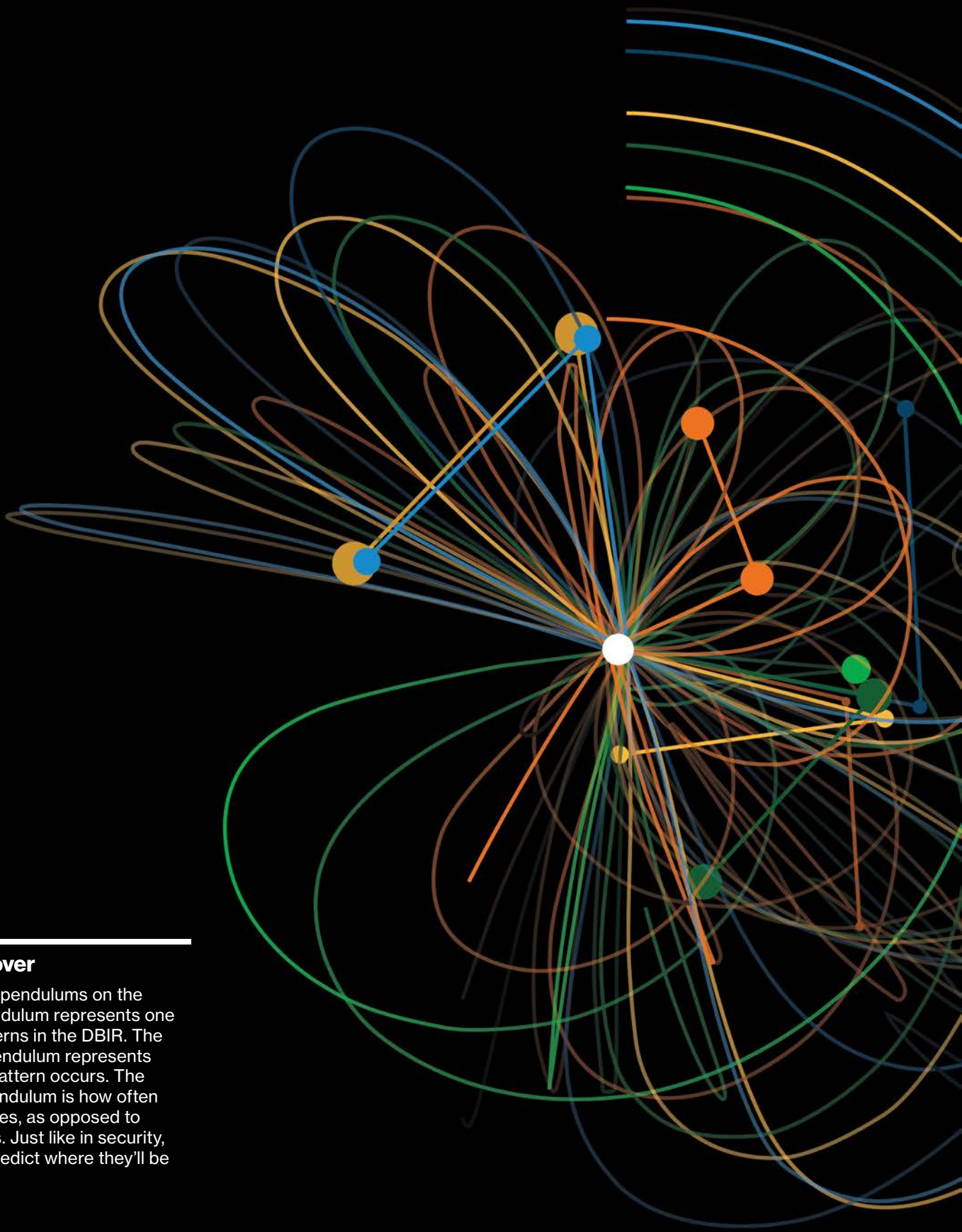

DBIR
2021 Data Breach Investigations Report
verizon



## About the cover

There are eight pendulums on the cover. Each pendulum represents one of the new patterns in the DBIR. The weight of the pendulum represents how often the pattern occurs. The length of the pendulum is how often they are breaches, as opposed to simply incidents. Just like in security, it's difficult to predict where they'll be in the future.

# Table of contents

# DBIR Master's Guide

**Hello first-time reader, and welcome to the 2021 Data Breach Investigations Report (DBIR). We have been creating this report for a while now, and we appreciate that all the verbiage we use can be a bit obtuse at times. We use very deliberate naming conventions, terms and definitions and spend a lot of time making sure that we are consistent throughout the report. Hopefully this section will help make all of those more familiar.**

## VERIS resources

The terms "action," "threat actor" and "variety" will be referenced often. These are part of the Vocabulary for Event Recording and Incident Sharing (VERIS), a framework designed to allow for a consistent, unequivocal collection of security incident details. Here is how they should be interpreted:

**Threat actor:** Who is behind the event? This could be the external "bad guy" who launches a phishing campaign or an employee who leaves sensitive documents in their seat back pocket.

**Action:** What tactics (actions) were used to affect an asset? VERIS uses seven primary categories of threat actions: Malware, Hacking, Social, Misuse, Physical, Error and Environmental. Examples at a high level are hacking a server, installing malware or influencing human behavior through a social attack.

**Variety:** More specific enumerations of higher-level categories, e.g., classifying the external "bad guy" as an organized criminal group or recording a Hacking action as SQL injection or brute force.

**Learn more here:**

- **github.com/vz-risk/dbir/tree/gh-pages/2021** includes DBIR facts, figures and figure data
- **veriscommunity.net** features information on the framework with examples and enumeration listings
- **github.com/vz-risk/veris** features the full VERIS schema
- **github.com/vz-risk/vcdb** provides access to our database of publicly disclosed breaches, the VERIS Community Database
- **http://veriscommunity.net/veris_webapp_min.html** allows you to record your own incidents and breaches. Don't fret, it saves any data locally and you only share what you want

## Incident vs. breach

We talk at length about incidents and breaches and we use the following definitions:

**Incident:** A security event that compromises the integrity, confidentiality or availability of an information asset.

**Breach:** An incident that results in the confirmed disclosure—not just potential exposure—of data to an unauthorized party.

## Industry labels

We align with the North American Industry Classification System (NAICS) standard to categorize the victim organizations in our corpus. The standard uses two- to six-digit codes to classify businesses and organizations. Our analysis is typically done at the two-digit level and we will specify NAICS codes along with an industry label. For example, a chart with a label of Financial (52) is not indicative of 52 as a value. "52" is the code for Finance and Insurance sector. The overall label of "Financial" is used for brevity within the figures. Detailed information on the codes and classification system is available here:

**https://www.census.gov/naics/?58967?yearbck=2012**

## Being confident of our data

Starting in 2019 with slanted bar charts, the DBIR has tried to make the point that the only certain thing about information security is that nothing is certain. Even with all the data we have, we'll never know anything exactly. However, instead of throwing our hands up and complaining that it is impossible to measure anything in a data-poor environment, or worse, simply making stuff up, we get to work. This year we continue to represent uncertainty throughout the report figures.

Figures 1, 2, 3 and 4 all convey the range of realities that could credibly be true. Whether it be the slant of the bar chart, the threads of the spaghetti chart, the dots of the dot plot, or the color of the violin chart, they all convey the uncertainty of our industry in their own special way.

The slant on the bar chart represents the uncertainty of that data point to a 95% confidence level (which is quite standard for statistical testing). In layman's terms, if the slants of two (or more) bars overlap, you can't really say one is bigger than the other without angering the math gods (and their wrath is terrible).

Dot plots are also frequently used, and the trick to understanding this chart is that the dots represent organizations. For example, if there are 200 dots (like in Figure 3), each dot represents 0.5% of organizations. This is a much better way of understanding how something is distributed among organizations and provides additional information than an average or a median. We added additional colors and callouts to make these even more informative this year.

Our newcomers this year are spaghetti and violin charts. They attempt to capture uncertainty in a similar way to slanted bar charts but are more suited for, respectively, data visualized over time and proportions of changes over a specific time period. For these charts, the darker area is more likely to be the correct value.

Let us know what you think of them.[1] We hope they make your journey through this complex dataset a little less daunting.

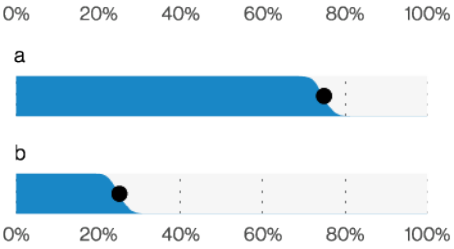



**Figure 1.** Example slanted bar chart (n=402)

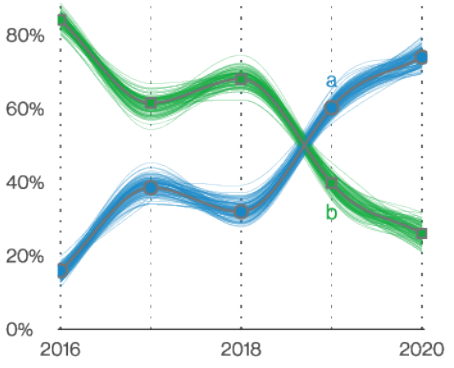



**Figure 2.** Example spaghetti chart

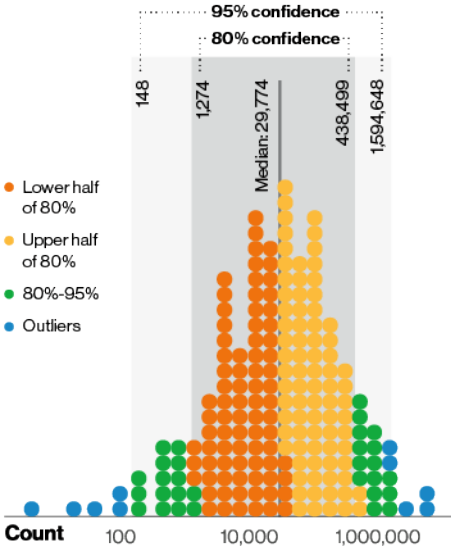



**Figure 3.** Example dot plot (n=672)
Each dot represents 0.5% of organizations

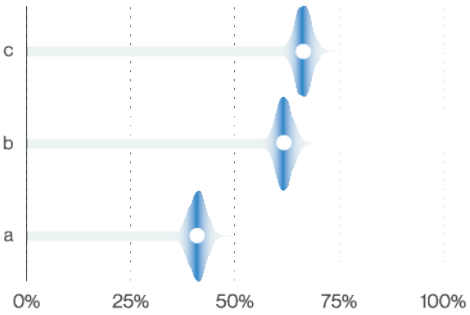



**Figure 4.** Example violin chart (n=581)

## Credit where credit is due

Turns out folks enjoy citing the report, and we often get asked how they should go about doing it.

You are permitted to include statistics, figures and other information from the report, provided that you (a) cite the source as "Verizon 2021 Data Breach Investigations Report" and (b) that content is not modified in any way. Exact quotes are permitted but paraphrasing requires review. If you would like to provide people a copy of the report, we ask that you provide them a link to verizon.com/dbir/ rather than the PDF.

## Questions? Comments? Upset there is no AR/VR version of the DBIR?[2]

**Let us know! Drop us a line at dbir@verizon.com, find us on LinkedIn, tweet @VerizonBusiness with #dbir. Got a data question? Tweet @VZDBIR!**

1 But only if you like them. Our figures guy is really thin skinned.
2 We REALLY want to make it happen!

# Introduction

Greetings! Welcome to the 2021 Data Breach Investigations Report (DBIR)! We always appreciate you, our readers, but this year we would like to say thank you for just showing up. Thanks for simply making it through the often frightening and always unpredictable dystopian wasteland that was 2020, and still having enough interest and energy to care about making the world a safer place. By the time you read this, it is devoutly to be hoped that we have moved on to a place of relative safety, somewhere beyond Thunderdome if you will.

Recent events around the world have been deemed by many to be sufficient cause to re-evaluate their priorities. In similar fashion, we have stepped back and taken another look at what we have been doing over the past few years. This exercise led to a revamp of our patterns, the creation of some shiny new ones and the recalibration of some others. It is our hope that doing this will increase awareness of where possible dangers lie, and how organizations may best avoid them. Perhaps we should say "probable dangers," since one lesson from 2020 is that many more things are possible than we might imagine. What is impossible is to accurately predict what those things might be. Therefore, we will not meddle with words like "possible," but will confine ourselves to what is "probable."

This year we analyzed 79,635 incidents, of which 29,207 met our quality standards and 5,258 were confirmed data breaches, sampled from 88 countries around the world. Once again, we include breakouts for 11 of the main industries, the SMB section, and we revisit the various geographic regions studied in the prior report to see how they fared over the last year. We also include our Center for Internet Security (CIS) Controls® recommendation mapping, because the world being unpredictable and uncertain doesn't mean your security strategy has to be.

As always, we wish to humbly say thank you to our 83 contributors, both old and new. This report would not be possible without you and we are always grateful for your continued support. Likewise, we thank you again, our readers, for continuing to share this journey with us.

Sincerely,
The DBIR Team

Gabriel Bassett
C. David Hylender
Philippe Langlois
Alexandre Pinto
Suzanne Widup

# Summary of findings

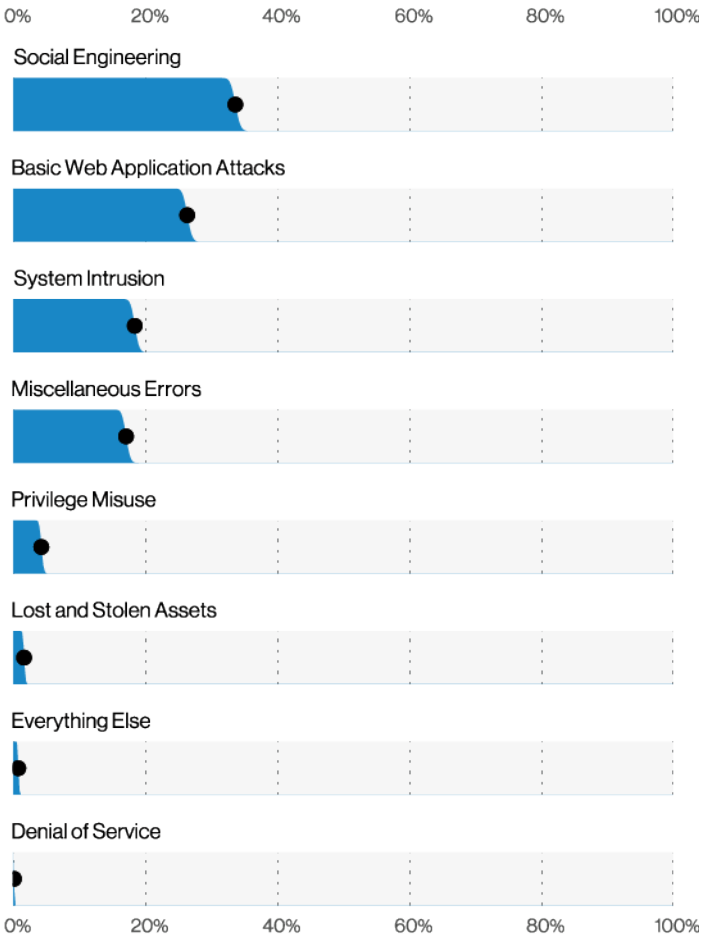



**Figure 5.** Patterns in breaches (n=5,275)

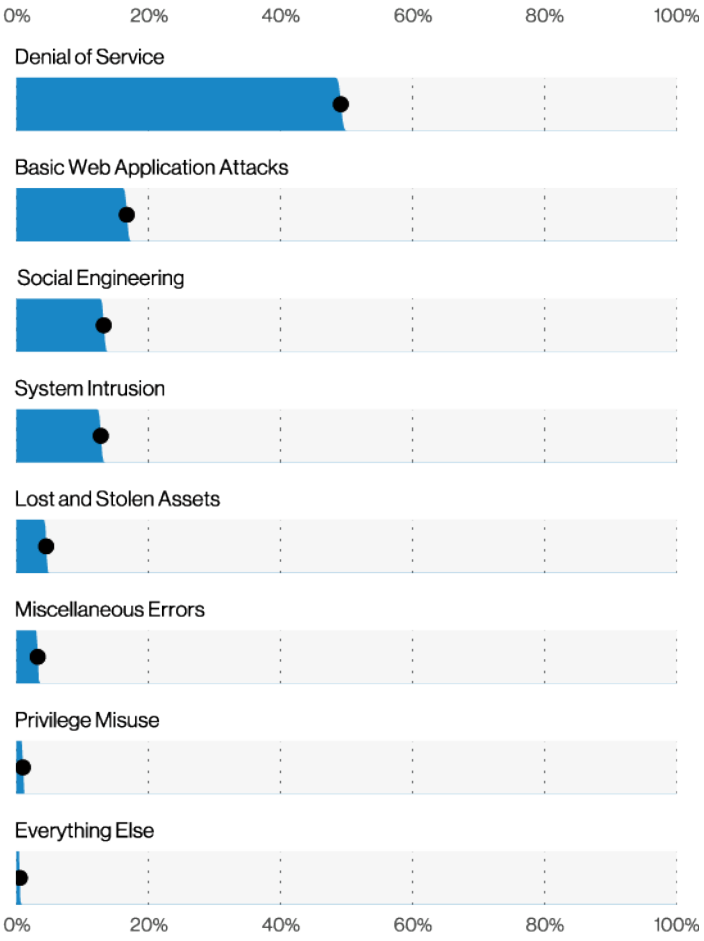



**Figure 6.** Patterns in incidents (n=29,206)

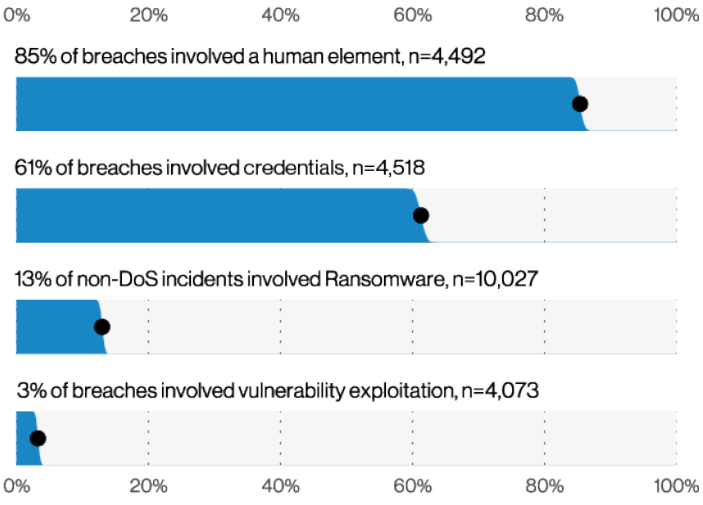



**Figure 7.** Select action varieties (n=4,073)

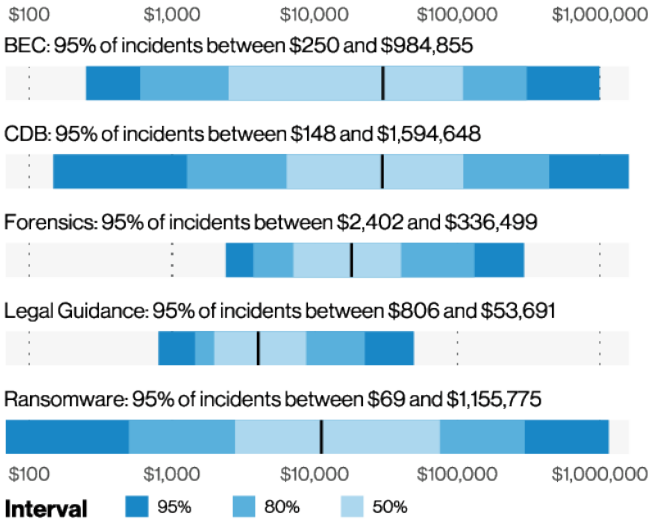
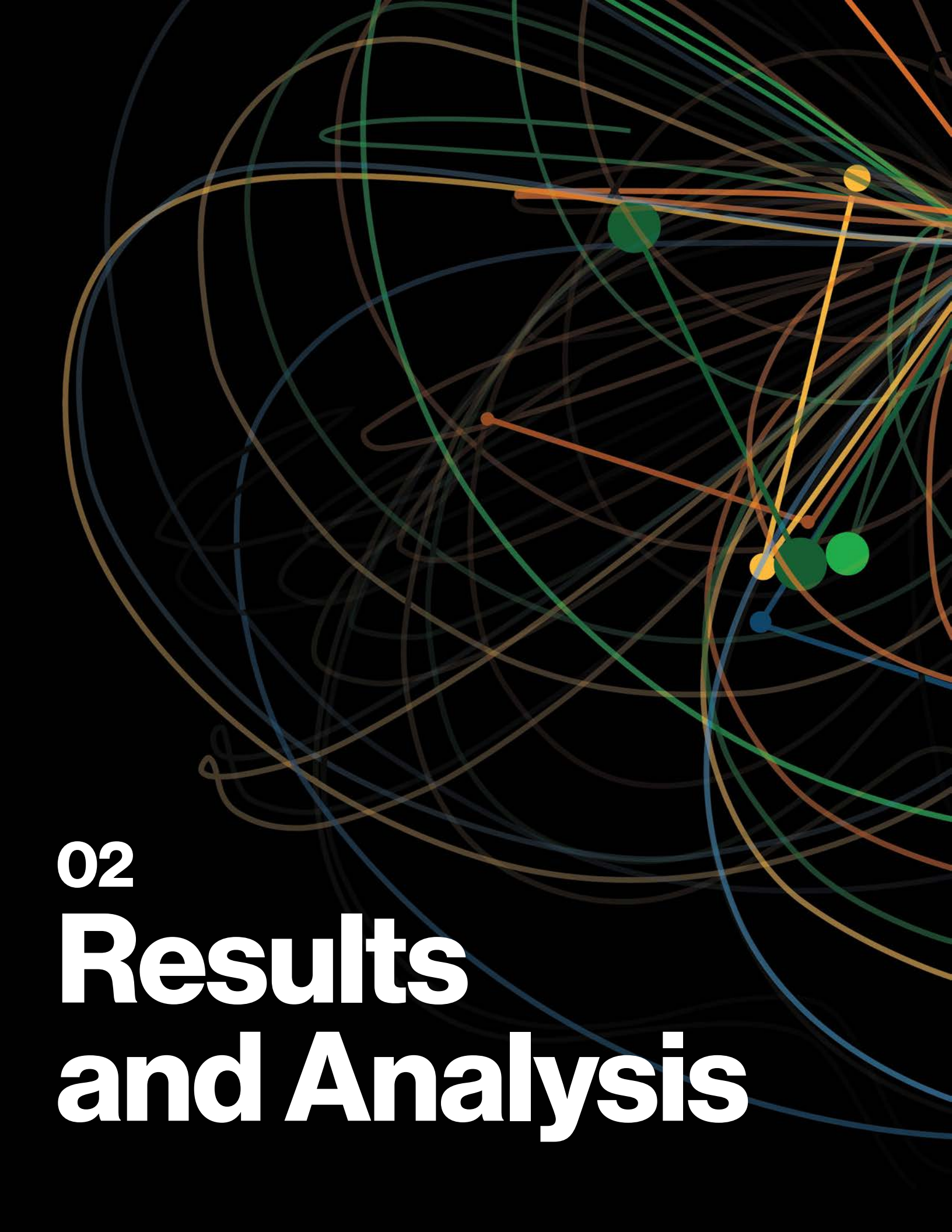



**Figure 8.** Select impacts of incidents



# 02
# Results and Analysis

# Results and Analysis: Introduction

The results found in this and subsequent sections within the report are based on a dataset collected from a variety of sources, including cases provided by the Verizon Threat Research Advisory Center (VTRAC) investigators, reports provided by our external collaborators, and publicly disclosed security incidents. The year-to-year data will have new incident and breach sources as we continue to strive to locate and engage with additional organizations that are willing to share information to improve the diversity and coverage of real-world events. This is a sample of convenience,[3] and changes in contributors—both additions and those who were not able to contribute this year—will influence the dataset.

Moreover, potential changes in contributors' areas of focus can shift bias in the sample over time. Still other potential factors, such as how we filter and subset the data, can affect these results. All of this means that we are not always researching and analyzing the same population. However, they are all taken into consideration and acknowledged where necessary within the text to provide appropriate context to the reader.

Having said that, the consistency and clarity we see in our data year-to-year gives us confidence that while the details may change, the major trends are sound.

**The DBIR is not in the business of prediction,[4] but it can go a long way to help you shape your response strategy in the face of an uncertain future.**

We believe it is fair to say that one of the primary lessons that 2020 had to teach us was that it is often futile to attempt to predict the future. However, not trying to predict it is not the same thing as giving up on scenario planning and preparing your organization for probable outcomes to the best of your ability. The DBIR is not in the business of prediction,[4] but it can go a long way to help you shape your response strategy in the face of an uncertain future.

