doesn't just show us what the attackers do, it gives us an idea of what mitigations make attacking toughest for the bad guys. Figure 16 shows the mitigations in prioritized order according to how much more difficult it will make things for the attacker. Mitigate the first two alone, and you've made your organization over three times harder to be breached than one with no mitigations.

18 With spins and maybe even a dip!

19 "Alter behavior" is a VERIS term. It's what happens when an attacker's actions cause someone to do something different. Normally, it's something they shouldn't do, like visiting a malicious link, which can then result in malicious software being installed on their computer.

20 Data mishandling could be due to someone who loads data onto a portable drive to take home to work on over the weekend (against policy). It could also be caused by someone who is sending sensitive data to their personal e-mail as a method of exfiltration.

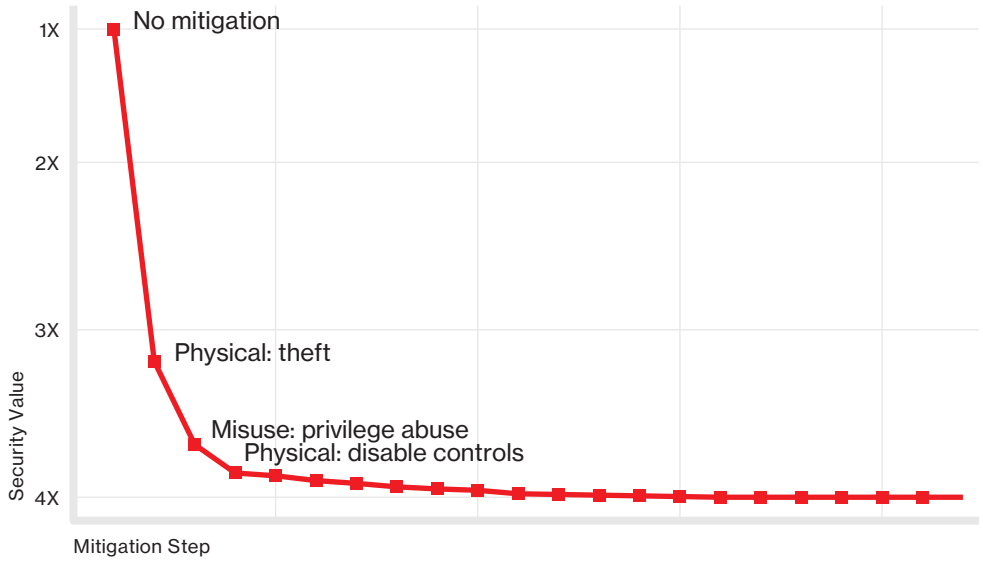


Figure 16.

Relative shortest path-length increase from mitigation

You can see from the figure above that when you take the first steps to make your environment more complex to attack, you get a big effect. But after those initial activities, you'll get less and less return for your efforts. Sure, you could keep applying mitigations, but the money might be better spent on ensuring those first few controls are properly implemented, that your operations team is well resourced to handle those breaches that do get through or maybe on that quadcopter for the office. You know, for morale.