Consider Figure 9 for instance; it's your run-of-the-mill DBIR chart with all the slanted bar-charted goodness, courtesy of our Misuse action varieties.[5] We have a few big things up top, and a lot of stuff near the end.

One valid way to interpret this is that the top bar or two are the norm of what may happen, namely in this example "Privilege abuse" and "Data mishandling." Those are the Action varieties that are understood to be so common that, if they were to cause a breach, someone (most likely on a bird website) would say, "That organization should have known better!"

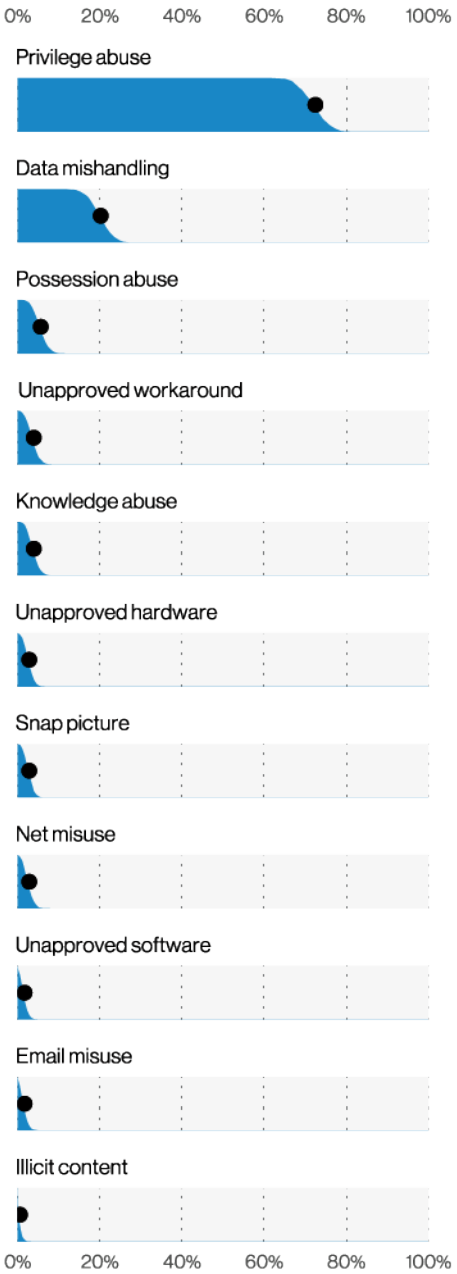



**Figure 9.** Misuse varieties in breaches (n=178)

3 Convenience sampling is a type of nonrandom sampling that involves the sample being drawn from that part of the population that is close at hand or available. More details can be found in our "Methodology" section.
4 Though we do suggest you put your money on "Trail Blazer" in the third.
5 Where are my insider threat fans at? Whoop whoop!

Suffice it to say, there's a great deal of inequality in the frequencies of the varieties shown. Those small bars are the extraordinary and uncommon attacks that could happen but are unlikely. If they were to cause a breach the victim would claim, "It was an advanced attack. There was nothing that anyone could have done."[6]

**The Gini coefficient is a measure of statistical dispersion most commonly used to represent the income or wealth inequality within a nation or other group of people.[9]**

However, if you take all those small bars on the Action varieties and add their breach frequencies together, you get Figure 10. Now it doesn't look quite so uncommon anymore, does it? In fact, in this example it appears that a breach is just as likely to be caused by one of our myriad exceptions as it is to be caused by our second most likely Action variety.

But does breach data always behave like this? Rather than show you lots of bar charts,[7] we're going to condense that concept down into a single number. Figures 11 and 12 show some data with different levels of inequality. We use the word "inequality" not by chance, but to introduce the fact that we can calculate the Gini coefficient[8] to represent this long tail behavior.

The Gini coefficient is a measure of statistical dispersion most commonly used to represent the income or wealth inequality within a nation or other group of people.[9] While it uses a lot of math none of us can be bothered with, it ultimately represents a completely equal outcome, where everyone has the same income (in other words, the "income per person" chart is a horizontal line), as a 0, and a world where one individual has all the income (in other words all we have on the chart is a huge vertical spike somewhere) as a 1.

Let's bring this closer to our subject matter by looking at some security-related data, like how often your SIEM generates a group of critical alerts that need immediate review. Anecdotally, you could attest that happens exactly "every time you are on-call," but humor us for a moment. In Figure 11, we generated some simulated example data that is perfectly smooth and looks horizontal on the chart—this one has an equality score of 0 (perfectly equal). Figure 12 has actual data representing the time interval between critical SIEM events, and it is extremely spikey.[10] It has a Gini equality score of 0.95, demonstrating a huge variation time between events. It's not just you: critical SIEM events fall into everyone's laps indiscriminately.

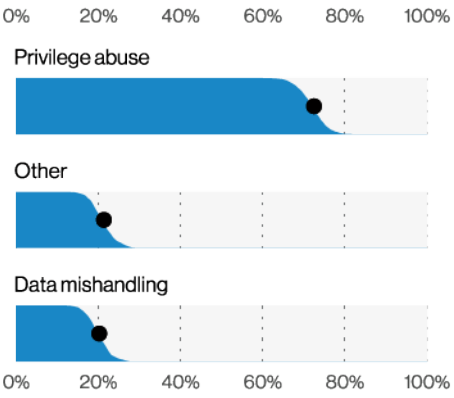



**Figure 10.** Top Misuse varieties in breaches (n=178)


**Figure 11.** Simulated time between SIEM events (n=1,335,343)

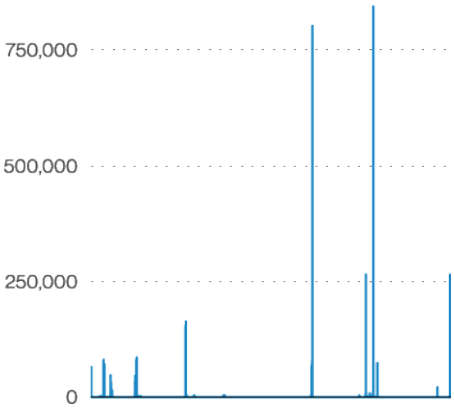



**Figure 12.** Time between SIEM events (n=1,335,343)

6 This report makes no claim about the validity of such a statement. Please refer to our official spokesperson and legal counsel. The data privacy of our readers is of the utmost importance to us.

7 And completely obliterate our page count budget.

8 https://en.wikipedia.org/wiki/Gini_coefficient

9 A less well-known fact is that the wish for wealth redistribution led to the phrase "Gini in a bottle." Not really, but it would have been cool if it did.

10 A technical term of art in Data Science, we assure you.

This complicated mathematical setup is to convey the reality that the DBIR data (incident and non-incident alike) is very unequal,[11] but at least we can measure it. Figure 13 shows the equality scores for Action, Actor, Asset, and Attribute varieties and vectors over the last seven years. The scores range from about 0.73 to 0.94, or as we would say here, "high." Breach data may seem likely to always be the same, but some varieties are more equal than others.

The reality is you don't need a crystal ball, a neural network or next-gen AI to tell you what the norm[12] is. You can do that for yourself and plan accordingly. On the other hand, you can't solution your way out of the long tail. It is made up of a legion of little things that happen only rarely—they are the exceptions to the norm. Well, maybe you can if you have enough money. And some organizations that are in critical roles to our society have no choice but to try to do so. But from a purely monetary value, if you look at what breaches cost in the Impacts section, it's not a wise use of your organization's resources to engineer solutions for every single possible exception.[13]

Armed with the knowledge of what is the norm and what is the exception, an ideally optimized solution would be to engineer solutions for the norm, and train your security operation teams to handle the exceptions. Turns out humans are very flexible problem-solvers, and most love a good challenge occasionally.

The next time we are up against a paradigm-shifting breach that challenges the norm of what is most likely to happen, don't listen to the ornithologists on the blue bird website chirping loudly that "We cannot patch manage or access control our way out of this threat," because in fact "doing the basics" will help against the vast majority of the problem space that is most likely to affect your organization.

Read on to learn what the normal actor has been up to for the last year, and pick out the areas where you can improve, against both the norm and the exception. Because the only way to predict the future is to change it yourself.

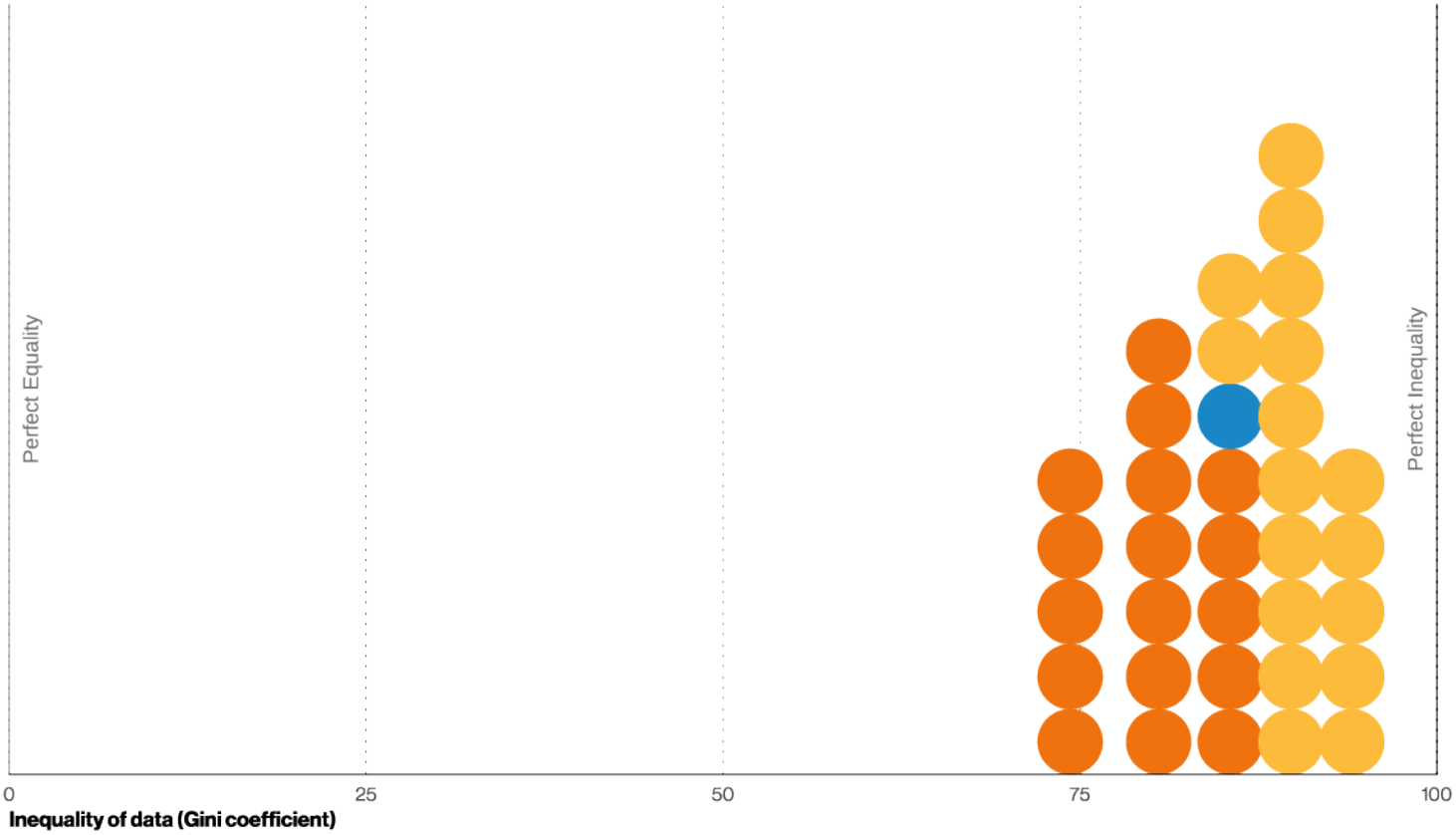



**Figure 13.** Inequality of enumerations in DBIR varieties and vectors for last 7 years

11 We deeply apologize to the junior U.S. senator from Vermont for the fact that the top 3% of varieties are responsible for 87% of the breaches.

12 You're reading the DBIR, and that is a great step in the right direction, if we may say so.

13 This argument does not consider potential incidents where loss of life or the security of individuals is concerned, as it would make no sense to assign a monetary value to that, and would, in fact, be callous and cruel.

# Actor

"All the world's a stage," and our threat actors "all have their exits and their entrances." We must admit that they seem to know their cues very precisely. However, at this point the analogy breaks down a bit, as rather than "playing their many parts"[14] we seem to keep viewing the same performance repeated ad infinitum, as if forced to endlessly re-watch a recorded musical theater presentation on a streaming service.[15]

It seems clear that our External actors are not giving up their close-ups, as they continue year after year to dominate the Actor types in breaches as illustrated in Figure 14. As a reminder to our readers, the Internal type shown here will include breaches in which both Misuse actions (where the mythical winged internal threats live in our taxonomy) and Error actions (the oopsies) occurred.

Of course, an External actor breaking into an organization by leveraging illicitly obtained credentials or other illegal access to pivot internally may initially resemble an internal threat before detailed incident forensics are engaged. But even though the call may be coming from inside the house, there is still a stranger on the line.

As in past years, financially motivated attacks continue to be the most common (Figure 15), likewise, actors categorized as Organized crime continue to be number one (Figure 16).

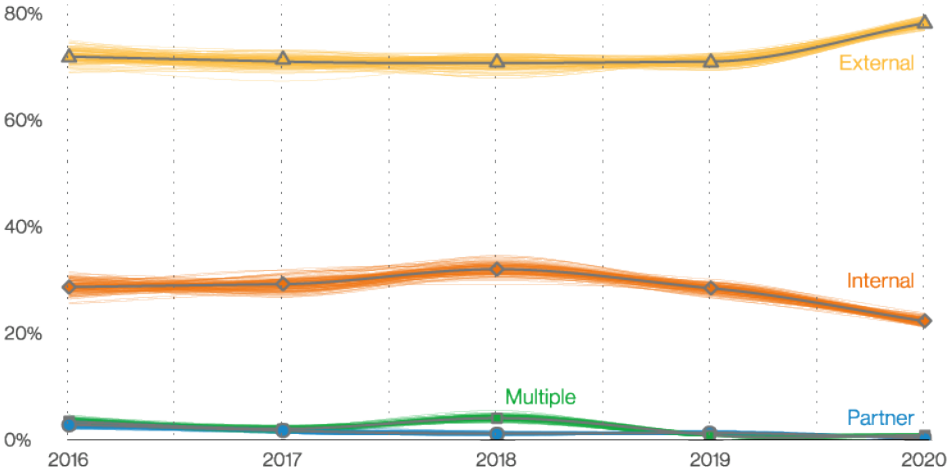



**Figure 14.** Threat actor over time in breaches

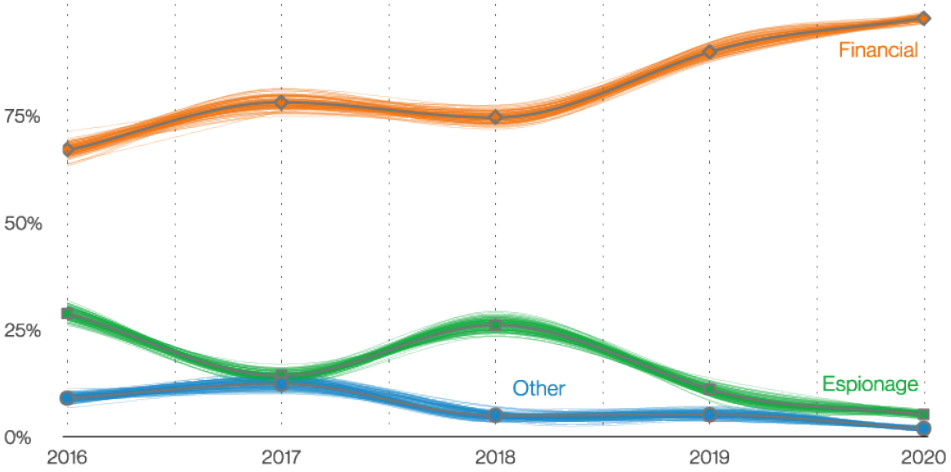



**Figure 15.** Top threat actor motive over time in breaches

**As in past years, financially motivated attacks continue to be the most common (Figure 15), likewise, actors categorized as Organized crime continue to be number one (Figure 16).**

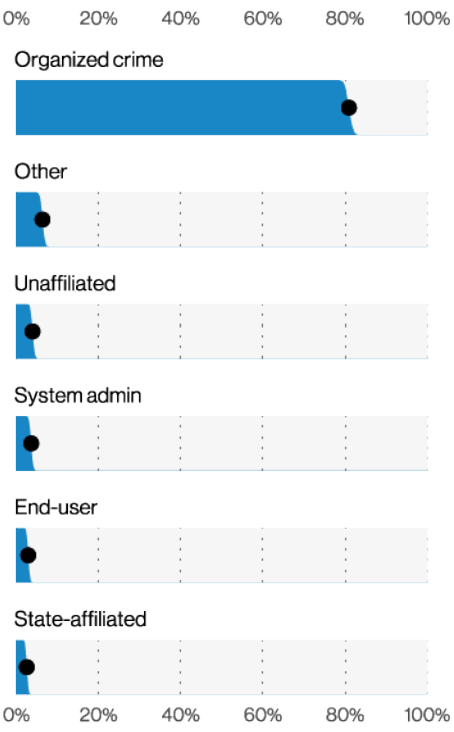



**Figure 16.** Top threat actor varieties in breaches (n=2,277)

14 As You Like It, William Shakespeare.
15 Anyone know if the Cyber+ trademark is available?

However, since 2015 it is relatively common for State-sponsored actors to also crave that cold hard cash[16] as the Financial motives for those actors have fluctuated between 6% and 16% of recorded breaches. Given this result, it should come as no surprise when you glance at Figure 17 and find that the two most common cybercrime terms found on criminal forums are bank account and credit card related.

Even as awareness of supply chain attacks has increased over the last few months, the overall percentage of incidents with a Secondary motive—where the ultimate goal of an incident was to leverage the victim's access, infrastructure or any other asset to conduct other incidents—has decreased slightly as a percentage from last year. There are two caveats here that should be kept in mind: The associated growth year-over-year of Financially motivated breaches, and that most Secondary motive breaches reported to us are simple in nature (which suggests the catastrophic ones on everyone's minds are still very much the exception).

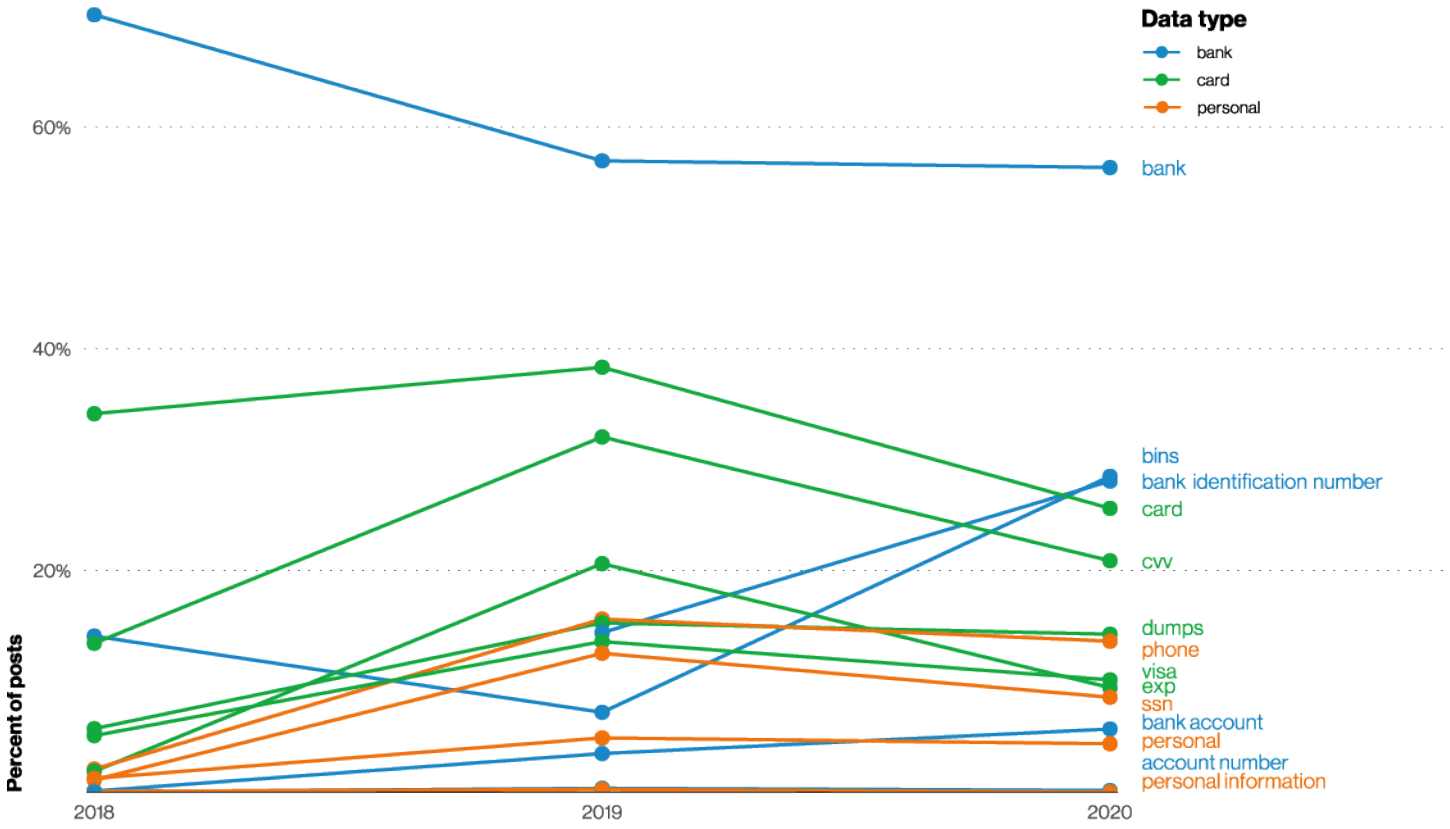



**Figure 17.** Terms over time in criminal forums and marketplaces

16 Or the hot ethereal cryptocurrency.

However, Secondary is still in second place (fittingly enough) as a top Actor motive, as Figure 18 demonstrates. So, if you are a software developer or service provider that has assets that could be repurposed in that manner, please make sure you are paying the proper attention to the operational parts of your organization.

In the same way automation may be helping you scale up your defensive operations, it can also help attackers scale up their offense. Figure 19 illustrates the relative occurrence of attack types in honeypot data. Near the top of the attacker's opportunistic sales funnel, we see scanners. Down near the bottom are where the Remote Code Execution (RCE) attacks reside. Regardless of their placement in the figure, automation is likely to assist attackers in moving potential victims from the top of the funnel to the bottom. As such, it's important to limit your public facing attack surface, through asset management, defensive boundaries and intelligent patching.

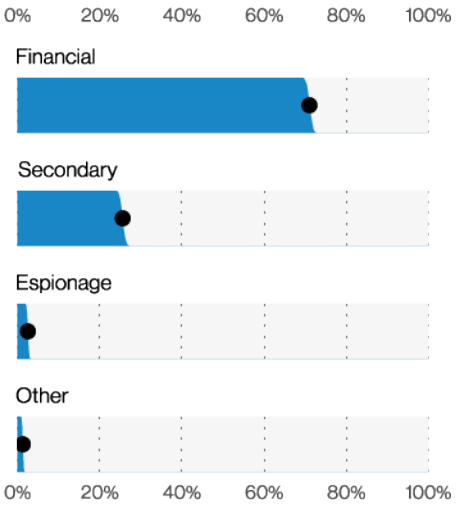



**Figure 18.** Top Actor motives in incidents (n=5,085)

**In the same way automation may be helping you scale up your defensive operations, it can also help attackers scale up their offense.**

# Secondary motive subset

In the Secondary Motive subset, we had an additional 24,913 incidents of which only one was a known breach. In all of these incidents, web apps were attacked with a secondary motive by External actors. Beyond that, we know very little.