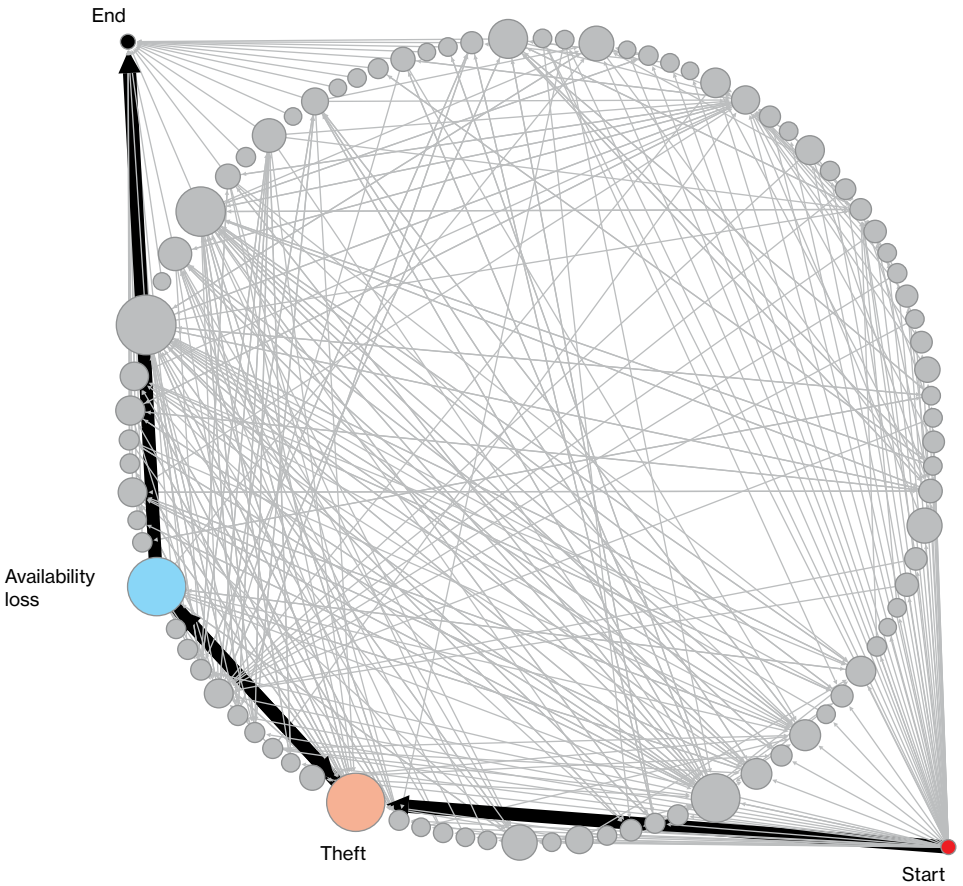


Figure 17.

Theft leading to loss path

Practical use case
So how can you use attack graphs in your work? Let's go over the most common path we saw in the PHI dataset—theft of Physical Assets leading to loss of a Physical Asset.

This is the shortest path in the attack graph for this dataset. To mitigate the attack shown in Figure 17, you'd be looking to put controls in place that stop the theft Action from successfully compromising the data on the asset stolen. So if we are looking at a laptop theft, encryption is a common control. You won't stop the theft from happening, and there will still be an availability loss because you no longer have the asset, but the data is no longer at risk.