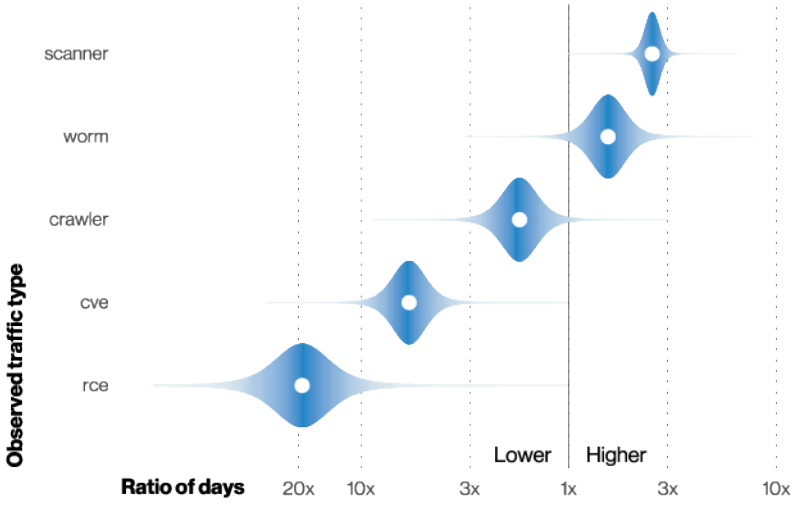



**Figure 19.** Ratio of days of high to low detection in honeypot data

# Action

Do we have an action-packed section for you, folks! Step right up, make room in the back so everyone can see! Figures 20 and 21 will reveal all you need to know about the frequency of Action varieties for the past year.

We do not want to divert all of your attention from the brand-new incident patterns. So we saved additional details on how those Actions manifested in the wild for you to dig your teeth into there.

## Talking the talk and acting the action

It would be impolite on our part not to address the virulent elephant[17] in the room, so we have centered this initial analysis of Actions on evaluating how adapting to life in a pandemic-stricken world has impacted the threat landscape. The DBIR team released a COVID-19 Threat Landscape Trends article[18] in the middle of last year, and it is only fair that we revisit how our speculations (see how we avoided the word predictions?) fared.

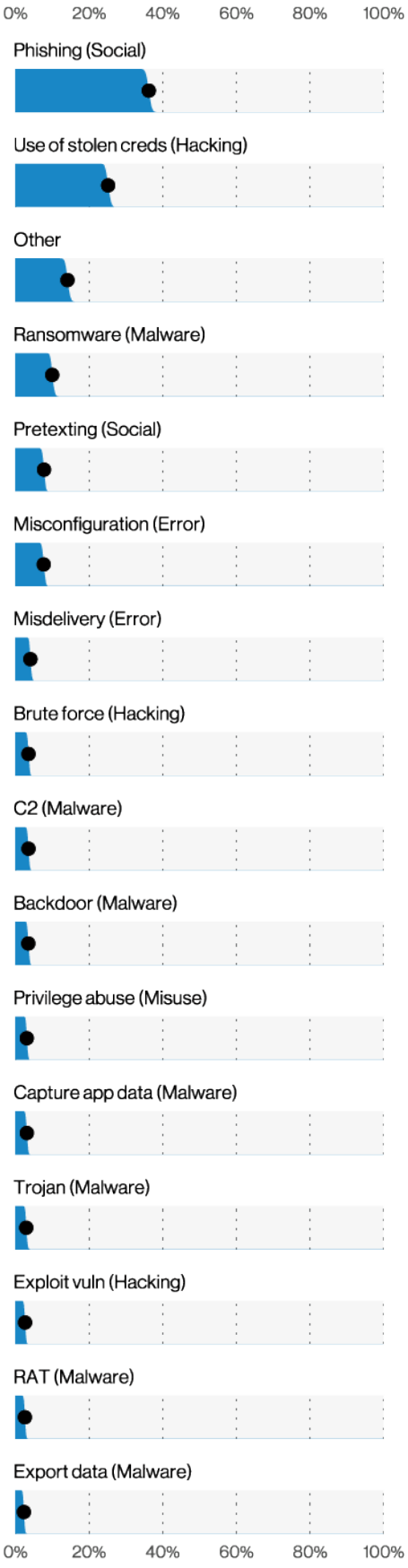



**Figure 20.** Top Action varieties in breaches (n=4,073)

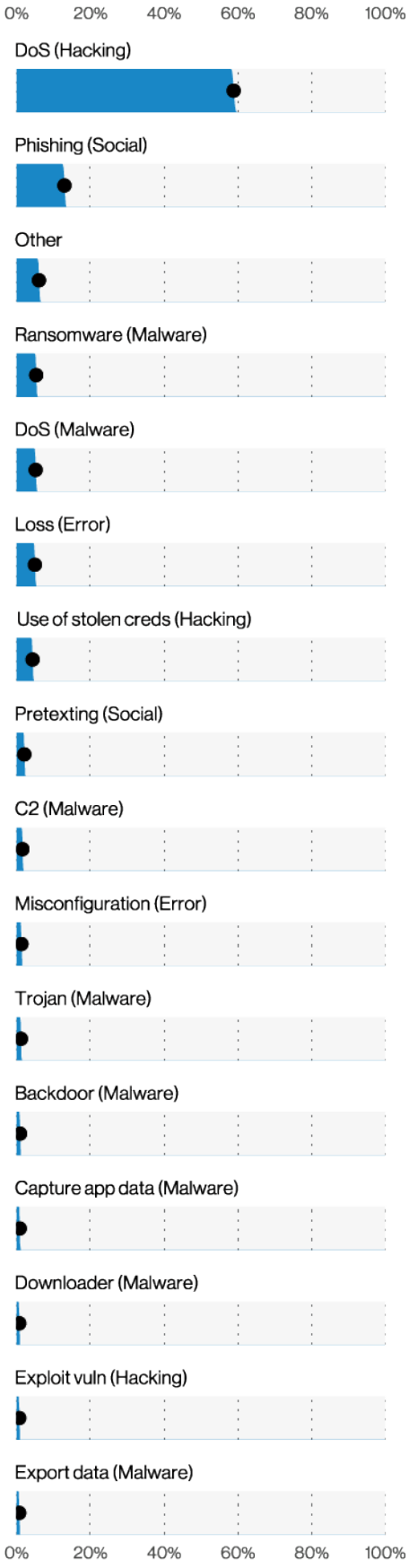



**Figure 21.** Top Action varieties in incidents (n= 24,362)

17 Viruphant? Eleplent?
18 https://enterprise.verizon.com/resources/articles/analyzing-covid-19-data-breach-landscape/

Figure 22 shows how the Actions we highlighted in that article varied in relation to last year's report. We highlighted Phishing, Use of stolen creds, Ransomware and Errors as Action varieties that could possibly increase.

Even in a year as unexpected as 2020, there are some things we can trust to stay the same. Phishing remains one of the top Action varieties in breaches and has done so for the past two years. Not content to rest on its scaly laurels, however, it has utilized quarantine to pump up its frequency to being present in 36% of breaches (up from 25% last year). This increase correlates with our expectations given the initial rush in phishing and COVID-19-related phishing lures as the worldwide stay-at-home orders went into effect.

Phishing continues to walk hand-in-hand with Use of stolen credentials in breaches as it has in the past. Admittedly, we expected to see an increase here due to a larger remote workforce. However, the numbers have remained in the region of 25% of breaches, which is still a significant number.

The major change this year with regard to action types was Ransomware coming out like a champ and grabbing third place in breaches (appearing in 10% of them, more than doubling its frequency from last year). This is also something we discussed, but this may have less to do with the changes in working arrangements than it does the shift in tactics of the actors who "named and shamed" their victims.

These actors will first exfiltrate the data they encrypt so that they can threaten to reveal it publicly if the victim does not pay the ransom. We are not sure if this breach double-dipping is permitted in the Threat Actor Code of Conduct, but there has been no evidence that they have one anyway.

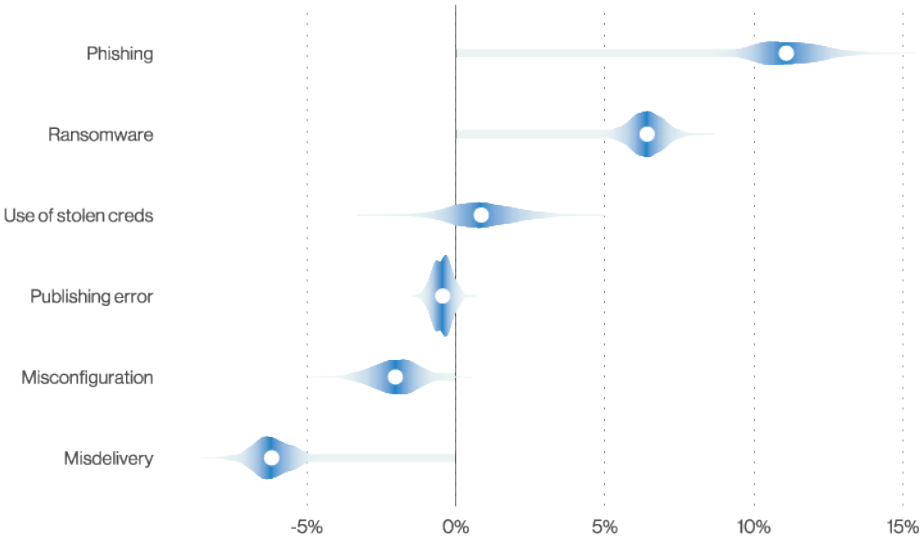



**Figure 22.** Change in COVID-19-related Action varieties

The final piece of this puzzle pertains to Error actions, where we opined that we would see an increase, but actually had a decrease this year to 17% of breaches (from 22%). This breaks a three-year streak of either staying the course or increasing. Granted, the absolute number of Error breaches did increase from 883 to 905. However, as a proportion of the dataset, Error decreased due to the rapid growth of Social breaches.

Of course, we here on the team secretly blame each other for this miscalculation on our part, as any team would. Still, both in relative and absolute terms, this is a significant value and is on par with Malware-related breaches as Figure 23 demonstrates, and it should certainly be front and center in your control definition strategy.

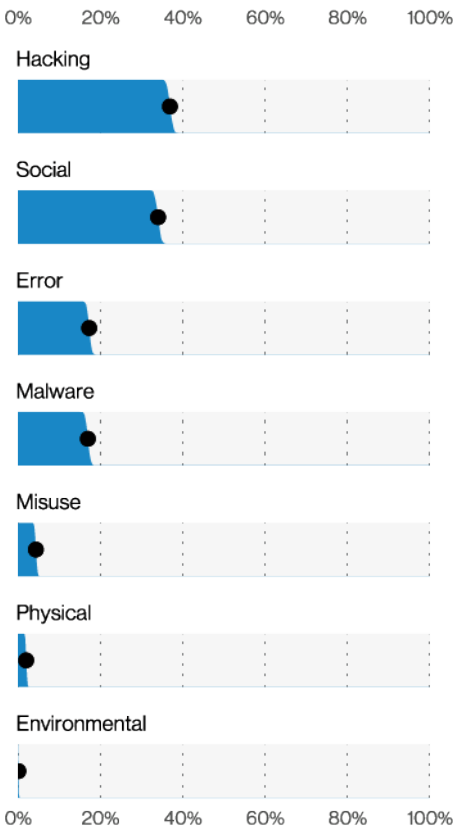



**Figure 23.** Actions in breaches (n=5,257)

# Actions have consequences[19]

A data point we started collecting over the past few years pertains to the results of Actions, which provide some interesting insights especially when you consider it as a complement to our ongoing attack chain research. For example, a threat actor might perform a Use of stolen credentials or Phishing action to Infiltrate a victim organization, but then deploy Malware in order to Exfiltrate the data they had their sights on.

The heatmap in Figure 24 shows how our most frequent results relate to our top-level Action categories.

Points of interest here are how well those findings align with the attack chain information that is present in some of the incidents we analyze. If an Action is concentrated into Infiltrate, it is closer to the top of the first actions in a chain chart, as shown in Figure 25, while Exfiltrate will correlate with the last one. Misuse actions are different, as they often assume or require legitimate access to the Asset that was breached, and, as such, are very focused into Exfiltration. With regard to Malware, well, given the Swiss Army Knife behavior of modern variants, it looks like you can eat your cake and have it too.[20]

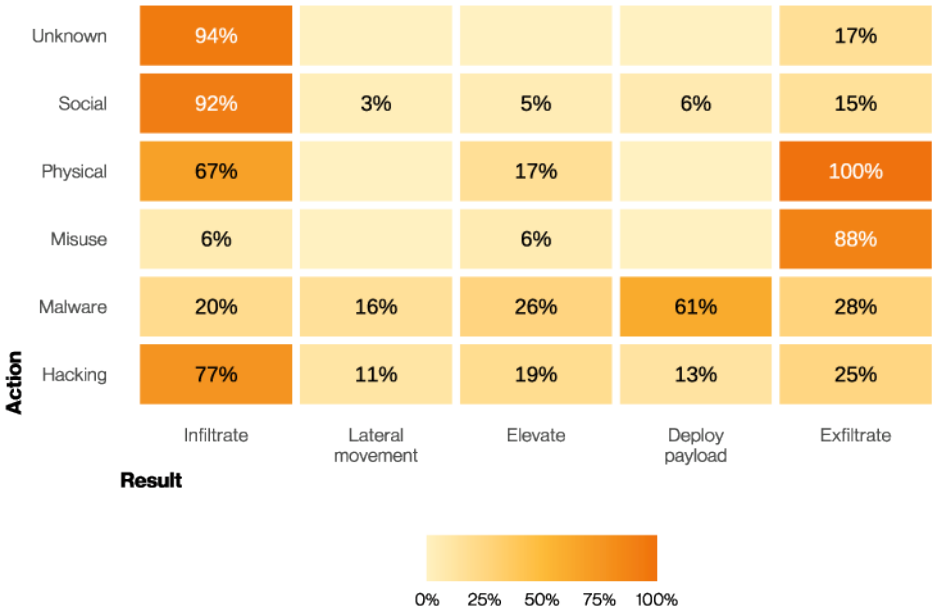



**Figure 24.** Results in breach Actions

19 Just like your Momma said.
20 Mmm...cake.

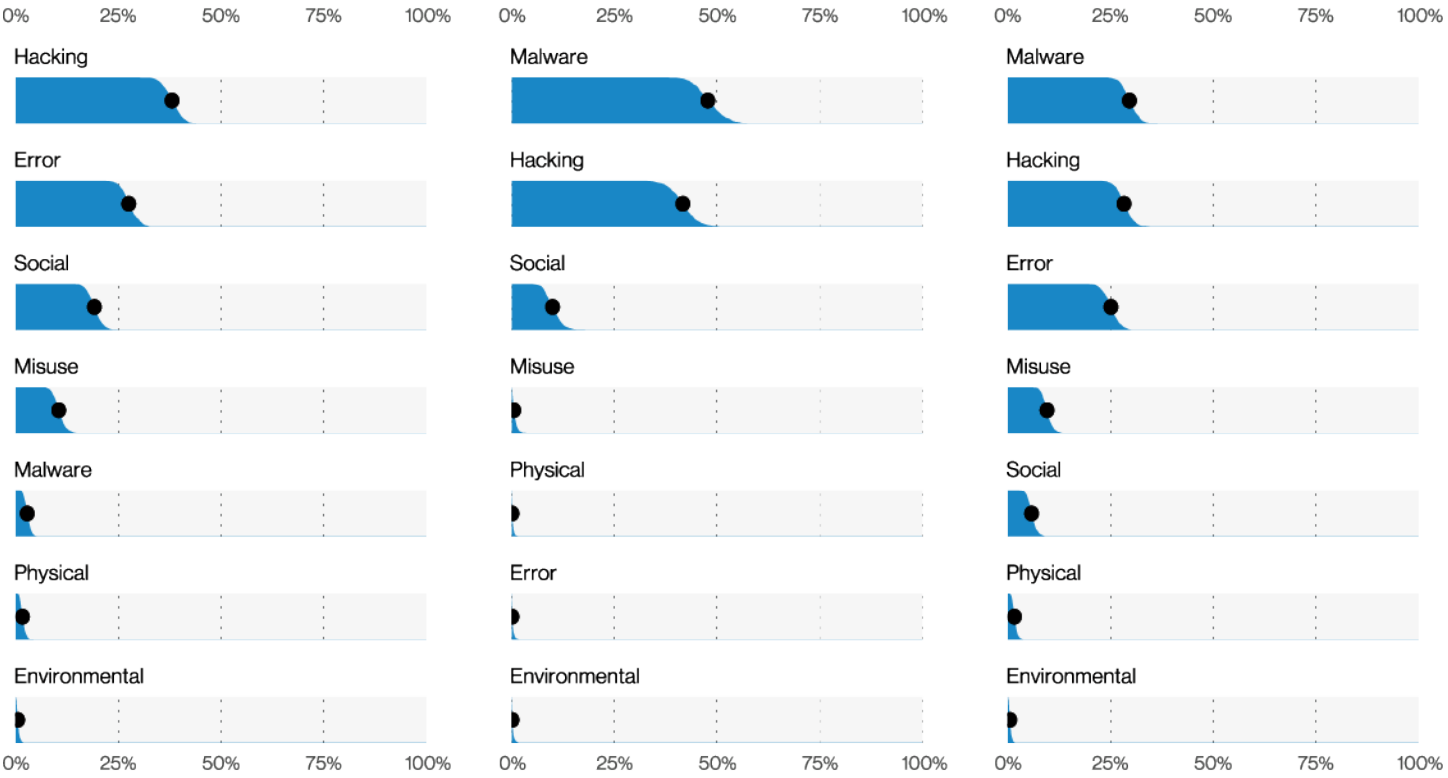



**Figure 25.** Actions at the beginning, middle and end of breaches

# Shared access is double access

Another noteworthy change this year is the increase in rank of Desktop sharing as the vector of a Hacking action to second place. As Figure 26 demonstrates, it is completely overshadowed by Web application as the attack vector. But it is now on the 5% threshold and we recommend attention to the authentication security of those. Notably, 89% of the Hacking varieties in this vector involved some sort of credential abuse (Use of stolen creds or Brute force).

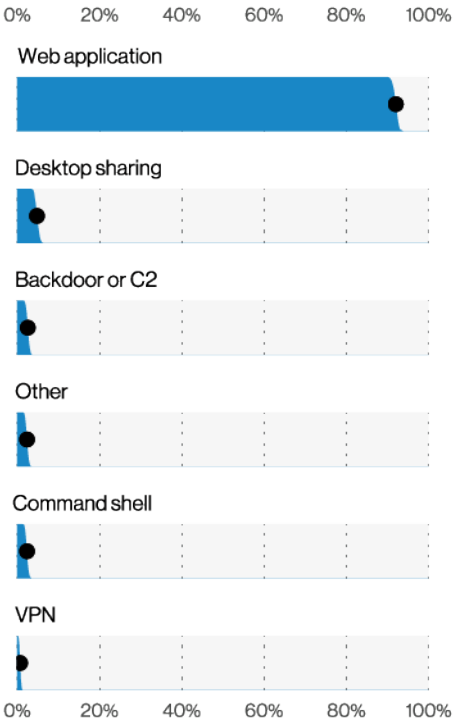



**Figure 26.** Top Hacking vectors in breaches (n=1,610)

# Assets

If, after looking at Figures 27 and 28, you had to double check that you weren't still in 2020, you would be forgiven. Servers are still dominating the Asset landscape due to the prevalence of web apps and mail services involved in incidents. And as social attacks continue to compromise people (they have now pulled past user devices), we begin to see the domination of phishing emails and websites delivering malware used for fraud or espionage.

However, we can glimpse the impact of a world where the flickering flames of digital transformation have slowly built into a sizable inferno when we review the Assets involved in breaches. Figure 29 shows that there is a large gap between

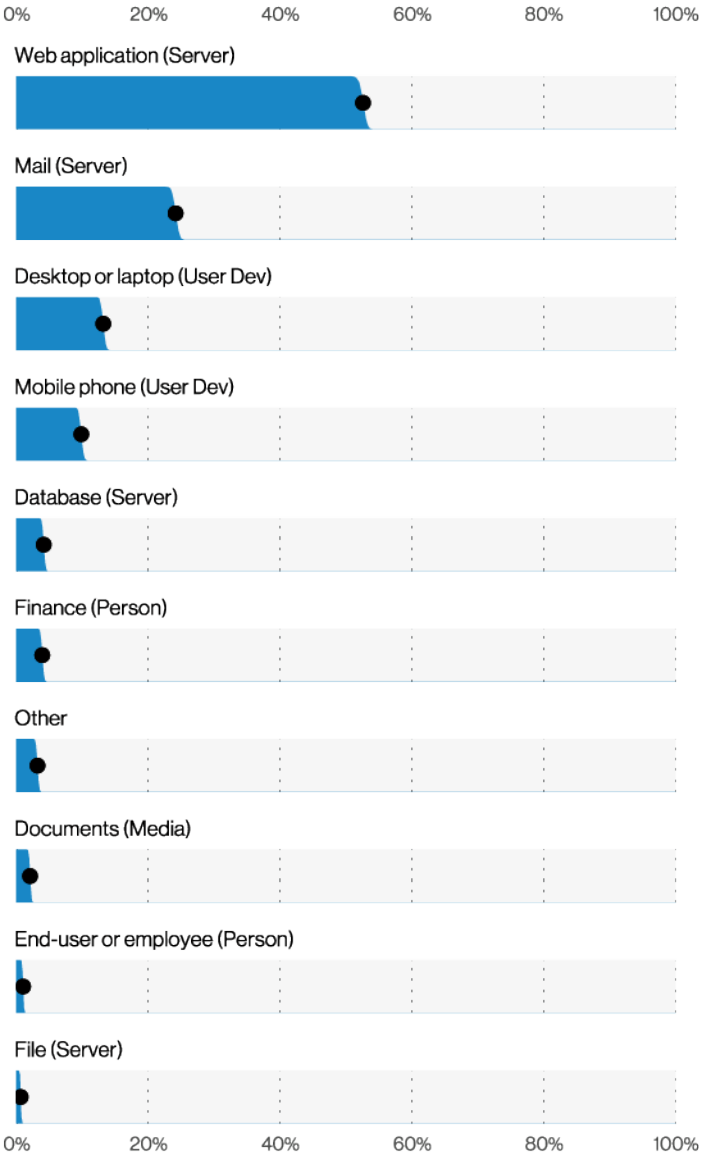



**Figure 28.** Top asset varieties in incidents (n= 9,188)

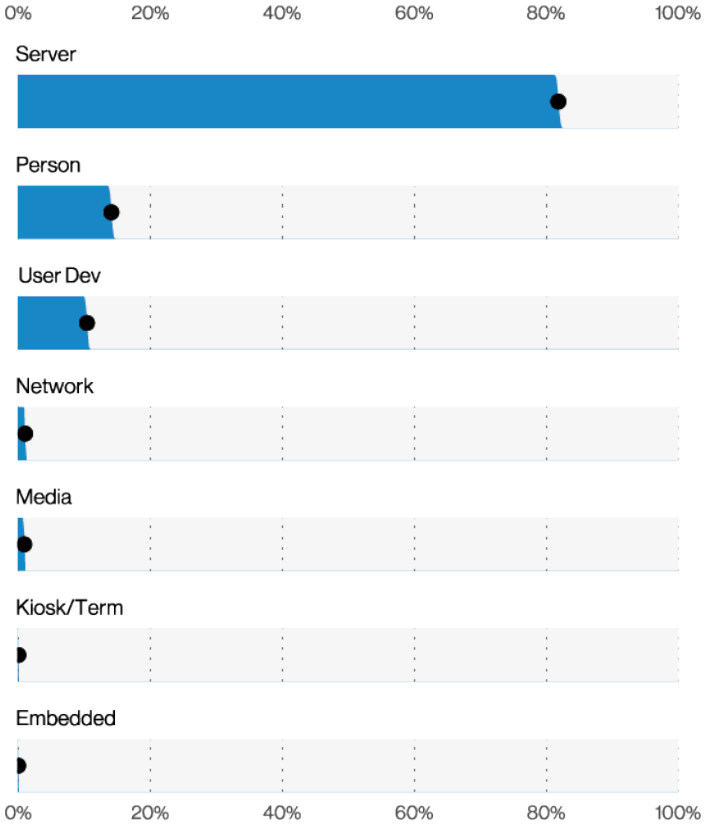



**Figure 27.** Assets in incidents (n=27,634)

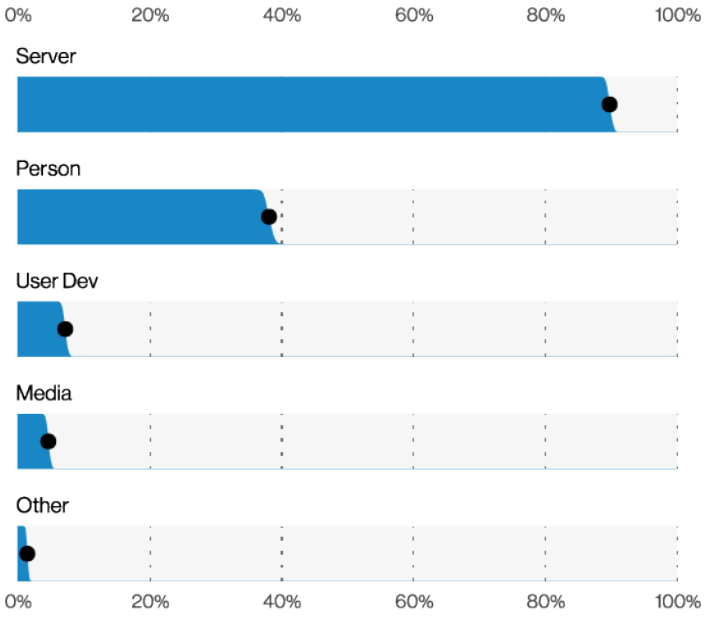



**Figure 29.** Top Assets in breaches (n=4,717)

Person and User devices as the most breached Assets, and the decline of User devices is statistically verifiable in relation to the previous two years. This result makes sense when we consider that breaches are moving toward Social and Webapp vectors, and those are becoming more server based, such as gathering credentials and using them against cloud-based email systems.