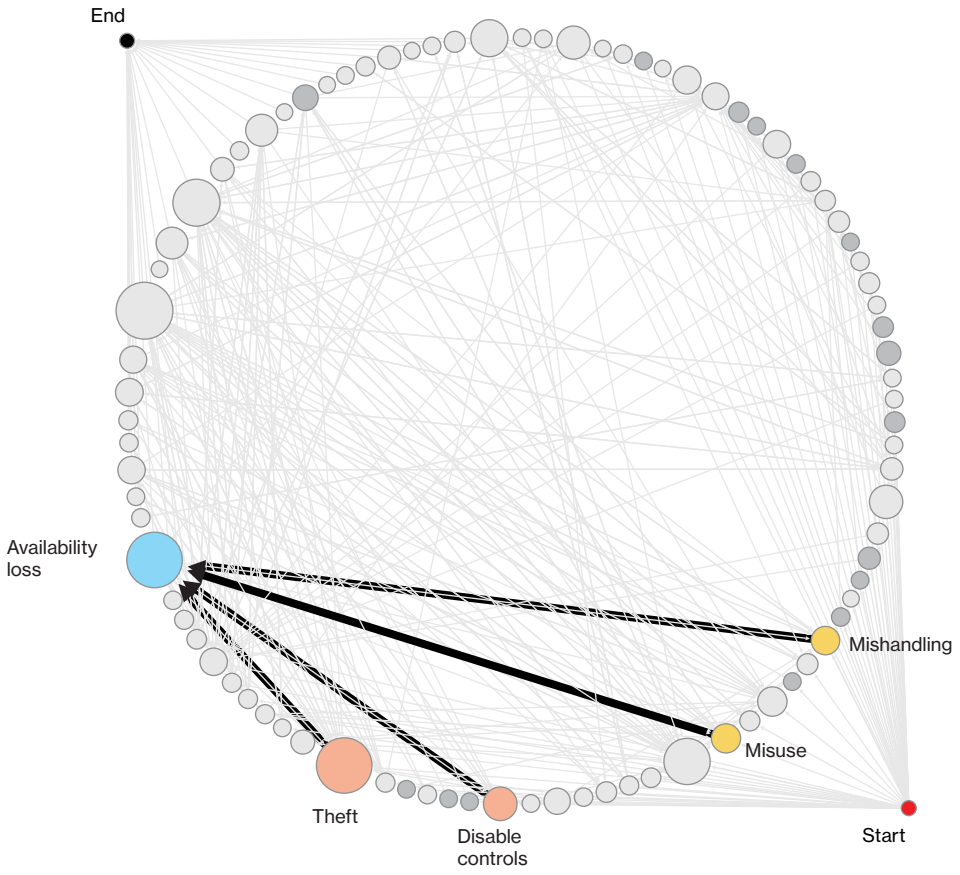


Figure 18.
Paths leading to loss

Figure 18 shows the other ways attackers got to the "availability loss" node, and illustrates that you have your work cut out for you, since you must mitigate each path. However, in doing this type of analysis, you can see where the controls need to be placed to mitigate each leg of the journey the attacker is taking.

If you want to do a few attack-graph what-if scenarios of your own, check out the DBIR attack graph tool.[21] Just watch the five-minute tutorial and you'll be ready to impress your friends and family with your attack graph prowess![22]

21 http://dbir-attack-graph.infos.ec/
22 https://securityblog.verizonenterprise.com/?p=7038

Triaging the outbreaks

(Timeline and discovery)

If a tree falls in the forest but there's nobody around to hear it, does it still make a sound? If you're pwned and can't see the alerts, do you still get a HIPAA fine? That is what comes to mind when we think of all the tools and data flooding into security operations centers (SOCs), but still see breaches going undiscovered for months.

There is a school of thought among information security professionals that you should simply work from the assumption that you've already been compromised. If that is the premise, how long does it take for you to detect a breach? In the DBIR, we showed what we call the "detection deficit," which illustrated a period of time (in this case, one week). Figure 19 shows the data for the overall DBIR dataset.

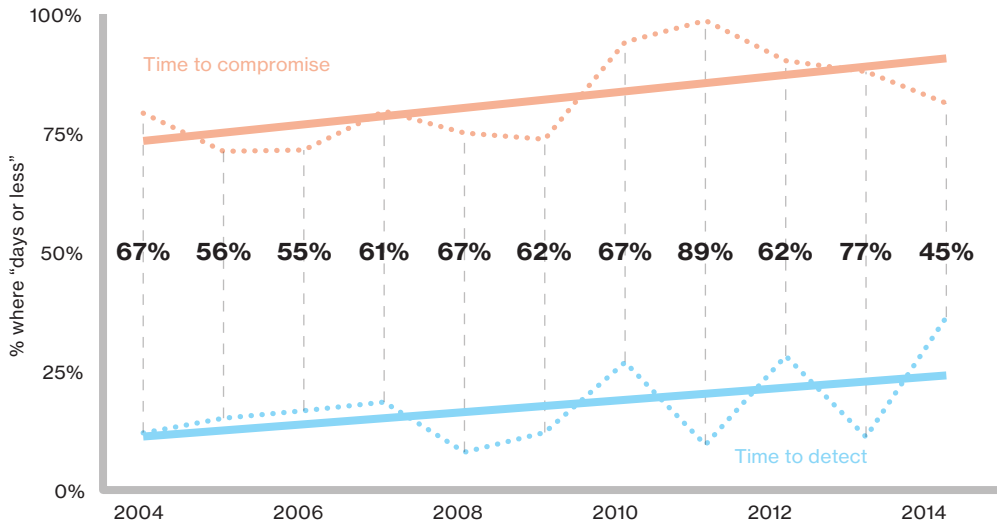
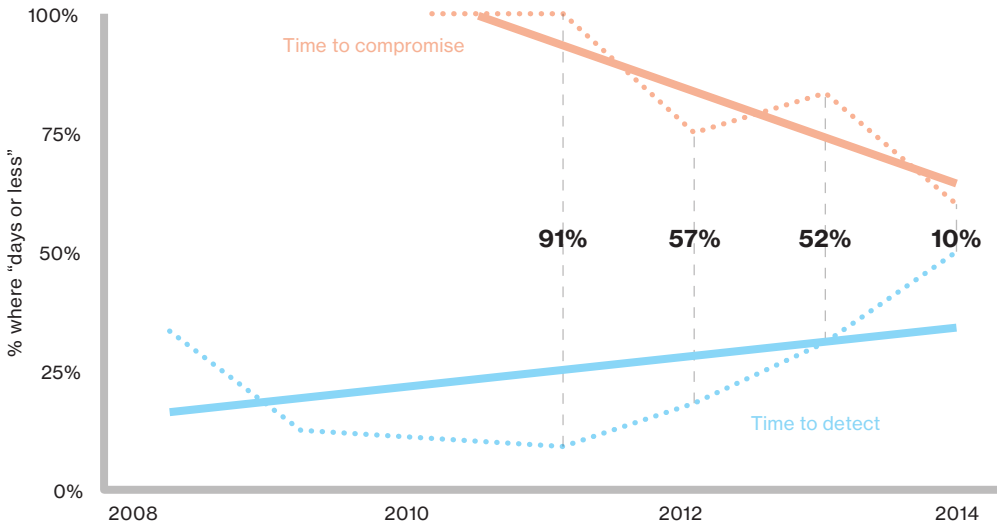


Figure 19.
2015 DBIR defender detection deficit

You can see that the bad guys (the top line) are (successfully) pwning people left and right, while the good guys (the bottom line) are having far less success at detecting these attacks (so many trees falling over with nobody to hear them). The space between is the "detection deficit," and while this past year showed us the best news we've seen since we started recording (45% is our smallest deficit to date), the news is not so great overall.

In contrast, look at the same graphic for this report's data (Figure 20):



In the healthcare industry in particular, unencrypted lost and stolen devices are a big problem.

Figure 20.
The PHI defender detection deficit

An inherent weakness in the PHIDBR dataset is that since it relies so heavily on the government reports and what reporters find interesting, we frequently don't get timeline information (the DBIR team would love to see the HHS Office for Civil Rights adopt VERIS). But for those incidents for which we have data, you can see that this graphic looks quite different from the DBIR. First, we see that 2014 had just a 10% deficit—down from 52% the year before. The most common method of discovery of these incidents was being found by an employee—which sounds great! However, when we look closer, we find that the reason they are being discovered by the employee is that they are the theft of physical assets. So yeah, if your laptop (thumb drive, etc.) gets stolen, you tell your boss about it pretty fast. This isn't the story we wanted to tell. We'd rather devices that were stolen were already encrypted and thus wouldn't end up in this dataset, but in the healthcare industry in particular, unencrypted lost and stolen devices are a big problem.

Figure 21 gives you the time to discovery of incidents over the past five years for the PHIDBR dataset. You can see how the incidents have been discovered over time, with most falling into the "months" category.