A related result that will likely not be surprising is that this year, external cloud assets were more common than on-premises assets in both incidents and breaches. Now before you put that in your marketing brochure for your next-gen AI[21] cloud security product, there were 10 times as many Unknowns (quite plainly incidents where the information on the location of the assets was not available) as there were cloud assets. That is more than enough to tip the scales in the other direction if we'd known more about what happened. Still, in a sample of random organizations, 17% that had a web presence had internet-facing cloud assets.[22] If it was not obvious by now, cloud assets deserve a seat at the grown-up security table and a piece of your budget pie.[23]

Even the median random organization with an internet presence has 17 internet-facing assets (Figure 30). Figure 31 gives you an idea of how vulnerable those organizations are. Most had no vulnerabilities at all. Furthermore, one might think that more recent vulnerabilities would be more common.[24] However, as we saw last year, it is actually the older vulnerabilities that are leading the way.[25]

**Rather than selecting out of something like the Alexa top 1 million domains, we randomly sampled a database of hundreds of millions of companies worldwide. Out of a million companies, only 1.4% had a web presence (a domain connected to the organization). It's easy to forget that the average security-conscious organization might be quite different from the average company.**

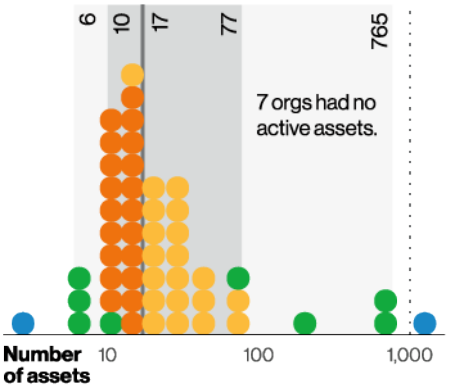



**Figure 30.** Number of internet-facing assets in randomly selected organizations (n=85) Each dot represents 2% of organizations

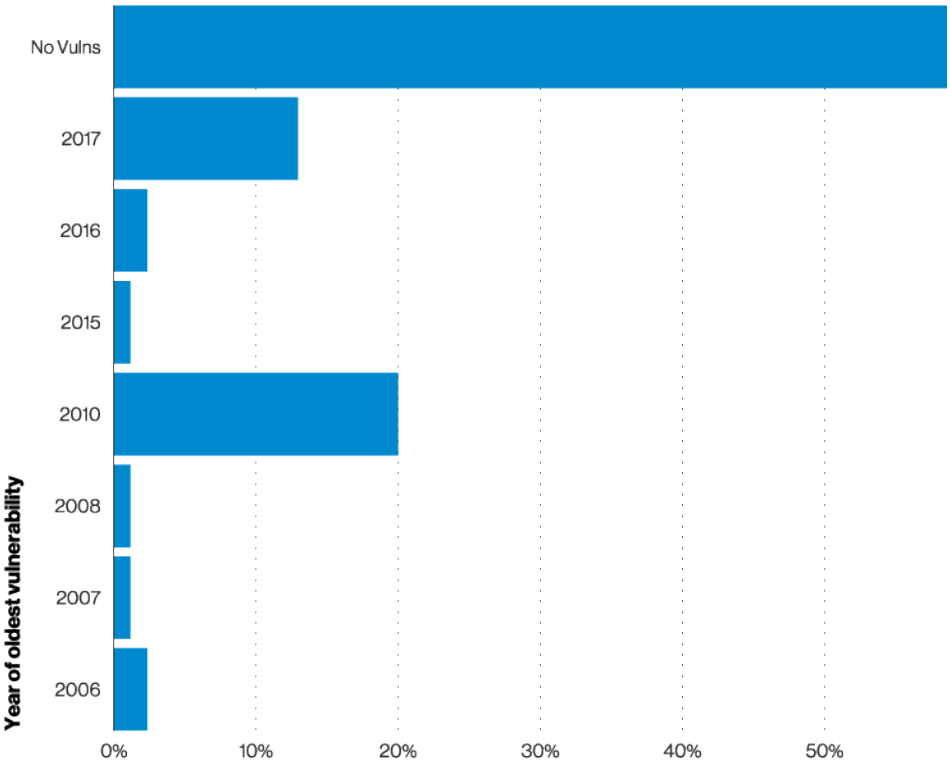



**Figure 31.** Organizations' oldest internet-facing vulnerability (n=85)

21 Emphasis on the "Artificial" not on "Intelligence."
22 See the sidebar for what we mean by "random organizations."
23 A terrible "pie in the sky" joke was edited out here. You are welcome!
24 You know, because of patching.
25 Just don't call them "boomer vulnerabilities," or you will start a fight. They might even tell you to get off their lawn.

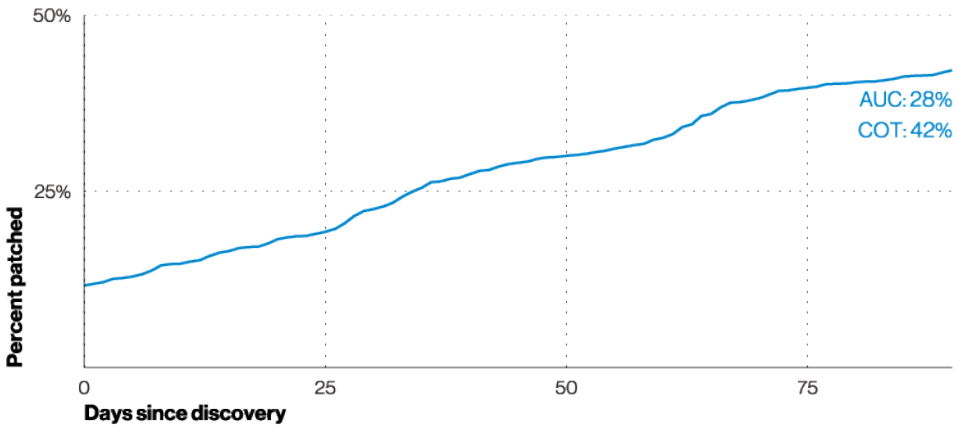



**Figure 33.** Patching in vulnerability scan data (n=110)

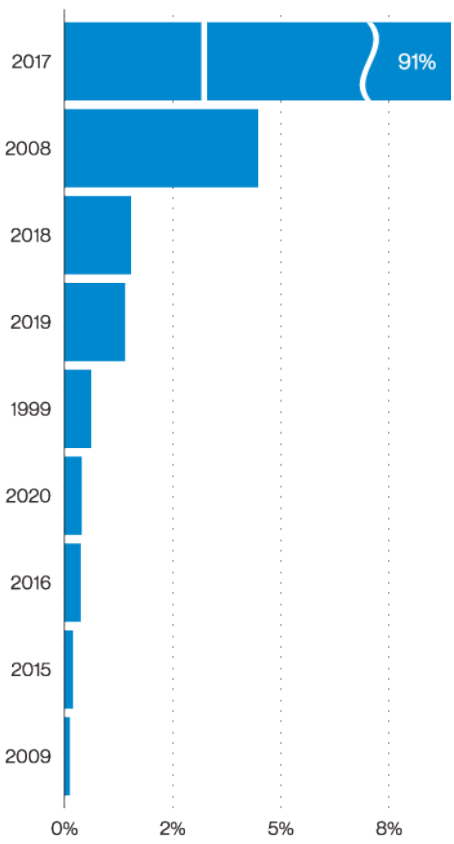



**Figure 32.** Percent of vulnerabilities by year in honeypot data (n=42,532,746) Eternal Blue is 37,217,565 of these. 2017 would be in 2nd place with 3% without it.

These older vulnerabilities are what the attackers continue to exploit. Figure 32 shows the discovery years of vulnerabilities that attackers attempted to exploit in bulk as seen from the perspective of honeypots. If Tom Brokaw were writing this report, he'd call them the greatest generation of vulnerabilities. Eternal Blue is a crowd favorite, which shows that the amount of time since discovery does not really factor into why actors target vulnerabilities. Instead, it seems to be simply a matter of what capabilities exploiting a vuln provides to the attacker, along with the robustness of current working exploits and payloads.[26]

So, what's a good, clean-cut, security-conscious organization to do? Based on Figure 33, the patching performance this year in organizations has not been stellar. Granted, it's never been great.[27] There are several likely hypotheses to explain why this year might be underperforming.

The ideal state for any organization is to patch smarter, not harder, by using vulnerability prioritization not necessarily to improve security, but to improve the organization's productivity. Every patch that has to be applied means you are that much farther from putting down the keyboard and picking up the d-pad.[28] Anything you can do to avoid patching vulnerabilities that do not improve your security keeps you just as secure but involves much less work (and less chance of burnout from your employees or service providers).

Mobile phones made the list in Figure 28 at the beginning of this section. As with last year, this finding is somewhat anticlimactic, as the vast majority are simply lost phones. Still, that's not quite the end of our mobile foray. We also have mobile data on malicious URLs and APKs[29] in Figure 34. What we found, in short, was that you don't have to be a large organization to have a good chance that one of your members has received a malicious URL or even installed a malicious APK.[30]

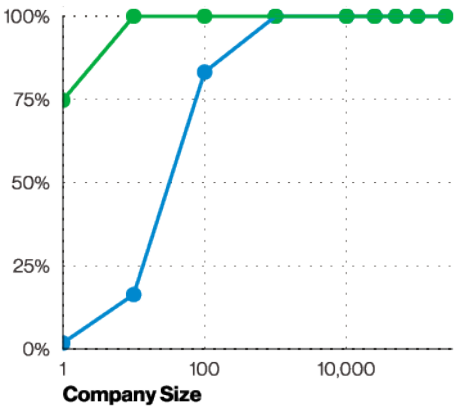



**Figure 34.** Probability that someone in the company will receive a malicious URL or install a malicious APK based on organization size (n=5,440,000)

26 As we write this section, a Microsoft Exchange Remote Code Execution Vulnerability (CVE-2021-26855) is being actively and massively exploited that has all the ingredients to also be part of this growing background noise of exploitation activity in the internet.

27 2017 DBIR, Figure 56.

28 Or your kid, or your running shoes, or something else that keeps you sane.

29 Android apps.

30 Observant readers may have noticed the assets section missing anything about Information Technology (IT) vs. Operational Technology (OT) assets. That's because it was largely missing from our dataset as well. We've heard those OT breaches are somewhere, but they're not in our dataset.

# Attribute

The Attributes are the Confidentiality, Integrity and Availability (aka the CIA[31] Triad) violations of the impacted asset. Whether it is a confirmed data breach in which the confidentiality of the data was compromised, or an integrity incident, such as altering the behavior of a person via phishing, the actions against the assets result in CIA violations. First, let's discuss Confidentiality and the types of data that are most frequently compromised.

As we have pointed out in previous reports, Credentials remain one of the most sought-after data types (Figure 35). Personal data is a close second. Considering that Personal data includes items such as Social Security numbers, insurance-related information, names, addresses and other readily monetizable data, it is little wonder that attackers favor them as they do. They are also useful for financial fraud further down the line, not to mention their resale value.

We do not mean to imply that attackers are the only way data is compromised. Sadly, we cannot discount the ability of our own employees to make mistakes, thereby contributing to the problem. However, they are less likely to involve credentials, and more likely to involve other data such as Personal information (Figure 36).

**As we have pointed out in previous reports, Credentials remain one of the most sought-after data types.**

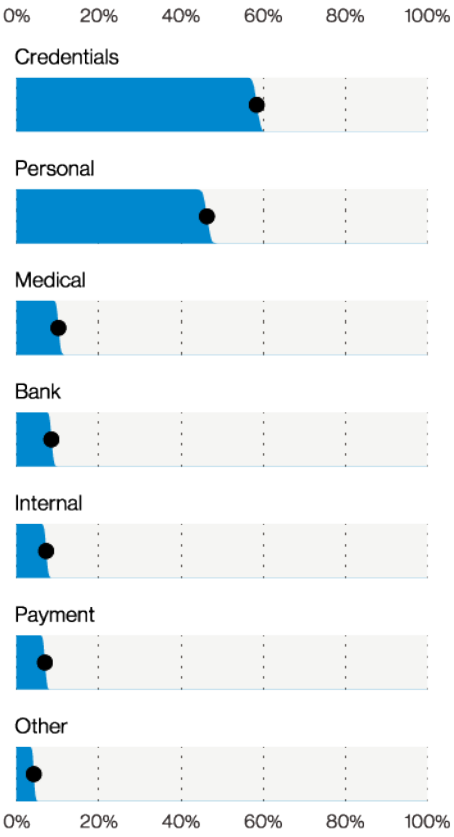



**Figure 35.** Top data varieties in breaches (n=4,552)

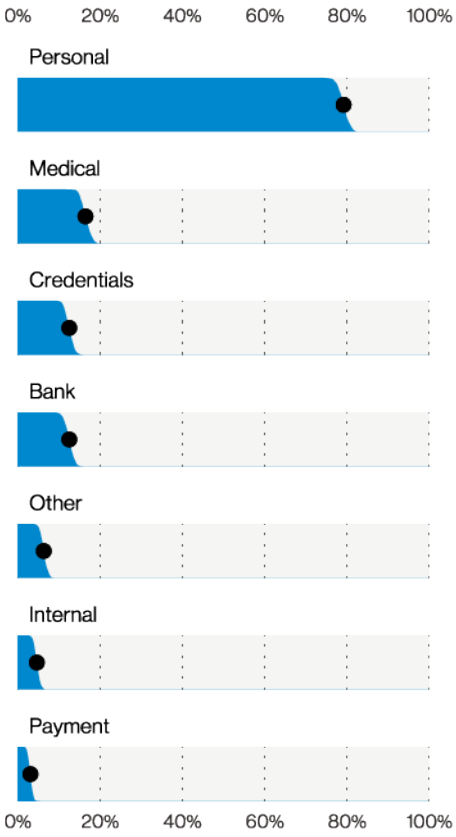



**Figure 36.** Top data varieties in Error breaches (n=839)

31 Not the CIA that keeps the alien presence on the DL, the other kind.

Moving on to Integrity violations (Figure 37), these are usually the result of a Social or Malware action. For the Social actions, Phishing and Pretexting will alter the behavior of their targeted victim. In some cases, Pretexting results in the initiation of a Fraudulent transaction, causing money to go where it was not supposed to. With the prevalence of Phishing and Pretexting in our dataset this year (43% of breaches) it is no surprise that Alter behavior ranks first among the Integrity violations.

But we must not forget the Malware actions. Software installation comes in second place due to the high number of System Intrusion pattern cases that had a Malware component. Most commonly these were directly installed by the actor after system access—usually after a Hacking action such as the Use of stolen creds or Brute force.

Finally, we arrive at our Availability violations (Figure 38). The most common is Obscuration, which is what you get when ransomware is installed and the encryption is triggered. Loss is our second most common violation, and results from either a lost or stolen asset, as you no longer have access to that data.

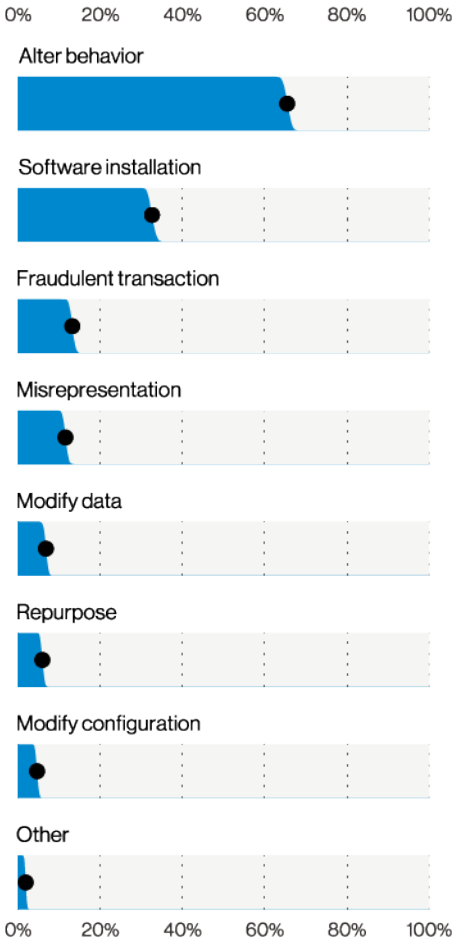



**Figure 37.** Integrity varieties in breaches (n=2,762)

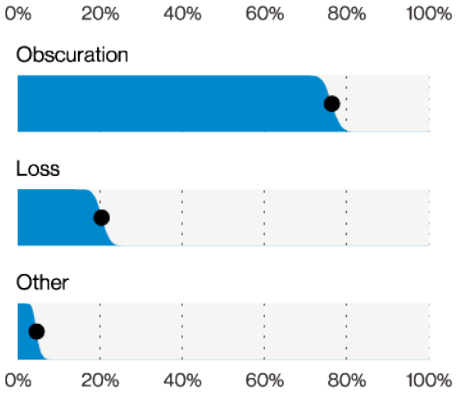



**Figure 38.** Top Availability varieties in breaches (n=541)

# Timeline

This year we decided to take a look at which breach types take the longest to discover (Figure 39). Traditionally, this has been insider Privilege Misuse. However, when looking at this year's data (largely due to the insight provided by the new patterns), we found that the differences between Privilege Misuse and System intrusion were negligible. Both were present in the longest to discover breaches.

In contrast, the breaches that are the fastest to discover appear to be those where it becomes readily apparent something is wrong. Examples include Stolen assets, because the employee found evidence of a break-in, and Errors, where the employee had that sinking feeling that they screwed up, and reported it in the hopes that it could be quickly contained. These are both internal methods of discovery, and if you don't already have an easy and fast way for your people to report these kinds of breaches, you should look into it. Why not cultivate your employees to be your early warning system when it can have a great return on investment?

The other end of the spectrum for discovery methods is when the threat actor involved makes the "notification" in the form of a ransom note that appears on screen.

Finally, we were also curious what kind of data was the fastest to be compromised, and that turns out to be Credentials. This is particularly the case in Phishing, which typically goes after the victim's credentials for use in gaining further access to their chosen victim organization.

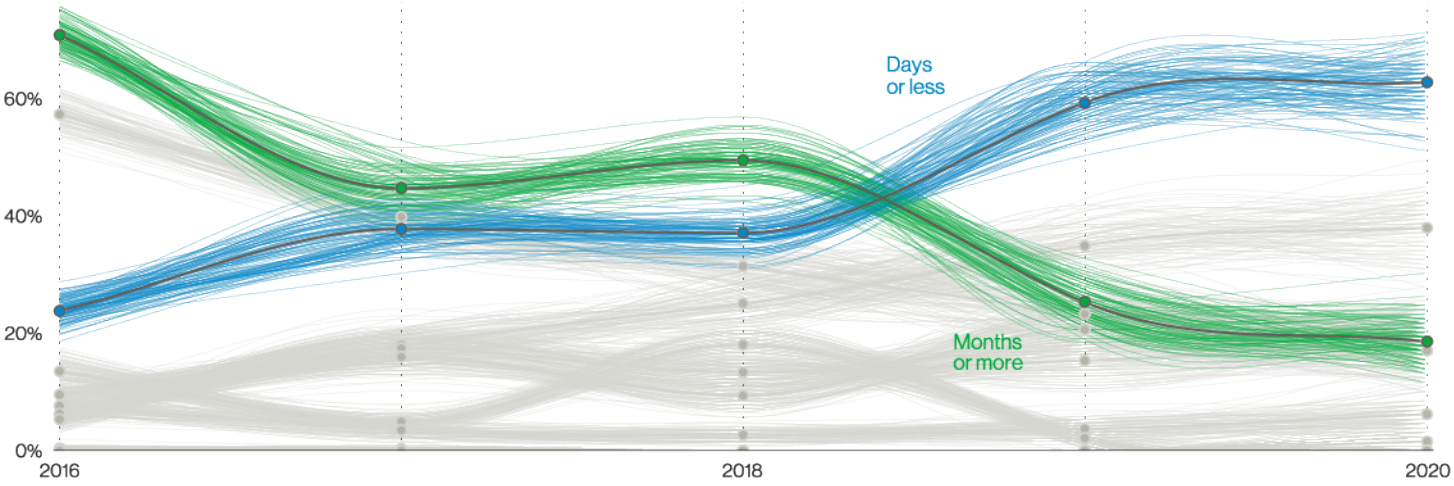



**Figure 39.** Discovery over time in breaches

# Impact

## Many hands make for light work

Attackers continue to profit substantially from the adversity that befalls breach and incident victims. And while that profit is certainly of interest,[32] what really concerns us is how the amounts tally up on the other side of the transaction. Figure 40 illustrates the range of loss from various types of incidents based on adjusted losses reported to the FBI Internet Criminal Complaint Center (IC3).[33] In this figure, each dot represents half a percent of incidents. First and foremost, according to IC3 data, is the fact that whether the attack was a Business Email Compromise (BEC), Computer Data Breach (CDB) or a ransomware attack, a large percentage of incidents did not actually result in a financial loss (42%, 76%, and 90% respectively).

When losses did occur, they were not of the one-size-fits-all variety. Following the rules of good business, we expect attackers to charge what the market can bear. For a small organization that is usually a small amount. For a large organization, however, losses can be much more substantial. When examining breaches that included a reported loss, 95% of BECs fell between $250 and $985,000 dollars with $30,000 being the median. That is a pretty big range, you say? Maybe so, but CDB ranges were even wider with 95% falling between $148 and $1.6 million, and a median loss of $30,000. Finally, for ransomware the median amount lost was $11,150, and the range of losses in 95% of the cases fell between $70 and $1.2 million.

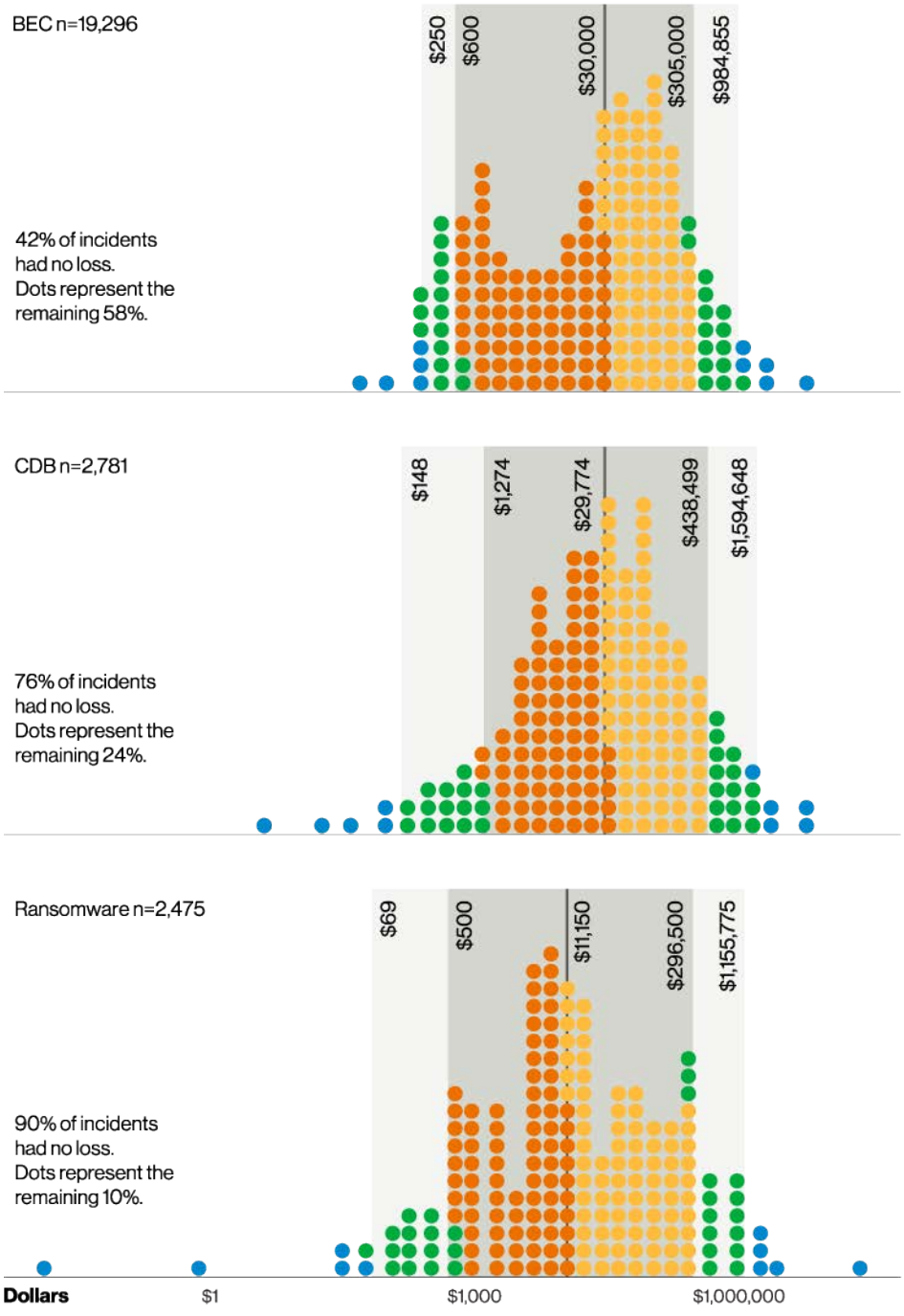



**Figure 40.** Loss by incident type
Each dot represents 0.5% of incidents

32 It would be fascinating to analyze profitability of different types of attacks from the perspective of the threat actors, but not only do we not believe we have the data necessary; we are not sure if this analysis would benefit the threat actors more than the defenders.