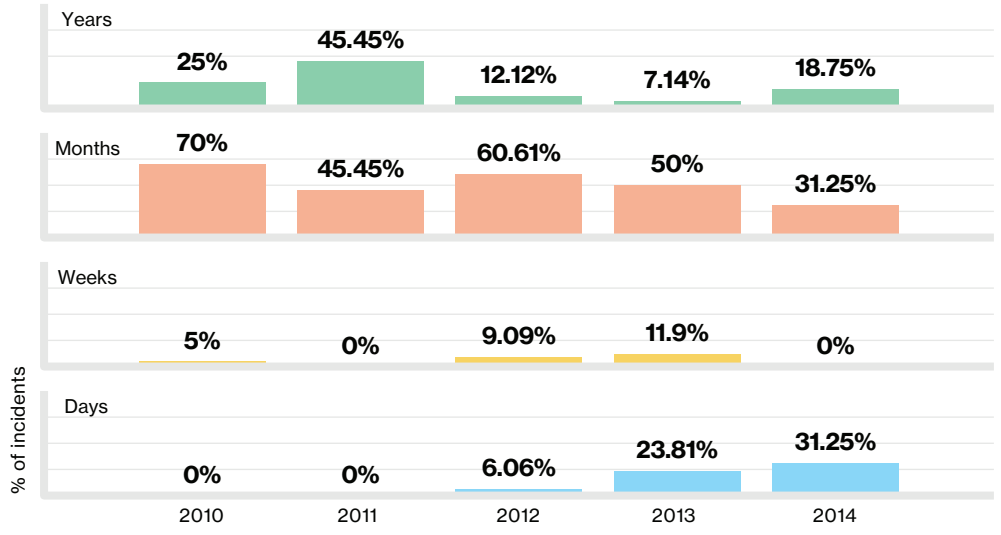


Figure 21.
2015 PHIDBR incidents time to discovery

The good news here is that we are seeing that over time, the "years" values are getting smaller. But there are still far too many incidents that are taking "months" to discover. We were curious to find out whether those incidents that took years to discover had anything in common. And yes, incidents in this dataset that took years to discover were over three times more likely to be caused by an insider abusing their LAN access privileges, and twice as likely to be targeting a server (particularly a database). How does this compare to the DBIR dataset? We found that while the DBIR incidents that took years to discover also tend to be Insider Misuse incidents, they are more frequently against either paper documents or media. The fact that both of these datasets (the PHIDBR and DBIR combined data is well over 100,000 incidents strong) show that the incidents that take the longest to detect are those being perpetrated by the organization's trusted insiders. This really speaks to the need for detective controls that can uncover this type of behavior.

The incidents that take the longest to detect are those being perpetrated by the organization's trusted insiders.

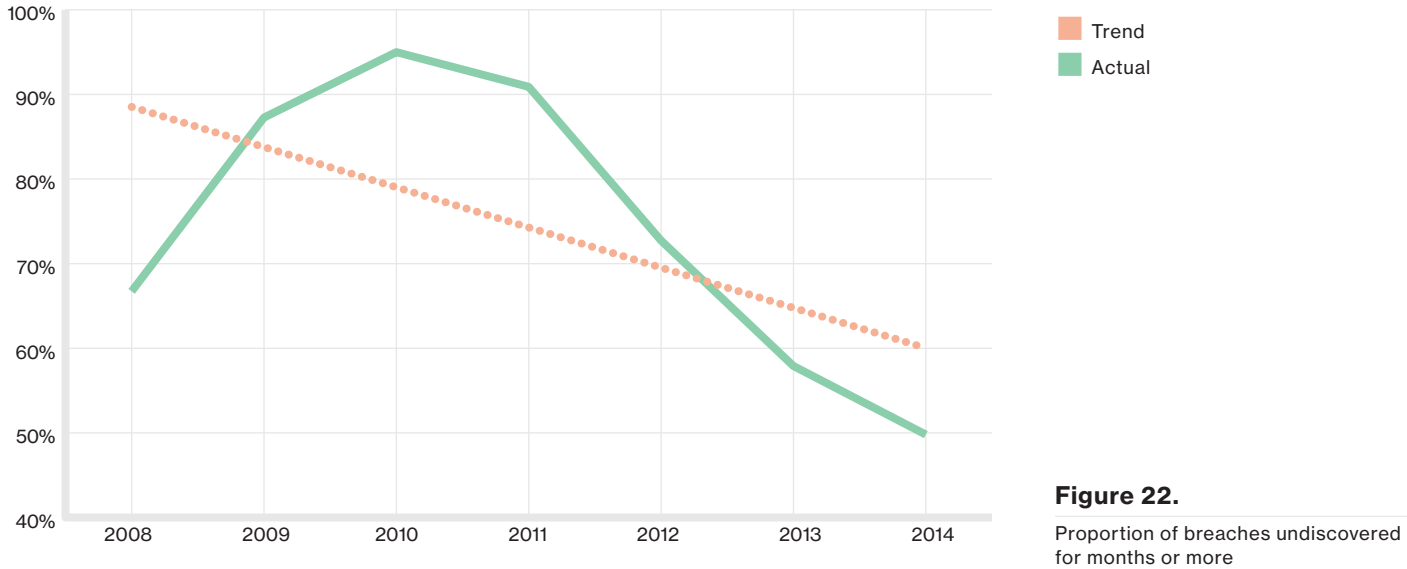


Figure 22.

Proportion of breaches undiscovered for months or more

This trend of discovering breaches more quickly is illustrated in the above graphic, Figure 22. The trend line is clearly sloping down, and you can see that the past two years in particular have shown improvement.

Incidents in this dataset that took years to discover were over three times more likely to be caused by an insider abusing their LAN access privileges, and twice as likely to be targeting a server (particularly a database).

In one memorable incident from the VCDB, a medical center had contracted a third party to handle disposal of its paper documents. Unfortunately, those documents ended up flying out of the vehicle as it drove down the road—to the extent that witnesses described the scene as "looking like a blizzard of white paper had struck the area."[23] Adding to this comedy of errors, an inmate work crew was dispatched to pick up the papers as part of their regular road cleanup in the area—what could possibly go wrong? Clearly, this is not how organizations would prefer to find out

23 http://www.wsbtv.com/news/news/local/medical-records-scattered-across-gwinnett-county-r/nbj8z/

about their incidents. This tree-based snowstorm would have been classified as having a third-party discovery method. And as you can see from the top line in Figure 23, it is the most common method for the PHI dataset (and is strongly trending upward in the DBIR dataset as well).

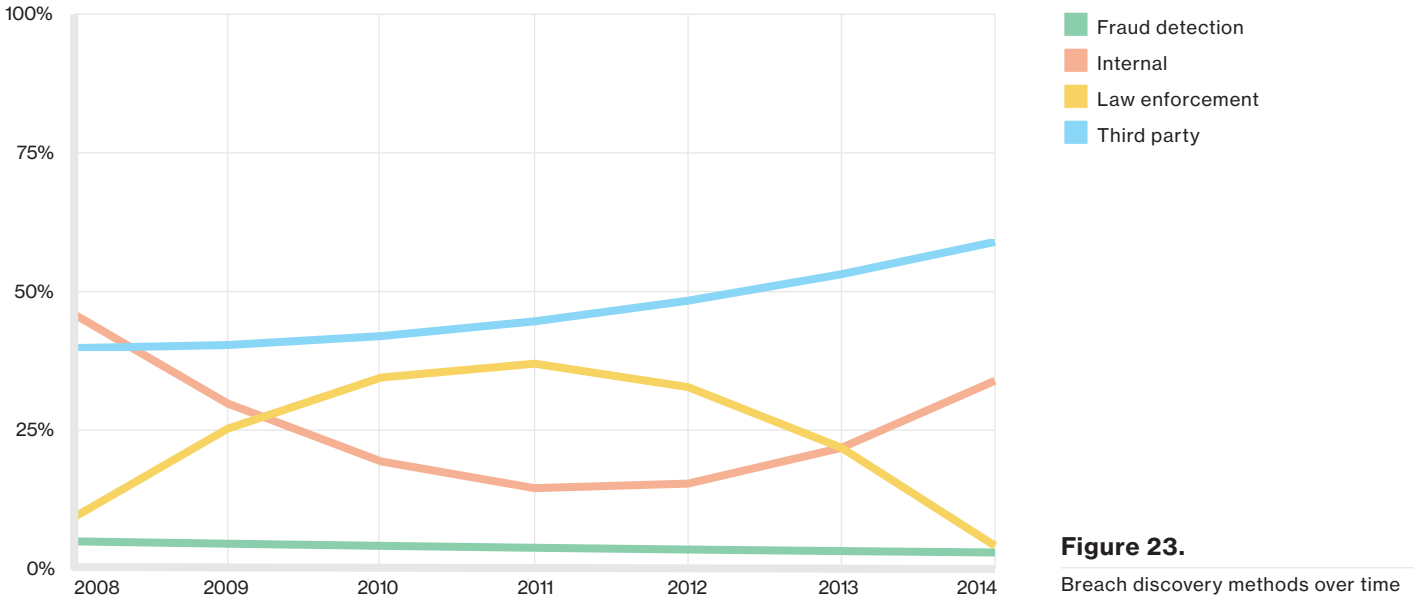


Figure 23.
Breach discovery methods over time

Fraud detection is most commonly associated with fraudulent use of payment card information that is tied to the Payment Card Industry's Common Point of Purchase (CPP) detection. Looking back at Figure 14, which presented the incident patterns over time, we see the expected correlation between this detection method and the rise and fall of POS intrusions. This plays less of a role in this report than the DBIR, as the payment cards (overall) have lower representation in this dataset. Law enforcement is also trending down as a discovery method.

Diagnosis and prognosis

News outlets, and even researchers like ourselves, regularly pontificate about devastating losses from breaches, such as the pilfering of the F-35 Lightning II fighter plans, or the seemingly ever-present large-scale credit-card attacks. Large-scale PHI-themed breaches in 2014 and 2015 have, however, helped put a spotlight on this highly personal loss of information confidentiality.

Detailed health records make it easier for criminals to engage in both identity theft and medical billing fraud—the former having direct impact on an individual or family, and the latter increasing healthcare costs for governments, organizations and individuals. Such private and potentially embarrassing (or worse) information can also be directly used against an individual, especially those in more sensitive positions. So what's the prognosis (doc)? As with any serious medical issue, there's good news and bad news.

PHI for half of the population of the United States has been impacted by breaches since 2009.

The bad news[24]