33 https://www.ic3.gov

Let us state this in a somewhat different manner: If you only consider the bottom half (everything below the medians that we just mentioned), CDBs are often associated with bigger losses than are ransomware events. This finding, when coupled with the 90% of ransomware incidents that did not result in any loss, could be telling the story that organizations are no longer paying the ransoms. It must also be kept in mind that this loss data includes individuals as well as organizations, which is another potential reason for the numbers being smaller. Unfortunately, we do not have a sufficient level of detail to distinguish between the two. There is also the specter of potential bias toward underreporting of larger ransoms. If, however, organizations are skipping the ransom, the low payout ranges could have been yet another contributing factor for the rise of the ransomware "name and shame" threat actors witnessed in late 2019.

In a "glass half full" view of the above situation, there is some possible good news in that there is a chance you can reverse the mass migration of your funds to other environs. The IC3 Recovery Asset Team (RAT) can sometimes assist victims in the freezing of lost funds for possible recovery. In Figure 41, we see that when the IC3 RAT acts on BECs, and works with the destination bank, half of all US-based business email compromises had 99% of the money either recovered or frozen, whereas only 11% had nothing at all recovered. If your organization experiences an incident, we highly recommend that you contact the local branch of your national law enforcement and seek their assistance. Or, better yet, get to know them before the breach occurs!

Of course, direct losses are not the sole cost one encounters due to a breach. Apart from the damage done by the attacker, there remains the expense of Digital Forensics and Incident Response (DFIR) and legal counsel. Figure 42 provides an idea of what to expect in these areas based on cyber insurance[34] claims. Each dot represents 2% of incidents. As you can see, 50% of incidents had no associated forensics costs. When forensics costs were present, 95% fell into the range of $2,400 to $336,500. Slightly fewer incidents had no associated legal costs, (36%). For the remaining 64%, 95% of the legal costs fell between $800 and $54,000.

It should be pointed out that insurance data can be somewhat biased. For instance, insurance may not cover legal costs or penalties. There may also be an additional deductible not covered in the overall costs. Of course, to address the elephant in the room,[35] it is unlikely that your insurance will cover the damage to your company's reputation. And depending on several factors such as disclosure requirements, the size of the breach, and other things hiding in the fine print, that damage can be considerable.

Various studies have arrived at very different conclusions regarding the impact on stock price from a breach in the days immediately following a breach, including 2.53% (Rosati, Cummins, Gogolin, van der Werff, & Lynn, 2017), 5% (Cambell, Gordon, Loeb, & Zhou, 2003), 2.1% (Cavusoglu, Mishra, & Raghunathan, 2004), and 1% (Goel & Shawky, 2009). The findings of these studies are helpful, but they don't shed much light on what happens in the long term. Figure 43 may help to illuminate the matter somewhat.

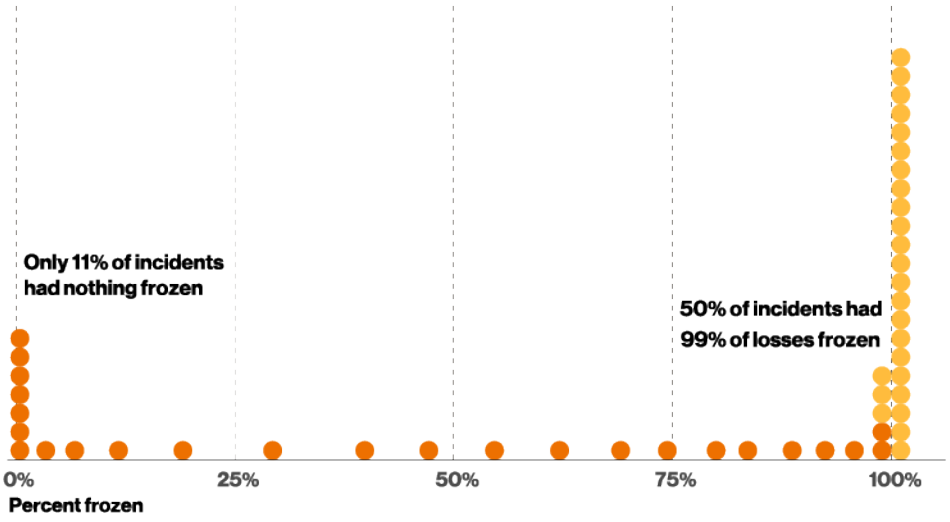



**Figure 41.** Percent of losses frozen for recovery (n=1,086)
Each dot represents 2% of incidents

34 For an additional fee, Verizon will provide a version of the DBIR that replaces all instances of "Cyber" with "Security." See your local Verizon representative for details.
35 Another elephant? This is a pachyderm-filled space!

Based on data collected by comparitech.com,[36] breached companies underperformed the NASDAQ (a U.S. Stock Market) by about 5% after six months, though if you look at 95% of companies the performance was anywhere from 48% under to 39% over performing. If we look two years into the future of those organizations (after the breach), those downward trends continued, suggesting that perhaps the breach wasn't actually the cause, but the symptom.[37]

To answer the question, "what might a breach cost in total?" we ran 1,000 Monte Carlo simulations using bootstrap sampling on breaches we had cost information about on this year's dataset like the good data nerds we are. Fourteen percent of the simulated breaches had no impact. Of the 86% that were impacted, Table 1 captures the results. What you do with these numbers is, of course, up to you. While you could plan for the median breach of $21,659, a better option might be to plan for the middle 80% of breach impacts, $2,038 to $194,035. Or better yet, be prepared for the most common 95% of impacts, between $826 and $653,587. If you add to that an organizational devaluation of around 5% (from Figure 43), then you just may have yourself a tangible figure you can plan around.

| Percent of breaches | Lower | Upper |
|---|---|---|
| Median | $21,659 | |
| 80% | $2,038 | $194,035 |
| 95% | $826 | $653,587 |

**Table 1.** Simulated breach costs

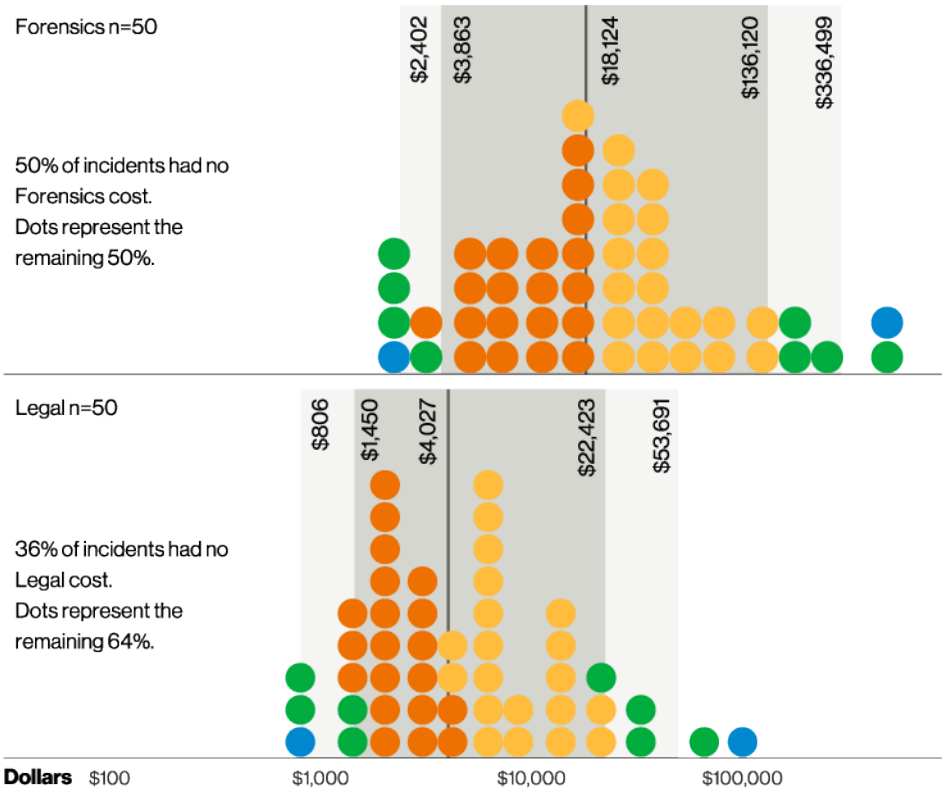



**Figure 42.** Cost by incident type
Each dot represents 2% of incidents

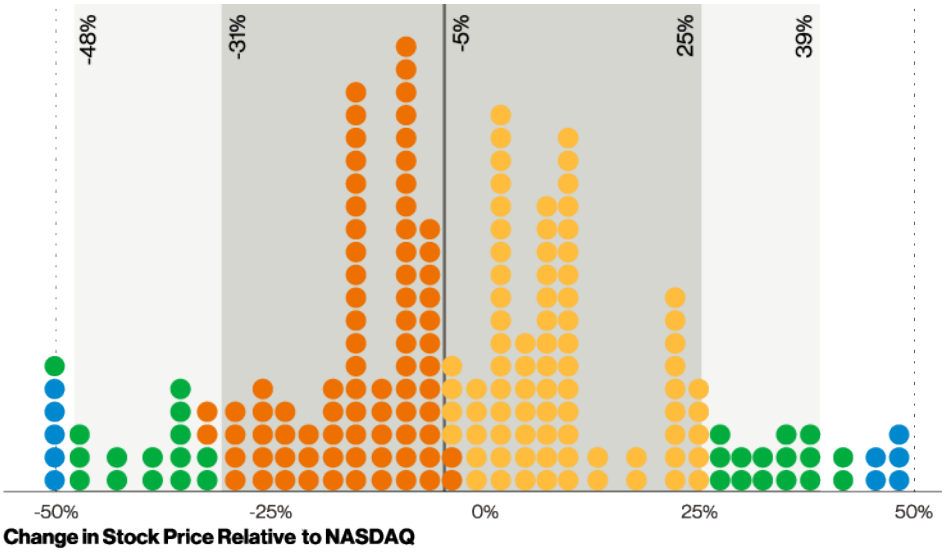



**Figure 43.** Changes in stock price for companies with breaches after six months (n=39)
Each dot represents 0.5% of incidents

36 More precisely, Paul Bischoff's (@pabischoff) blog post at https://www.comparitech.com/blog/information-security/data-breach-share-price-analysis/
37 Dr. Frank N. Furter nods approvingly.

# About the FBI

**Herbert Stapleton**
**Deputy Assistant Director, FBI Cyber**

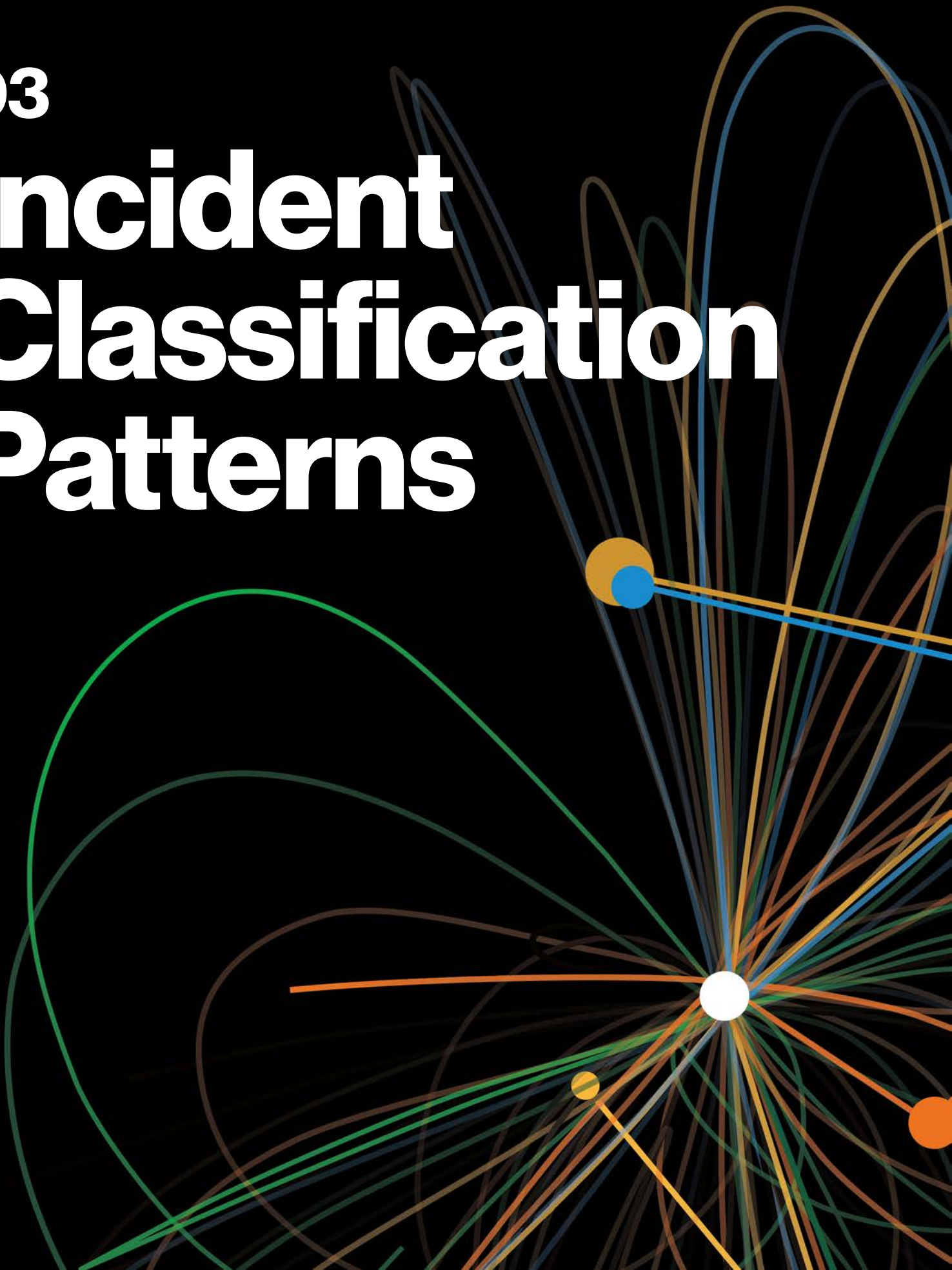
Over the past decade, the cyber threat has grown exponentially with nation state and cyber criminals increasing the scale, scope and level of sophistication of their cyber attacks. Addressing this kind of complex and agile environment requires a more comprehensive response than any one single government agency, business, technology, or data source can provide. Instead, an interwoven architecture of combined capabilities from across public agencies and the private sector must be leveraged to protect critical infrastructure and impose risk and consequences on attackers.

The FBI is committed to sharing as much as possible about cyber threats as quickly as possible so the public is alerted and prepared. We strive to be viewed as an indispensable partner, using our unique authorities as a law enforcement agency and member of the United States Intelligence Community to enable government operations against our cyber adversaries and allow the public to enhance their security posture. Because of our unique authorities, world-class capabilities, enduring partnerships, and presence we can conduct investigations, collect intelligence, and interact with victims – all in pursuit of attribution. Attribution is what allows the U.S. government to impose risk and consequences on our adversaries and prioritize our operations with our partners, including the private sector. Cyber [Combatting cyber crime] is the ultimate team sport and we all must be committed to using every tool we have at our disposal to address the cyber threat.

Of utmost importance to the FBI, and a key component of our foundational cyber strategy, is the ability to share relevant and actionable information with our government partners, the international community, private industry, and the public. But, we also rely on the information received from our partners, private industry, and victims to develop a broader picture of cyber threats. The Internet Crime Complaint Center (IC3) serves as a reliable, convenient, tool for submitting information to the FBI about suspected internet-facilitated criminal activity, while also developing effective partnerships with law enforcement and private sector entities. Information provided to the IC3 is further analyzed, resulting in investigative leads or the identification of new or emerging cyber threats. We share what we've learned through our analysis of IC3 data with the public and private industry through PSAs, alerts and reports such as the DBIR.

For the 2021 DBIR, the FBI's IC3 focused on supplying data specifically for business email compromises/ email compromises (BEC/EAC), and other data breach incidents reported to IC3. In recent years, the FBI's IC3 has observed that BEC/EAC and data breach incidents trend more towards victimizing corporations and/or private sector entities and less on targeting a single individual. IC3 recognizes that the public plays a central role in IC3 being able to understand how cyber criminals are evolving. By submitting a cyber related complaint, the public is assisting the FBI in addressing those specific complaints, as well as, identifying the critical details of developing cyber threat trends.

# 03
# Incident Classification Patterns



# Incident Classification Patterns: Introduction

## The times they are a-changin'

Remember 2014? Uptown was funky, Pharrell Williams was happy, and if you had a problem, you could shake it off. The DBIR first introduced the Incident Classification Patterns in 2014, as a useful shorthand for the sometimes complex combinations of VERIS Actors, Actions, Assets and Attributes that frequently occur. The threat landscape has changed a bit since then, and we are now happy to introduce a refresh of the DBIR patterns.

As you can imagine, this was a very hard decision for the team, but we were able to find strength and courage from the leadership shown by big, bold, refreshing business moves such as the release of New Coke and Crystal Pepsi.

Our new patterns explain 99.3% of analyzed breaches and 99.6% of analyzed incidents this year. They also explain 95.8% of quality breaches and 99.7% of quality incidents over all time.[38]

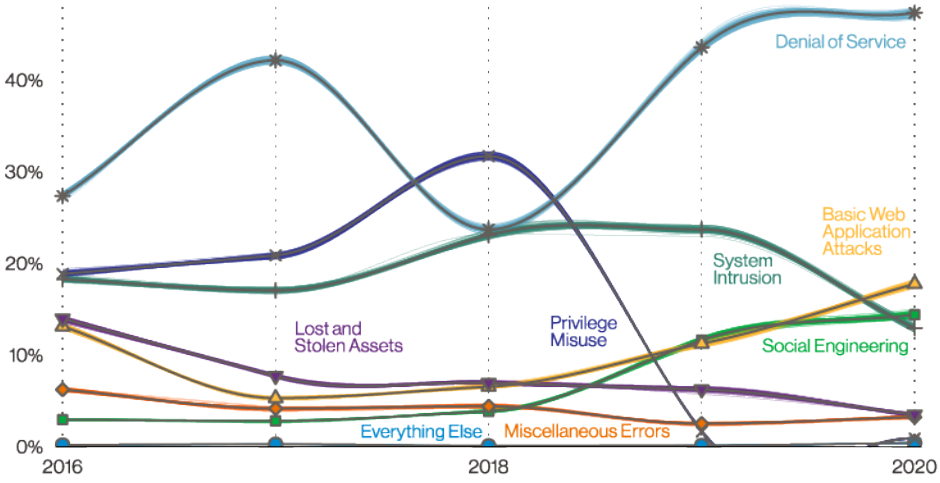



**Figure 44.** Patterns over time in incidents

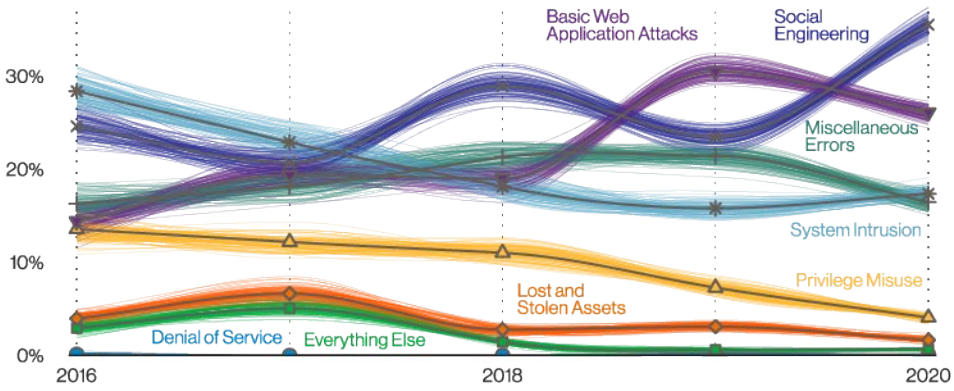



**Figure 45.** Patterns over time in breaches

38 Last but not least, it kills 99.9% of germs on contact! Ok not really.

Of course, not everything has changed. Denial of Service, Basic Web Application Attacks,[39] Lost and Stolen Assets, Miscellaneous Errors, Privilege Misuse, and Everything Else renew their contract for another season. Payment Card Skimmers, Crimeware, Cyber-Espionage, and Point of Sale are the MVPs[40] retired to make room for a couple of seasoned minor league patterns, ready for the big leagues: Social Engineering and System Intrusion.

Now just because some names haven't changed it does not mean the patterns are the same. What has been currently assigned to the 2021 version of Miscellaneous Errors (for example) is not necessarily what was in the 2014 Miscellaneous Errors.

The original patterns were based on a hierarchical clustering approach that helped derive some simple rules used to assign incidents to patterns. It was a very prescriptive process that worked quite well at the time, but we could see the strain starting to show.[41]

The new patterns are based on an elegant machine-learning clustering[42] process. Making this decision was a gamble in many ways, as we were committed to trust the data on this process, and it paid off. The new patterns clearly fell around the same ones that had been prescriptive before, but also better capture complex interaction rules the old ones were unlikely to handle.

Figures 46 and 47 give an idea of where incidents and breaches went between the old and new patterns. First, the easy-to-explain changes. Lost and Stolen Assets are still mostly in the Lost and Stolen Assets pattern. The same can be said for Miscellaneous Errors, Privilege Misuse, Basic Web Applications Attacks, and Denial of Service. What has changed starts with Payment Card Skimmers, which now falls squarely into Everything Else. It originally had some similarities to the current System Intrusion pattern in that that is where non-webapp payment card breaches ended up. Obviously, skimming isn't really what we think of when we picture a popped system, so over to Everything Else it goes.


**Figure 46.** Old patterns mapped to new patterns in incidents

39 Which, after going through an incredibly scientific, focus-tested rebranding, briefly became "The pattern formerly known as Web Applications," but then Verizon Branding and Communications said we couldn't do that either. We were bummed—we had a symbol picked out and everything.

40 Most Valuable Patterns

41 Like a three-day holiday visit with your in-laws.

42 We will be talking about it in way more detail than necessary in the next part of the section.

Of more interest are Point of Sale, Crimeware, Cyber Espionage and Everything Else. They are now defined by the characteristics of the breach. Was Social Engineering the significant aspect? To the new Social Engineering pattern it goes! Was it a simple attack where the initial intrusion point was the web application? To Basic Web Application Attacks it goes! Or was it more of an elaborate system intrusion where the attacker gained access and poked around, maybe without us even knowing how they gained access? System Intrusion is just waiting to welcome those incidents with open arms like an old Journey song. Those big changes weren't exactly planned (quite frankly nothing in the DBIR ever is in regard to what the data is going to tell us).

Thanks to the re-focused patterns, we can provide better guidance when one of those patterns appear at the top of your industry. Cyber Espionage and Crimeware could suggest a different complexity of the incidents in most cases, but your controls don't care if the threat actor has a cushy government job or if they are a free-market enthusiast entrepreneur.

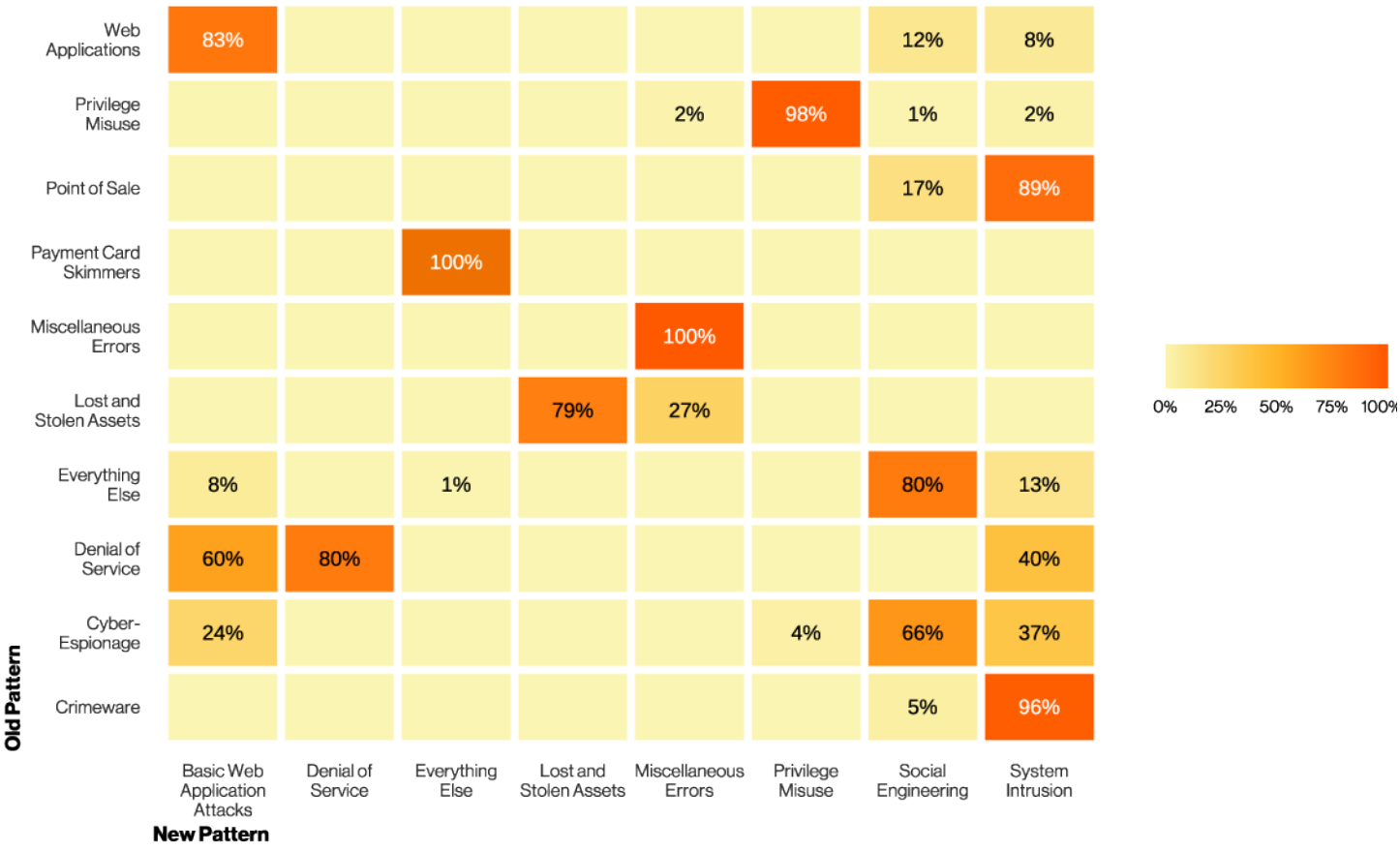



**Figure 47.** Old patterns mapped to new patterns in breaches

# This is the way

Coming up with new patterns was not a superficial process. It has been in the works for some time. Clustering DBIR data is not quite as straightforward as it might seem. First, we have almost 2,600 columns in the dataset leading to almost assured overfitting. Second, our data is mostly logical rather than categorical or continuous, limiting the approaches that are likely to work. Third, we have over 800,000 rows in our dataset, again limiting the approaches that would work. Fourth,[43] we are well aware that our clusters would be imbalanced. There would be some clusters with far fewer incidents and breaches than others. Fifth, the results needed to be somewhat explainable, always a fun proposition on large-scale, machine-learning endeavors. Sixth,[44] whatever approach we took would have to provide rules we could use to classify data later since we shouldn't be re-clustering things every single year. Seventh, we want it to be possible for an incident to be able to fit into two or more patterns in order to better capture the nuance of more elaborate incidents. All of these, plus the importance of getting it right, meant we've taken it slow and steady.

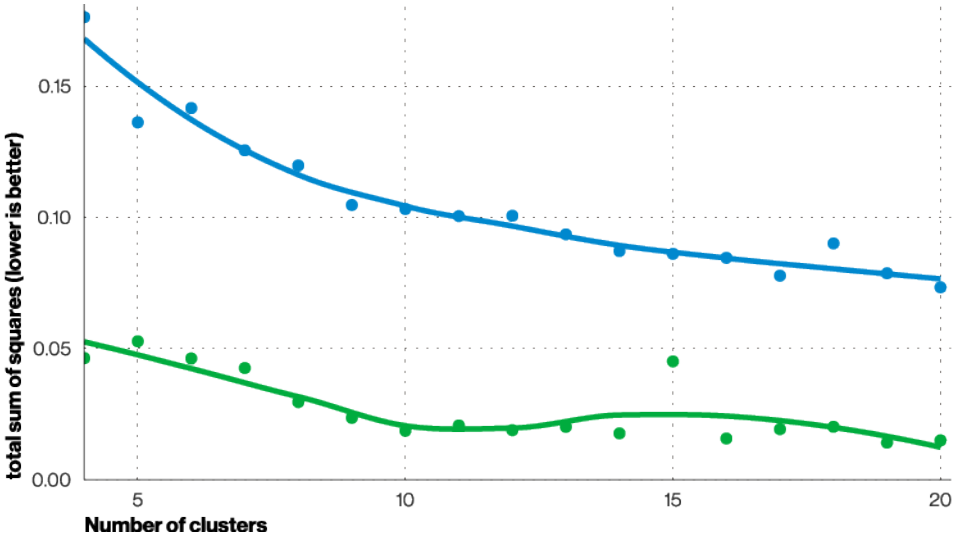



**Figure 48.** Model rating by cluster number

Before we get to what did work, let's talk about some of the things that didn't work. We started with hierarchical clustering similar to the 2014 pattern's original methodology. Unfortunately, it was too unbalanced, finding small, highly similar things instead of bigger trends; kind of like seeing the trees but not the forest. K-means clustering would have been ideal, however given the size of our data, it's simply too memory intensive due to the number of all-to-all comparisons needed. Principal Component analysis didn't penalize using lots of features enough for our needs. Latent Dirichlet Allocation was slightly better, but still not good enough. Lasso and Ridge Regression didn't converge well. Association Rules did not differentiate clusters well and would have had to be paired with a predictor. Artificial Neural Networks (ANNs) would have provided prediction but not clustering.[45] We even tried Gaussian Finite Mixture Model clustering, but it had the opposite problem of hierarchical clustering in that all the clusters were minor variants on the big themes; kinda like seeing the forest, but not the trees.[46]

What we eventually settled on was spherical k-means. It provided us the clustering benefits of k-means (ability to classify new data, find both small and large clusters, handle high dimensionality without overfitting, handle logical data, and be explainable) while also not choking on our rather large dataset. Normal k-means calculates the distance between all the rows in the dataset in the dimensional space of the number of columns. It randomly creates a set number of cluster centers and assigns the points to the closest center. It then recalculates the center of each cluster, and repeats the two steps until there are no significant changes in the cluster memberships. All those distance calculations take a lot of time and memory. Spherical k-Means improves on that by calculating cosine distance, and taking advantage of that special structure to avoid calculating full object-to-object distance matrices.[47]

43 Even I thought this list would only be three items long, but man did we have a lot of challenges.
44 Another one? We get it. It was hard.
45 We also tried ANNs for clustering, specifically Self-organizing Maps, but that didn't work either.
46 You have our permission to read this out loud as many times as you would like on first dates and/or at family get-togethers and Super Bowl® parties.
47 http://www.stat.cmu.edu/~rnugent/PCMI2016/papers/SphericalKMeans.pdf

**The new patterns provide a clear framework for us to explain the threat landscape and for you to bring it to the stakeholders in your organization.**

Even then, it would be 10 hyperparameter variations until we were sure the approach would work and an additional six cluster versions based on the 2021 DBIR data to finalize the model. We settled on 517 columns to cluster, primarily dealing with the VERIS 4A's (Action, Actor, Asset and Attribute), victim, targeted, timeline and discovery method.

We wanted to prioritize more recent incidents over older ones, and while we tried just using the last few years of data, we ultimately settled on an exponential weighting function. We used a Lloyd-[48]Forgy[49] style fixed-point algorithm with local improvements via Kernighan-Lin chains.[50,51] While we wanted pattern overlap, the spherical k-Means fuzziness parameter yielded poor results, so instead we set it to hard partitions and, after clustering, included incidents in multiple clusters if the next closest cluster(s) were almost as close as the main cluster. And voilà. We have some new patterns to play with.

From experience, we know that incident and breach data can be very different. Breaches are a subset of incidents, but many times more important than incidents for our analysis.[52] To ensure that both incidents and data breaches were reflected in the patterns, we ran clustering twice, once for each. To pick the best number of clusters, we calculated the total sum of squares (a measure of success in clustering) for several different numbers of clusters.

We then manually examined the patterns generated around the "bend" in the lines (around five for incidents and eight for breaches; see Figure 48). Eventually we settled on eight breach clusters and 10 incident clusters. After clustering, the clusters were examined and some were grouped together (five in System Intrusion; three in Privilege Misuse and Miscellaneous Errors; two in Basic Web Application Attacks, Social Engineering, and Lost and Stolen Assets; and one in Denial of Service) and then named, forming the new patterns.[53]

Table 2 is what we got for all of that work. In some places, nothing has changed. In some places, everything has changed. But, more importantly, the new patterns provide a clear framework for us to explain the threat landscape and for you to bring it to the stakeholders in your organization.

| | |
|---|---|
| **Social Engineering** | Psychological compromise of a person, which alters their behavior into taking an action or breaching confidentiality. |
| **Basic Web Application Attacks** | Simple web application attacks with a small number of steps/additional actions after the initial web application compromise. |
| **System Intrusion** | System Intrusion captures the complex attacks that leverage Malware and/or Hacking to achieve their objectives including deploying ransomware. |
| **Miscellaneous Errors** | Incidents where unintentional actions directly compromised a security attribute of an information asset. This does not include lost devices, which is grouped with theft instead. |
| **Privilege Misuse** | Incidents predominantly driven by unapproved or malicious use of legitimate privileges. |
| **Lost and Stolen Assets** | Any incident where an information asset went missing, whether through misplacement or malice. |
| **Denial of Service** | Attacks intended to compromise the availability of networks and systems. Includes both network and application layer attacks. |
| **Everything Else** | This last "pattern" isn't really a pattern at all. Instead, it covers all incidents that don't fit within the orderly confines of the other patterns.[54] |

**Table 2.** New Incident Classification Patterns

48 Lloyd, Stuart P. (1982)
49 Forgy, Edward W. (1965)
50 Dhillon, Guan and Kogan (2002)
51 Did that make sense to you? We'll be honest, we didn't read the papers. We just chose that option on the software.
52 It is the Data Breach Investigations Report, not the Data Incident Investigations Report, after all.
53 Astute readers may notice that we did not actively attempt to retain old patterns. The fact that so many of them remain is a testament to the relevancy of the 2014 patterns.
54 Like that container you keep all the cables in for electronics you do not own anymore just in case.

# Denial of Service

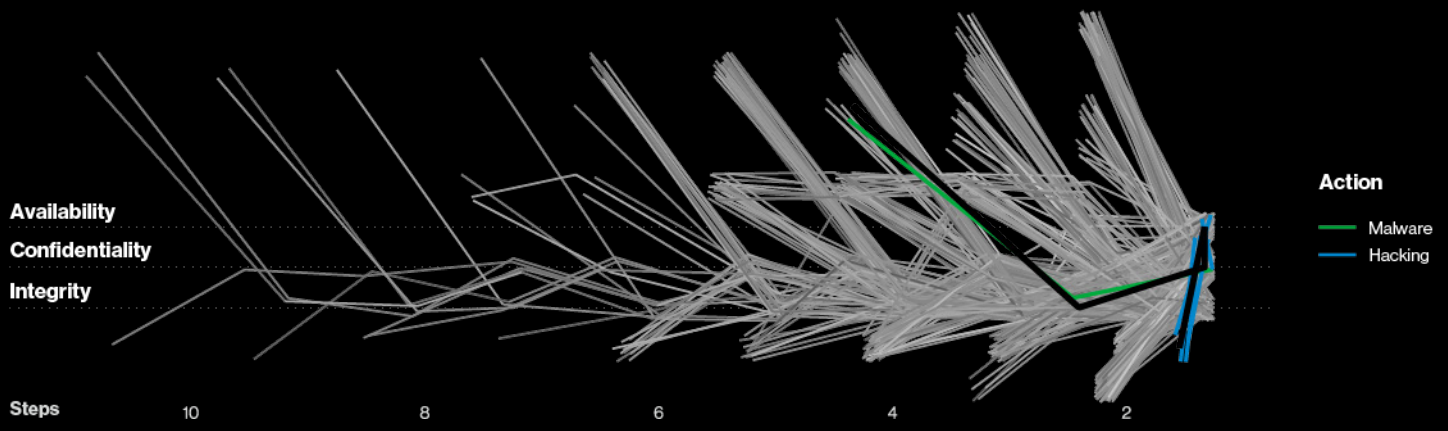



**Figure 49.** DoS incident paths (n=5)

## Summary

The Denial of Service pattern consists of attacks intended to compromise the availability of networks and systems. This pattern includes both network and application layer attacks, and is the most common pattern across incidents. However, don't let its volume concern you, as this is often one of the easiest threats to mitigate effectively.

| | |
|---|---|
| **Frequency** | 14,335 incidents, 4 with confirmed data disclosure |

Denial[55] of Service (DoS) is one of those infosec threats that actually can be addressed. This is the one you do something about for an injection of self-empowerment when you're feeling down about the latest threat du jour that you have no clue how to stop. Admittedly, as we can see in Figure 50 it's not a small threat. In fact it's the most common pattern across all incidents.

But when you look at Figure 51, you'll notice that the median bits per second (bps) of 1.3 Gbps may be only a bit (no pun intended) more than your home internet connection. Ninety-five percent of incidents fell between 13 Mbps and 99 Gbps, an easily mitigatable range. So, sign up for a DoS mitigation service and reward yourself with that cannoli you've had your eye on.

55 It's not just a river in Egypt, Harry.

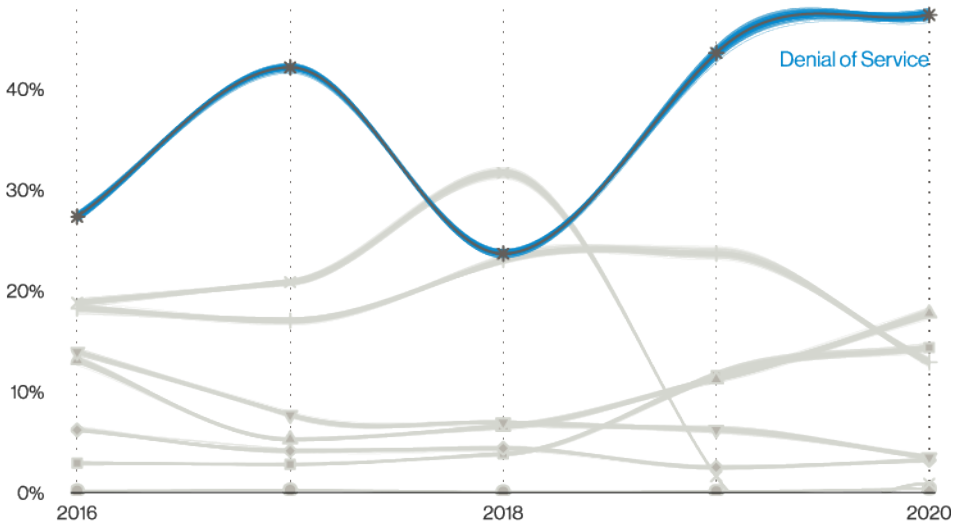



**Figure 50.** Patterns over time in incidents

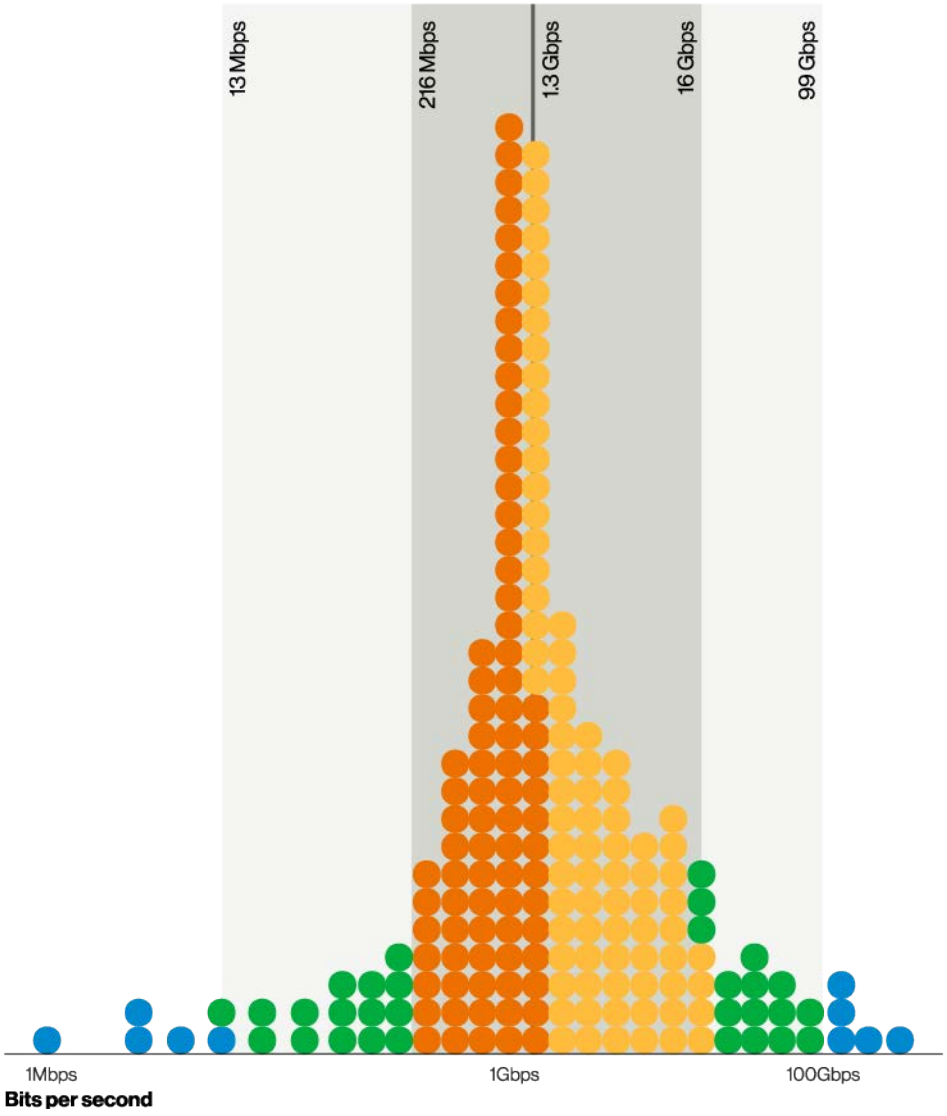



**Figure 51.** Bits per second in DDoS Incidents (n=11,306)
Each dot represents 0.5% of organizations

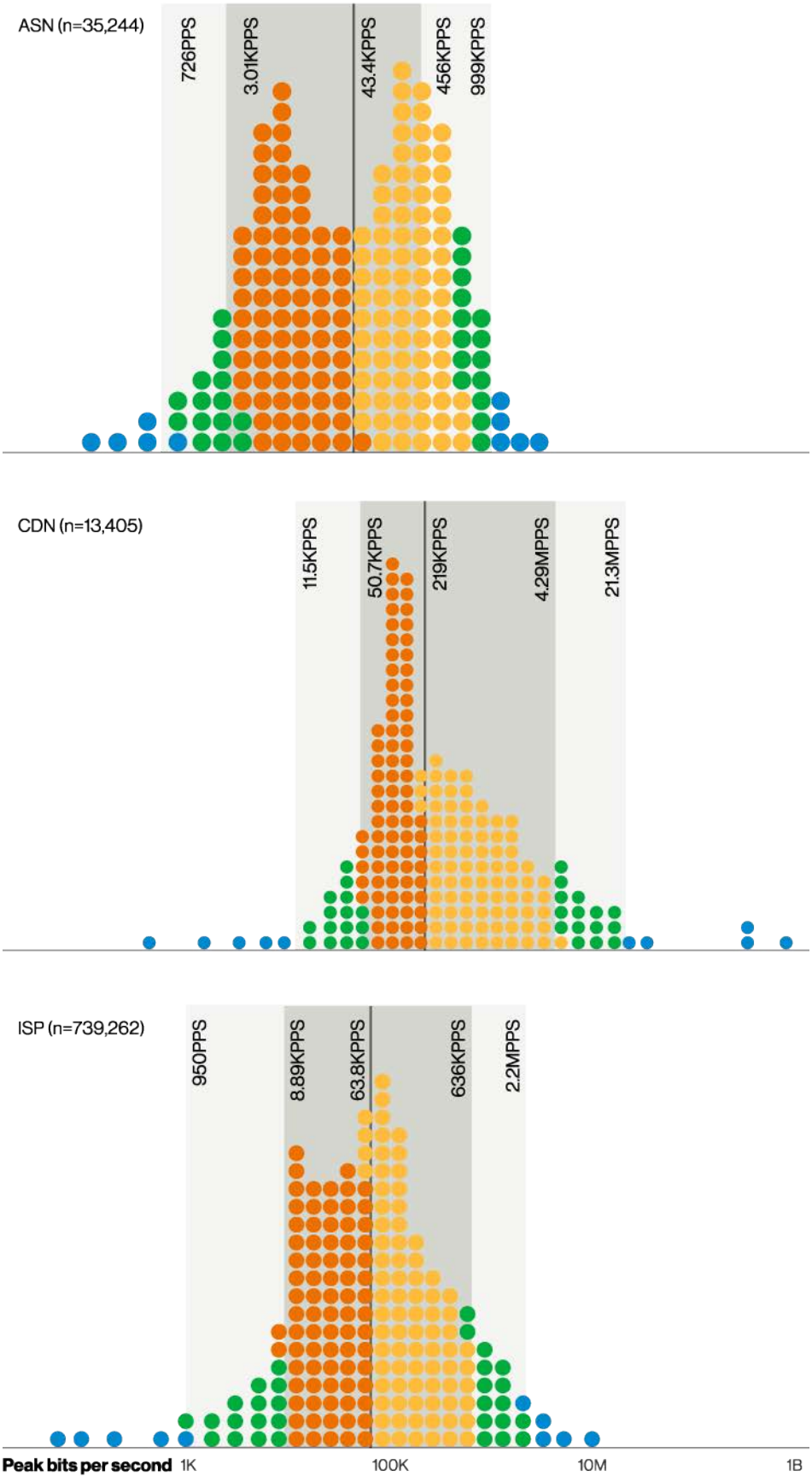



**Figure 52.** Peak PPS in various DoS locations
Each dot represents 0.5% of organizations

One reason DDoS attacks aren't more of a threat is that those mean[56] packets have to cross a lot of internet to get to you. Figure 52 covers just how much DDoS is getting blocked at various places, from Internet Service Providers (ISPs) at the start of the trip, to Autonomous System Numbers (ASNs) in the middle, to Content Delivery Networks (CDNs) that your site might sit behind. All have a hand in mitigating the attack.

In Figure 53 we take a quick look at a couple of different types of attacks. DoS attacks can be direct (packets come directly from the actor or their botnets) or reflected (actor sends packets to a vulnerable service that then reflects the packets to the victim). They can also be intended for resource exhaustion (send packets that cause abnormal load on memory or processing) or volumetric (lots and lots of packets). What we see is that there aren't many differences between the different attack types (and frankly, a single DDoS attack[57] can use multiple).

We bounce back and forth a bit between packets per second (PPS) and bits per second (BPS). We do so largely based on the data we have available, but in case that is what's keeping you up at night right now, we'd like to put your fears to rest.[58] For any given packet type (and there are several), there's a fixed range of how many bytes you can expect in the packet. You can see that in the linear nature of Figure 54. And so, whether we're using BPS or PPS, the conclusions are still the same.