By just examining the HHS data alone, we can see that PHI for half of the population of the United States has been impacted by breaches since 2009.[25] Furthermore, the FBI issued a warning to healthcare providers in early 2015 stating that "the healthcare industry is not as resilient to cyber intrusions compared to the financial and retail sectors, therefore the possibility of increased cyber intrusions is likely."[26] And, as of January 1, 2014, all public and private healthcare providers must have adopted and demonstrated "meaningful use" of electronic medical records (EMR) in order to maintain their existing U.S. government reimbursement levels.

So, large PHI breaches have happened; more of our medical history is in electronic form than ever before; and the U.S. government itself is concerned about the safety of this information across all healthcare organizations. We've seen that criminals have shown both a desire for this data and the ability to whisk it away—seemingly at will. Unlike some measures one can take with a credit or debit card, there is virtually nothing we can do to safeguard our own medical information except rely on the individuals and organizations we've trusted to keep it secret.

As we've demonstrated in this study, it is not just the healthcare industry exposing this data, given that nearly all industries were represented in the dataset. Even those industries that have medical data but are not one of the HIPAA-covered entities must do their part to safeguard this type of information for the good of all.

24 In a recent study, 75% of folks want the "bad news" first (who are we to argue with the stats?): http://psp.sagepub.com/content/early/2013/10/30/0146167213509113.abstract#aff-1

25 169,700,764 records as of October 12, 2015

26 http://www.reuters.com/article/2014/04/23/us-cybersecurity-healthcare-fbi-exclusiv-idUSBREA3M1Q920140423

The good news
There are some bright spots to hold on to. Organizations with PHI are detecting incidents faster than other industries and closing the detection deficit. Enforcement (at least in the United States) in terms of fines and penalties is also increasing,[27] both in frequency and severity, which creates additional incentives for a healthcare provider or insurer to focus more on cybersecurity.

An ounce of prevention
Since there isn't much an individual can do, healthcare providers can use the information in this report—and the DBIR itself—to better proactively defend patient data from prying eyes. Just as a doctor might counsel a patient that there is no "miracle pill" and that they should eat better, exercise more and maintain a proper sleep schedule, the same is true for ensuring the confidentiality, integrity and availability of these records. Assess processes, procedures and technologies that affect the security of these records and prescribe a proactive treatment that will help the "cyber immune system" better protect the data entrusted to them.[28]

27 http://www.hhs.gov/ocr/privacy/hipaa/enforcement/examples/
28 We shy away from prescribing generic "best" practices, but take a look at the "Wrap-up" section in the 2015 DBIR (http://verizonenterprise.com/DBIR/2015/) for some potential "easy wins."

verizonenterprise.com