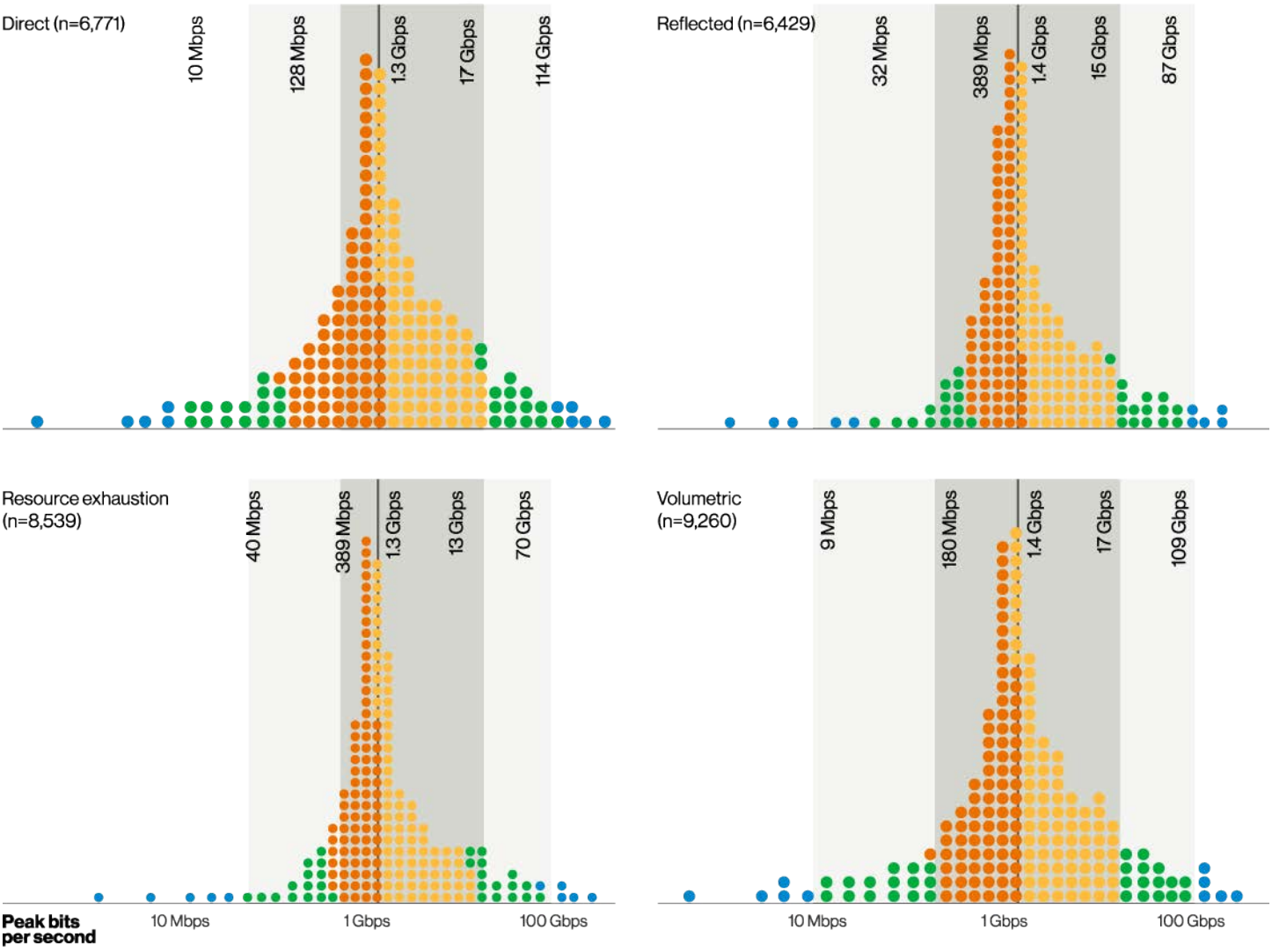



**Figure 53.** Peak BPS in various DoS types

56 Malevolent mean, not average mean.

57 In fact, what is a DDoS attack really? Does it start with the first packet and end with the last? How would we know? What if it's a different botnet at the same time? Or if it stops for a few seconds and starts again? Or... or.... When did the DBIR footnotes become the Wikipedia discussion page?

58 Metaphorically and literally.


**Figure 54.** Relationship between PPS and BPS per DoS

Figure 55 gives you an idea of the equality in DDoS packets per second. It shows that for the majority of organizations, the data is pretty spikey. Figure 56 shows predictions of a Recurrent Neural Network (RNN) trained on 450,000 DDoS attacks. All it does is predict the average DDoS timing and fails if the DDoS is anything but average. Don't spend your time worrying about predicting the next DDoS. You can't predict it. Hire a service to handle it for you and it's cannoli time.

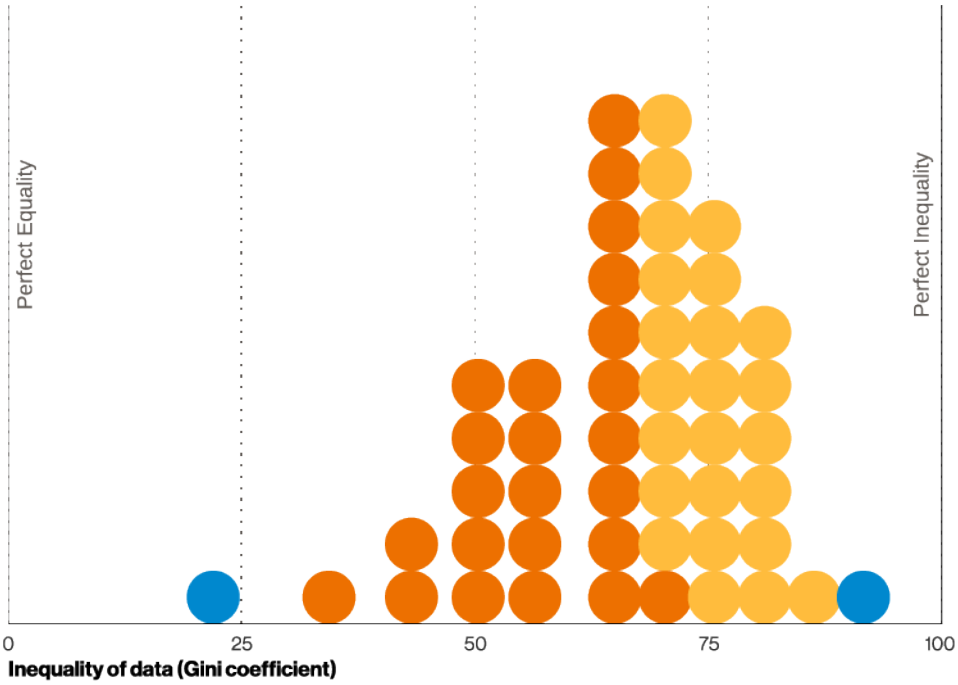



**Figure 55.** Inequality of DDoS PPS by organization (n=54)
Each dot represents 2% of organizations

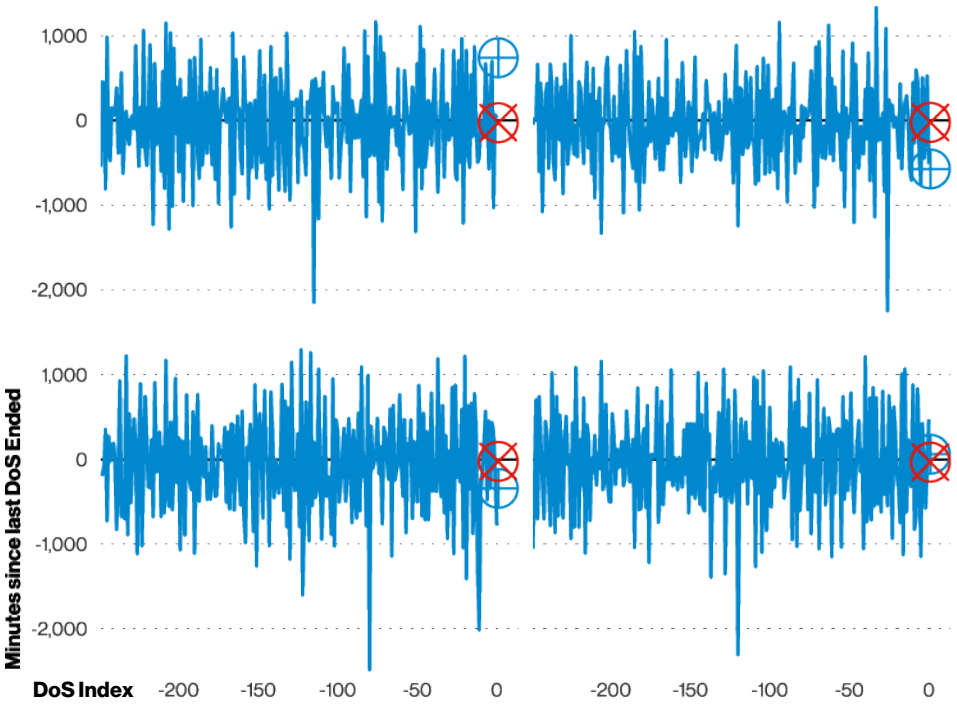



**Figure 56.** Predictions of RNN trained to predict the next DDoS

# Lost and Stolen Assets

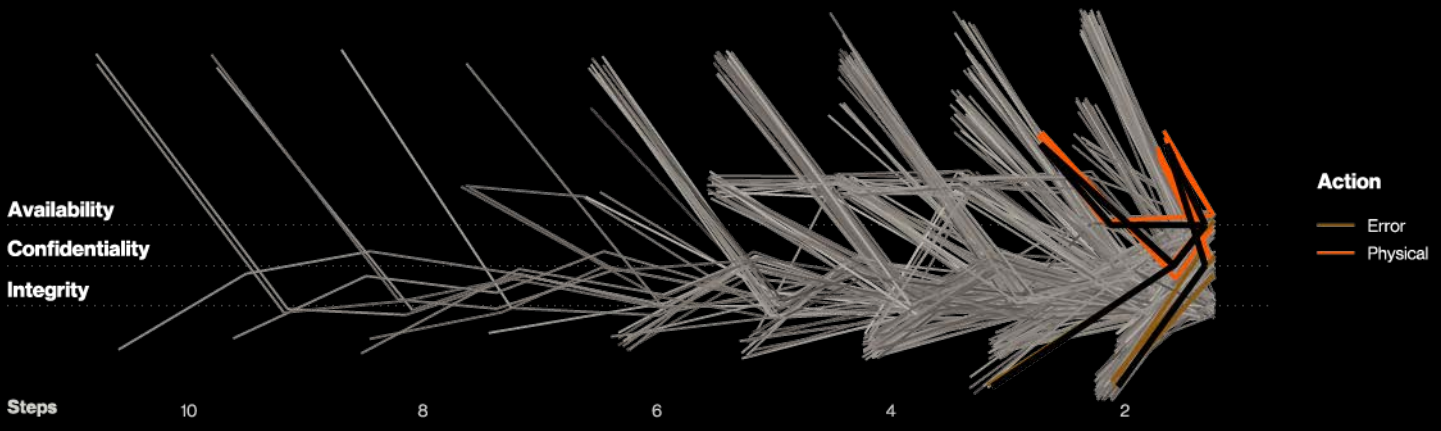



**Figure 57.** Lost and Stolen Assets incident paths (n=13)

## Summary

Devices continue to be lost or stolen, a pattern that is unlikely to change anytime soon. While the actor may be Internal (for loss) or External (for theft), the controls to protect the data on these devices remain constant.

| | |
|---|---|
| **Frequency** | 1,295 incidents, 84 with confirmed data disclosure |
| **Threat Actors** | External (87%), Internal (17%), Multiple (5%), Partner (1%) (breaches) |
| **Actor Motives** | Financial (100%) (breaches) |
| **Data Compromised** | Personal (80%), Medical (43%), Bank (9%), Other (7%) (breaches) |

We are all perhaps too familiar with that sinking feeling of reaching for your cellphone in your pocket or purse, only to find it missing. After frantically tearing the house apart, flipping seat cushions and asking anyone in close proximity to call your phone, you probably found out that you were holding it all along, or is that just us?

Anyway, this primordial fear of misplacing tiny devices that contain thousands of personal and work-related files is one of the common themes for the breaches and incidents in this pattern. Computers, documents, USB devices and cell phones end up disappearing, accidentally or otherwise. Like many of the patterns and incidents that we're covering this year, bear in mind the unique circumstances of how we've evolved our work habits over the course of 2020.

**This primordial fear of misplacing tiny devices that contain thousands of personal and work-related files is one of the common themes for the breaches and incidents in this pattern.**

This is especially true when it comes to where and how we work. The findings here might need to be taken with the tiniest speck of salt, as this is not necessarily going to be a representative year. Let's take a dive into the data.

# Steady-state thefts and error

While many things have changed over the last year, some things haven't changed a great deal in this pattern. One of those things is that Error trumps Theft in incidents. In our data, much like previous years, Errors in which some Internal user accidentally mislays an asset and reports the loss is significantly more common than someone reporting an asset stolen. However, for an organization this is more or less the same problem: You now have to know what was on that device, how was it protected, and how you are going to respond. The distinction in cases like this is often a moot point since you're probably going to have to remotely wipe the device either way.

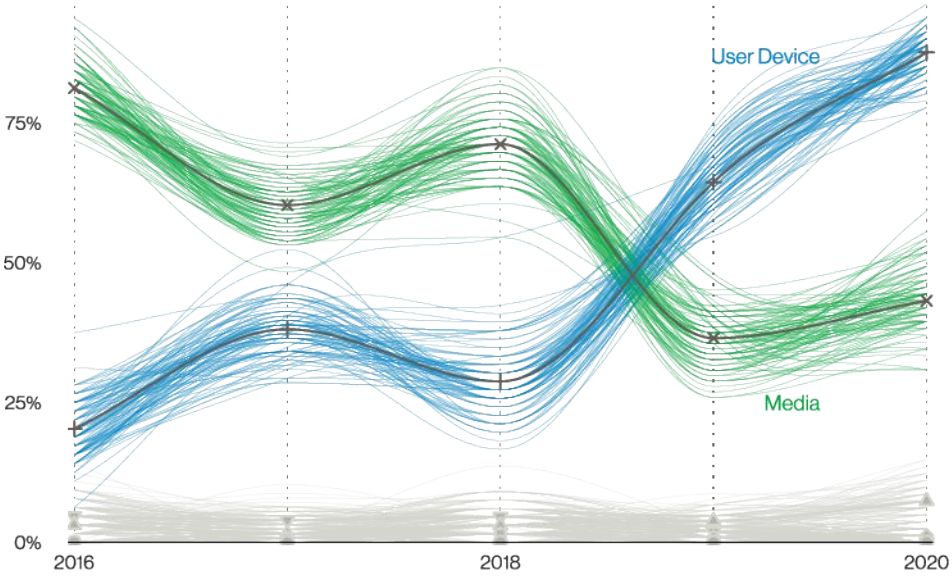



**Figure 58.** Assets in Lost and Stolen Assets breaches over time

# Would you like paper or silicon for your data breach?

One of the trends that we have noticed over the last few years is the transition from Media (such as Documents) to User devices (such as Mobile phones) being the main assets involved in Lost and Stolen breaches. If we needed a barometer as to when digital transformation occurred, we could probably point back to 2019 when, for the first time in our dataset's history, User devices were more frequently stolen and lost than Documents. This year about 43% of the breached assets with known data disclosure were Media while the rest are Desktops and laptops (Figure 58). For incidents where we don't know if there's a confirmed breach, cell phones were lost or stolen the most. Not that we're gambling people, but if we were to place money on whether or not this trend will continue, we would probably take the over, since many new organizations, schools and businesses had to quickly pivot to a remote workforce.

The type of data lost with the majority of known data breaches involves loss of Personal data, quickly followed by Medical data, which really shouldn't be too surprising. The amount of legislation regarding privacy breach disclosure (medical and otherwise) would explain why we see this in our data. And lastly, when it comes to discovering that an asset is lost or stolen (Figure 59), your best line of detection won't be the next-gen AI, but your employees themselves. Make sure that they are provided with a means to easily report any lost or stolen assets to your organization. For instance, if they lose their phone they have a number they can call...wait, never mind. The quicker the organization knows, the better position they'll be in to respond. Something...something...obligatory "hindsight is 2020" joke.

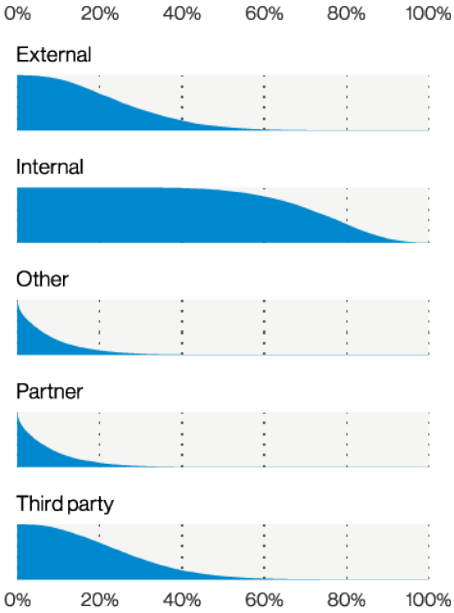



**Figure 59.** Discovery methods in Lost and Stolen Assets breaches (n=9)

# Miscellaneous Errors

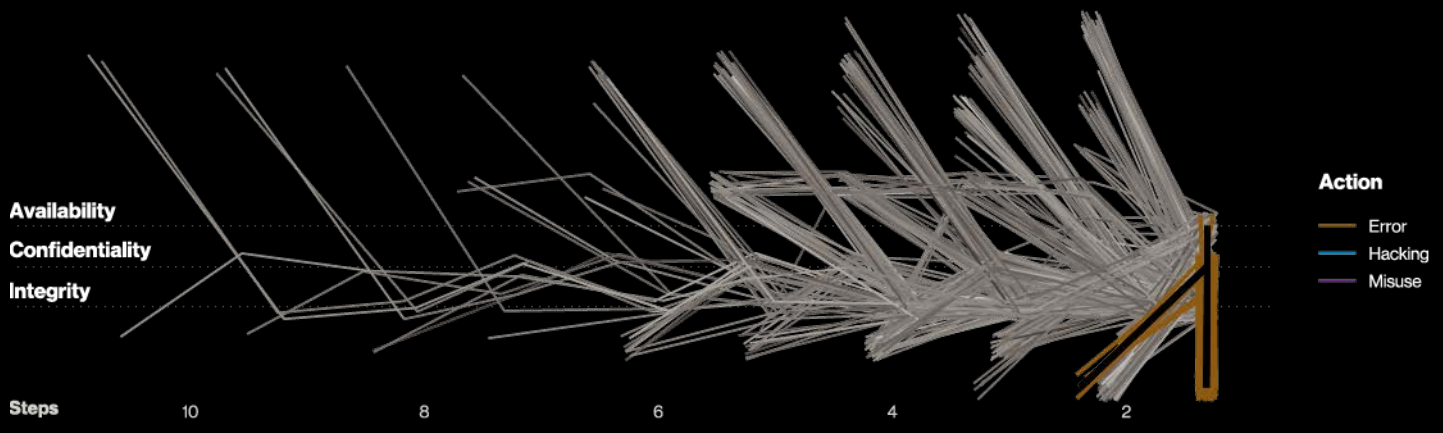



**Figure 60.** Miscellaneous Errors incident paths (n=126)

## Summary

Errors are unintentional actions, typically taken by an Internal actor, but Partner actor errors also occur. Misconfiguration of database assets being found by Security researchers is a growing problem. Employees sending data to the wrong recipients also continues to be a significant issue.

| | |
|---|---|
| **Frequency** | 919 incidents, 896 with confirmed data disclosure |
| **Threat Actors** | Internal (99%), Partner (1%), Multiple (1%) (breaches) |
| **Data Compromised** | Personal (79%), Medical (17%), Other (13%), Bank (13%), Credentials (13%) (breaches) |

The Miscellaneous Errors pattern should be a familiar frenemy from years gone past. We have included this pattern since the beginning, and the errors have remained constant. What can we really say about this pattern? Humans make mistakes, often at scale. This pattern consists of Internal and/or Partner actors only.

We show the breakdown for Internal actors in Figure 61, and they are relatively intuitive since both system administrators and developers typically have privileged access to data on the systems they maintain. However, the adage of "to whom much is given, much is expected" assuredly applies here. When people in these roles do make mistakes, the scope is often of much greater significance than the foibles of an average end-user.

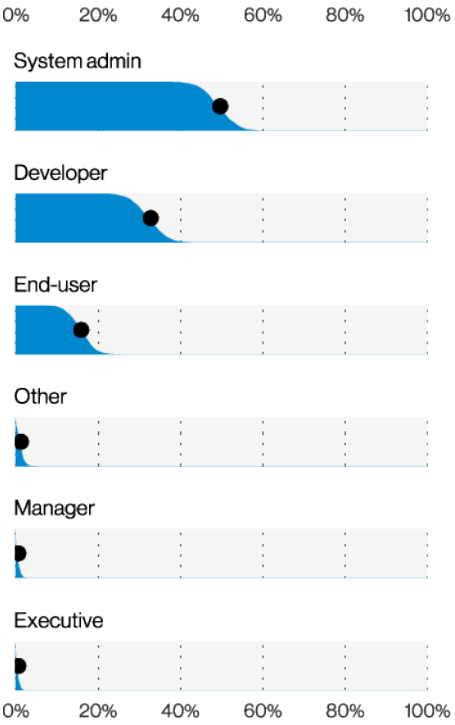



**Figure 61.** Internal actor varieties in Miscellaneous Errors breaches (n=157)

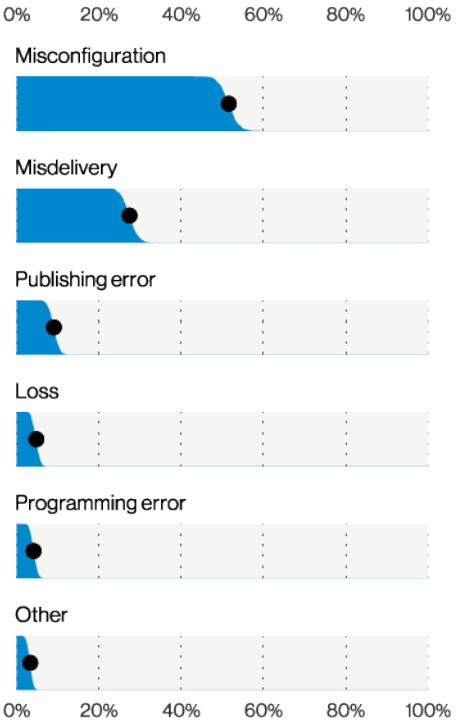



**Figure 62.** Top Error varieties in Miscellaneous Errors breaches (n=609)

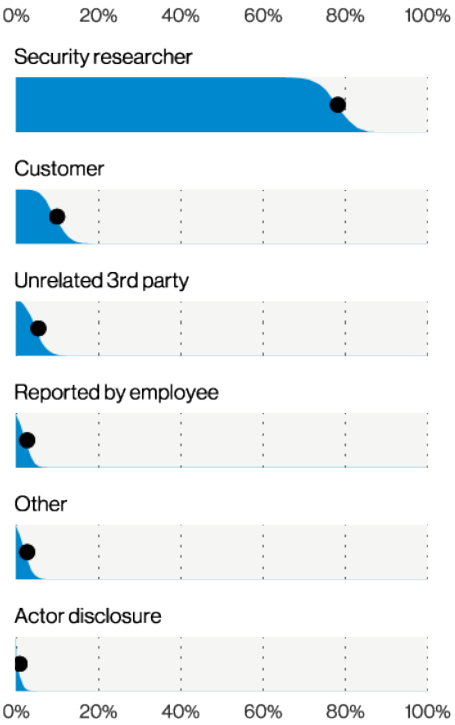



**Figure 63.** Discovery method varieties in Miscellaneous Errors breaches (n=110)

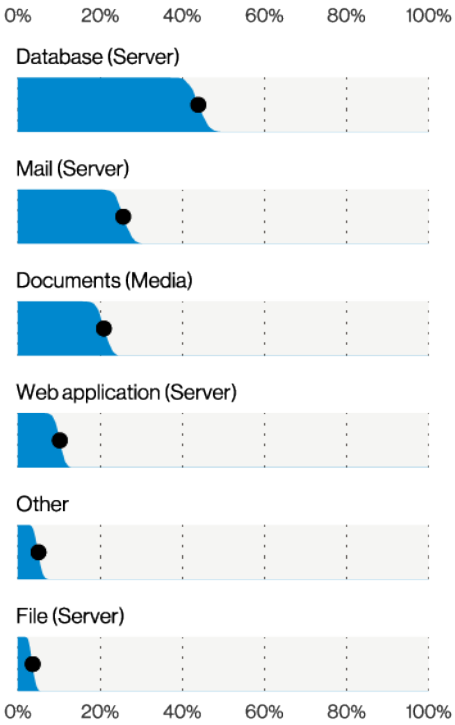



**Figure 64.** Top Asset varieties in Miscellaneous Errors breaches (n=635)

**Sadly, Misdelivery remains alive and well in our dataset, and while a number of these breaches are electronic data only (e.g., email to the wrong distribution list), there remains a significant number that involve paper documents.**

Allow us to take you on a tour of pairings—no, not wine and cheese, but Actors and Actions. Given the pairing of sys admins and developers with the Misconfiguration action varieties (Figure 62), you can imagine that this combination can wreak havoc on the confidentiality of an organization's data, or that of their customers' or employees'.

The other pairing we frequently observe is data stores (such as relational or document databases, or cloud-based file storage) being placed onto the internet with no controls, combined with the security researchers who search for them (Figure 63). These rather undesirable combinations have been on the rise for the past few years.

Sadly, Misdelivery remains alive and well in our dataset, and while a number of these breaches are electronic data only (e.g., email to the wrong distribution list), there remains a significant number that involve paper documents (Figure 64). These are particularly common in industries in which large mass mailings are a preferred method of getting information to the customer base. One example being when the envelopes become out of sync with the contents. Many of these events could be avoided by a basic sample check at different points during the mailing process. Nevertheless, we continue to see this occurring regularly, but rarely with any of our bills (those always seem to arrive on time).

## Personal data is the most commonly disclosed data type in these cases by a wide margin.

The Assets involved in Error actions run the gamut, from the aforementioned misconfigured databases to physical documents and user devices (Figure 64). A certain portion of this is from Asset loss, although if the device is configured such that unauthorized data access cannot be confirmed, it would be considered an incident rather than a breach.

Personal data is the most commonly disclosed data type in these cases by a wide margin (Figure 65). Medical data is also exposed in this manner, but not nearly as often. The other data varieties represented appear in much smaller quantities.

Just take a gander at that lovely Discovery timeline in Figure 66. See how it flexes all of those breaches discovered within hours and days of the event? Surely this is the story of successful detective controls! Actually, it may be because people usually realize they goofed fairly quickly. But just in case they don't, they have the added safety net of legions of devoted Security researchers out there scouring the internet with their specialized search engines just looking for mistakes.

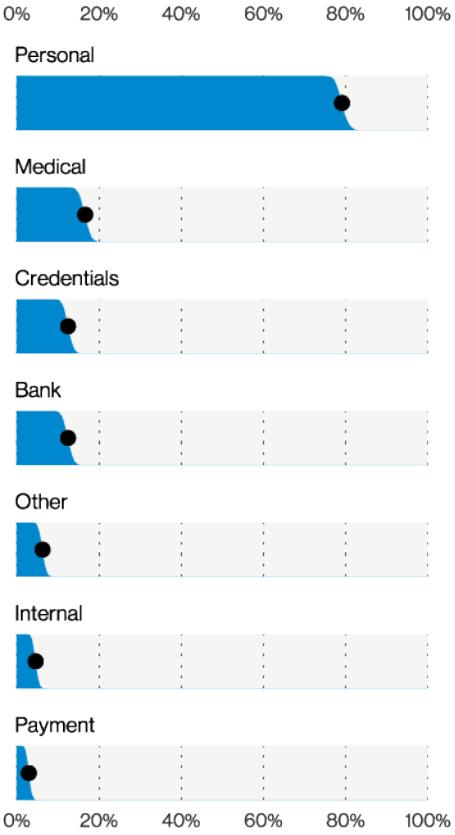



**Figure 65.** Top Data varieties in Miscellaneous Errors breaches (n=839)

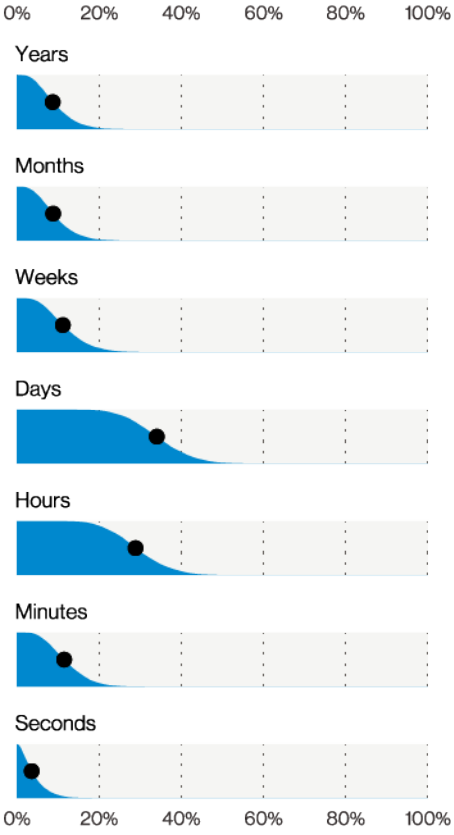



**Figure 66.** Discovery timeline in Miscellaneous Errors breaches (n=39)

# Privilege Misuse

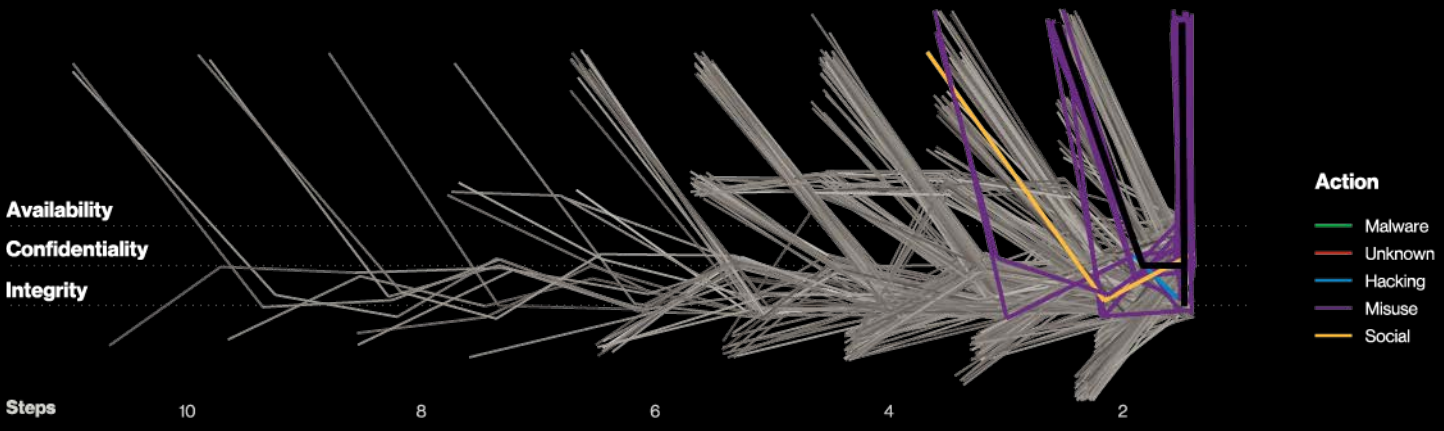



**Figure 67.** Privilege Misuse incident paths (n=51)

## Summary

Privilege abuse was the most common action type for this pattern, with the majority of actors being Financially motivated. The most common data type stolen was Personal information, and somewhat surprisingly, the rise in remote workers did not appear to have a noticeable effect on Misuse.

| | |
|---|---|
| **Frequency** | 265 incidents, 222 with confirmed data disclosure |
| **Threat Actors** | Internal (99%), Multiple (9%), External (8%), Partner (2%) (breaches) |
| **Actor Motives** | Financial (64%), Fun (17%), Grudge (14%), Espionage (9%), Convenience (3%), Ideology (1%) (breaches) |
| **Data Compromised** | Personal (64%), Other (35%), Medical (27%), Internal (19%) (breaches) |

This pattern is an uncomfortable one—this is where the people we trust betray us. Privilege Misuse is our colleagues deciding (for a number of reasons) to take their access and use it to pilfer data they are not authorized to take, or use it in ways they really shouldn't.

This is the malicious Internal actor pattern—the wicked stepsister of the innocent Miscellaneous Errors pattern. While Miscellaneous Errors is perhaps a bit of a klutz, Privilege Misuse is actively piling chores on us to make sure we don't get to attend the ball.

Now that we've stretched that metaphor right to the breaking point, let's move on. You can see in the At-a-Glance table that most of the cases in which there is Misuse there is also a confirmed data breach. While these are almost exclusively perpetrated by Internal actors (or occasionally by Partners), this is the pattern where we most frequently see evidence of multiple types of actors working in concert.

Most Internal actors are motivated by greed—they're trying to cash in on the data they steal. A much smaller percentage are in it for the LOLs. Fewer still are holding a grudge against their employer. And finally, we get to those who are doing this to start a competing business or benefit their next employer. The last three make up a small percentage of the whole, and the main takeaway here is that people are frequently financially motivated—whether they have trusted access or not.

# How they do what they do

The most common variety of Privilege Misuse is Privilege abuse (Figure 68). The second-place spot went to Data mishandling. Note, the Other bar is a combination of the remaining varieties added together. The majority of vectors for those were described as network-based access of some sort to the assets. We would have expected an appreciable increase in people performing Misuse from home, given the increase of those who are working remotely due to the pandemic. However, we did not see an increase from Remote Access as a vector, but it may simply be that the detail was left out of the data when the cases were worked, or organizations aren't able to detect and report on this vector of access.

There were a variety of data types stolen in these cases, with Personal being in the lead, as shown in Figure 69. But others included Medical, Internal, Bank and even Secrets. It usually comes down to the type of data the individual can access that drives which variety they take.

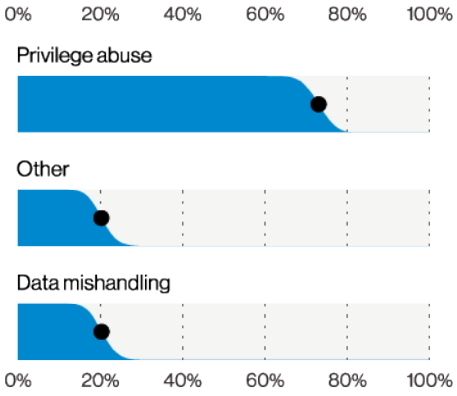



**Figure 68.** Top Misuse varieties in Privilege Misuse breaches (n=175)

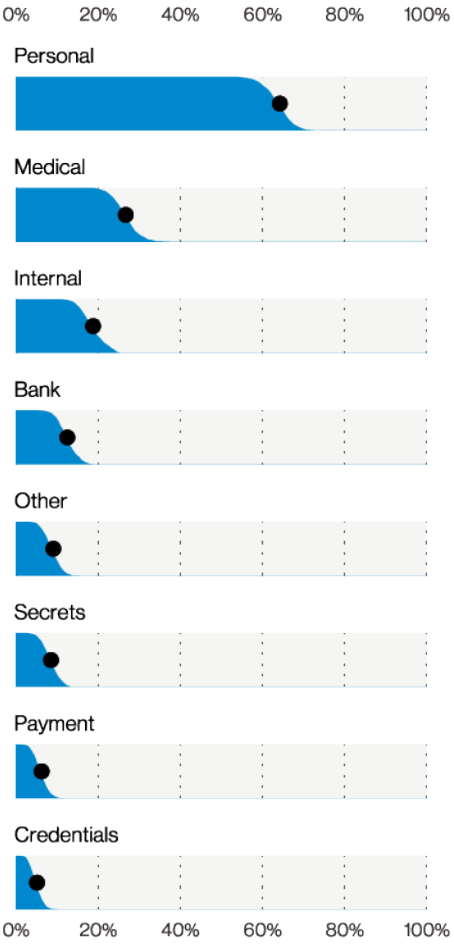



**Figure 69.** Top Data varieties in Privilege Misuse breaches (n=176)

# Discovery all

As we mentioned in the Timeline section, Misuse breaches can be difficult to detect. When one compares the Discovery timeline for this pattern vs. the overall dataset, it really illustrates that point, with more Privilege Misuse cases taking years to discover than non-Privilege Misuse cases (Figures 70 and 71).

The three longest timelines (weeks, months and years) show up even with each other for Misuse cases this year. In reality, most organizations have tailored their controls primarily to find people trying to get in from the outside. But for organizations that have especially sensitive data, such as Healthcare, along with regulatory requirements that make reporting mandatory, it showcases the need for detective controls that can quickly catch this kind of misuse. Until they are in place, and tested, people will continue their thieving ways.

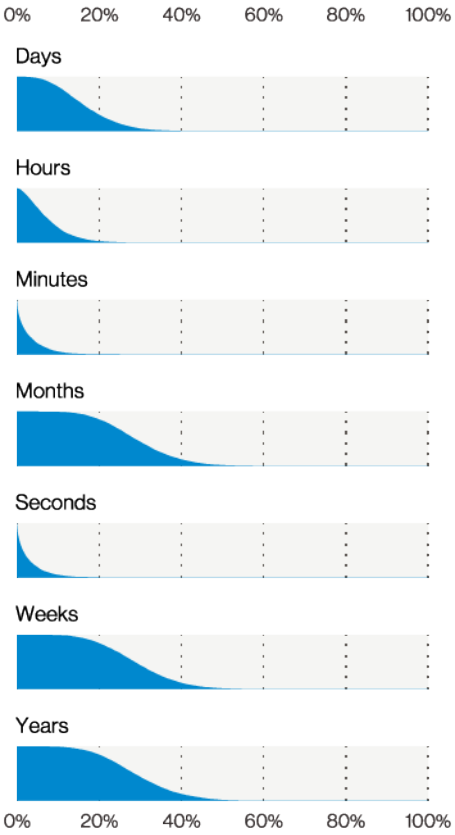



**Figure 70.** Discovery timeline in Privilege Misuse breaches (n=22)

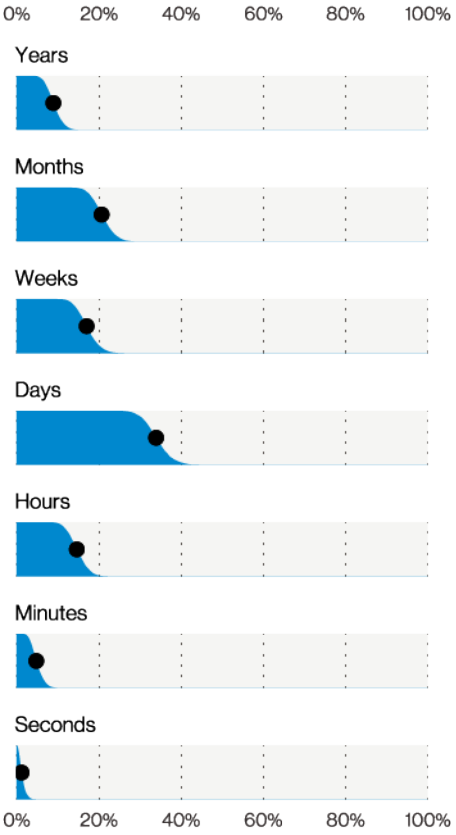



**Figure 71.** Discovery timeline in 2021 breaches (n=195)

# Social Engineering

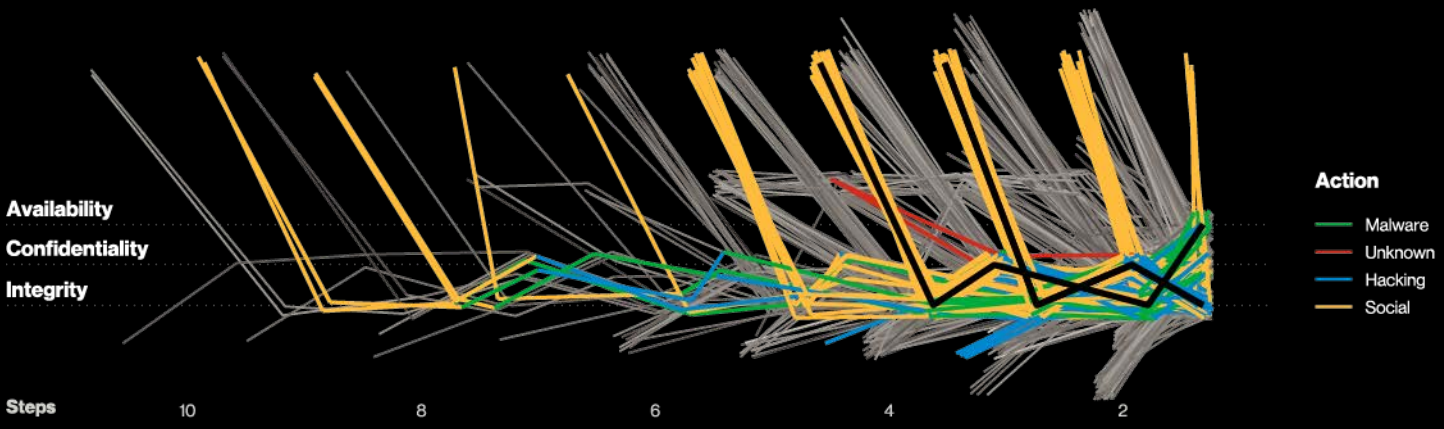



**Figure 72.** Social Engineering incident paths (n=103)

## Summary

Phishing is responsible for the vast majority of breaches in this pattern, with cloud-based email servers being a target of choice. Business Email Compromises (BECs) were the second most common form of Social Engineering. This attack scenario reflects the meteoric rise of Misrepresentation, which was 15 times higher than last year in Social incidents. Additionally, Social Engineering attacks often result in the loss of Credentials. This pattern saw those stolen credentials used in both Hacking and Malware attacks.

| | |
|---|---|
| **Frequency** | 3,841 incidents, 1,767 with confirmed data disclosure |
| **Threat Actors** | External (100%) (breaches) |
| **Actor Motives** | Financial (95%), Espionage (6%) (breaches) |
| **Data Compromised** | Credentials (85%), Personal (17%), Other (9%), Medical (4%) (breaches) |

Anyone who has been around children for an extended period of time is well acquainted with social engineering. Watching them trying to convince a parent (or sibling) to see things their way can be quite entertaining. Not that you can blame them. We're all trying to get ahead. But none of us wants to be the one handing over something we'd rather keep just because the actor, whether they are three years old or 30, has a really good story about why they need it.

We've definitely seen a jump in Social Engineering breaches as a pattern from last year with an overall upward trend since 2017. For the past couple of years, it appears to be correlated to an uptick in the compromise of cloud-based mail servers. What we cannot say is why email is so enticing to threat actors.[59] Maybe it's for the email addresses themselves. Maybe it's for the internal information they contain. Maybe it's for the creds, personal and other monetizable information. Or it could simply be that they want to repurpose the server to send more malicious emails out. Sometimes it's best to admit when you just don't know.

Hopefully it is not a surprise that all Social Engineering incidents have a Social action,[60] but as you can see in Figure 72, Malware and Hacking pop up as well.

59 Just like the old Defcon adage the person on stage (or in this case writing the report) is probably not the smartest person in the room. Maybe in this case, the smartest person is you. If you have data showing what threat actors are doing with all the email accounts they're compromising, give us a holler.

60 Mostly delivered by email.

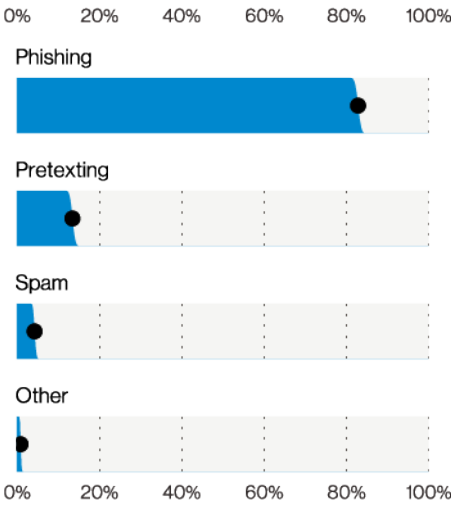



**Figure 73.** Top Social varieties in Social Engineering incidents (n=3,810)

A lot of Social Engineering breaches steal Credentials[61] and once you have them, what better thing to do than to put those stolen creds to good use, which falls under Hacking. On the other hand, that Phishing email may have also been dropping Malware, which tends to be a Trojan or Backdoor of some type (Figure 74), a trap just waiting to be sprung.

As with past years, Social actions are predominantly Phishing, though Pretexting, normally associated with the BEC,[62] also makes a strong showing. Remember those children and their great stories? This is the grownup version of why they need what you have.

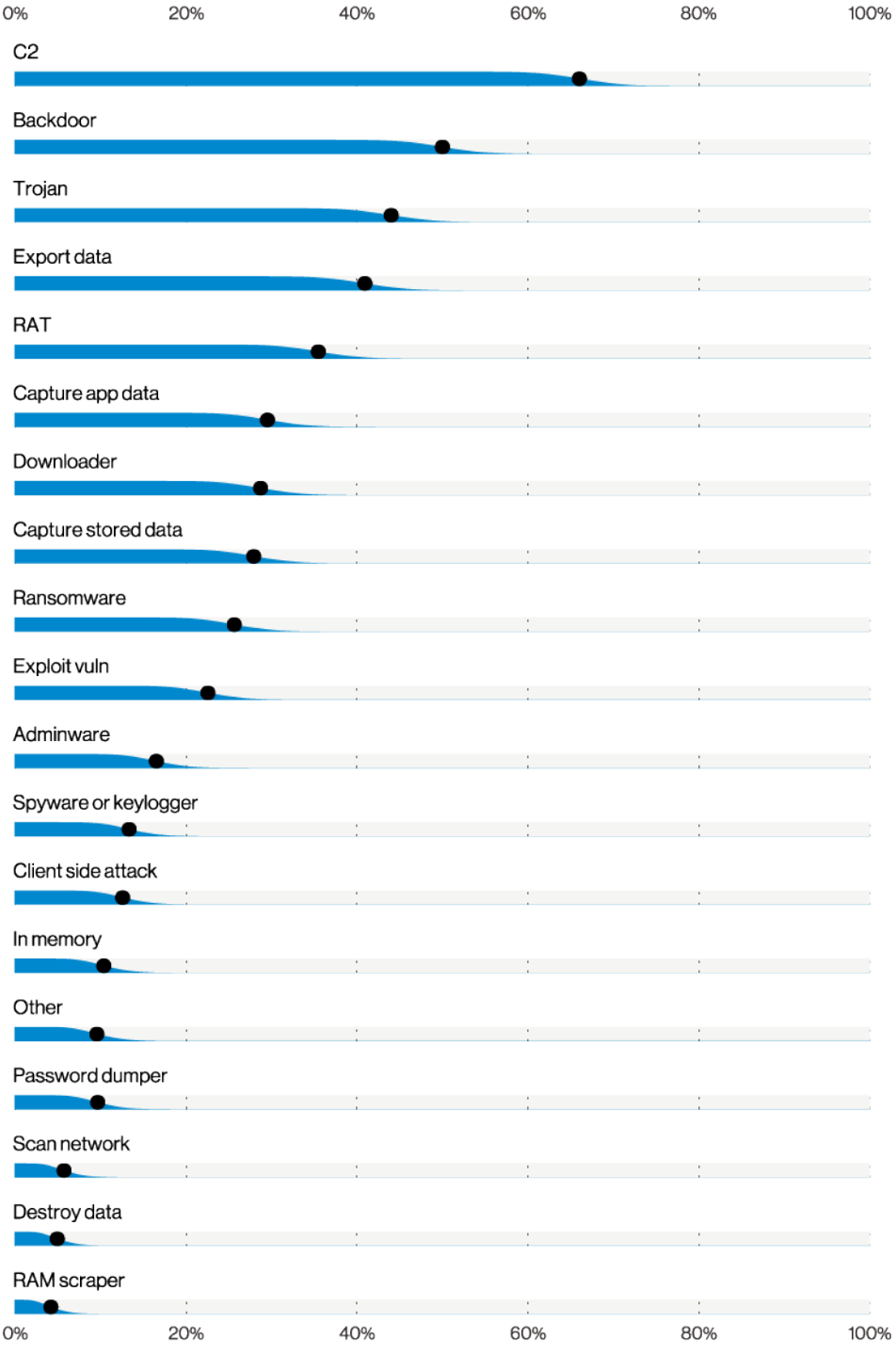



**Figure 74.** Malware varieties in Social Engineering incidents (n=130)

61 Though we'd be remiss to overlook the second most common data variety: Personal. It's just that it's kinda obvious that if someone's got your email, they've probably also got personal info.

62 Fun fact, BEC doesn't even have to compromise a business email address. Your.CEO@davesmailservice.com comes up all too often in our dataset.

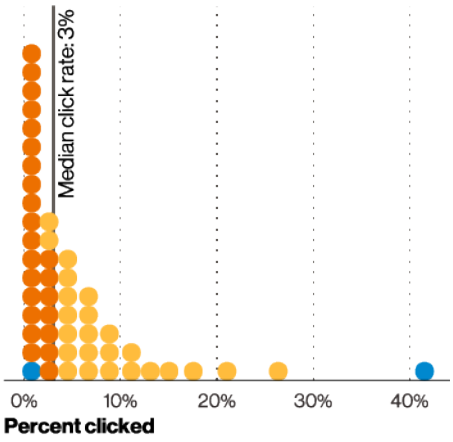



**Figure 75.** Click rate for organizations in their last phishing campaign (n=18,177) Each dot represents 2% of organizations.

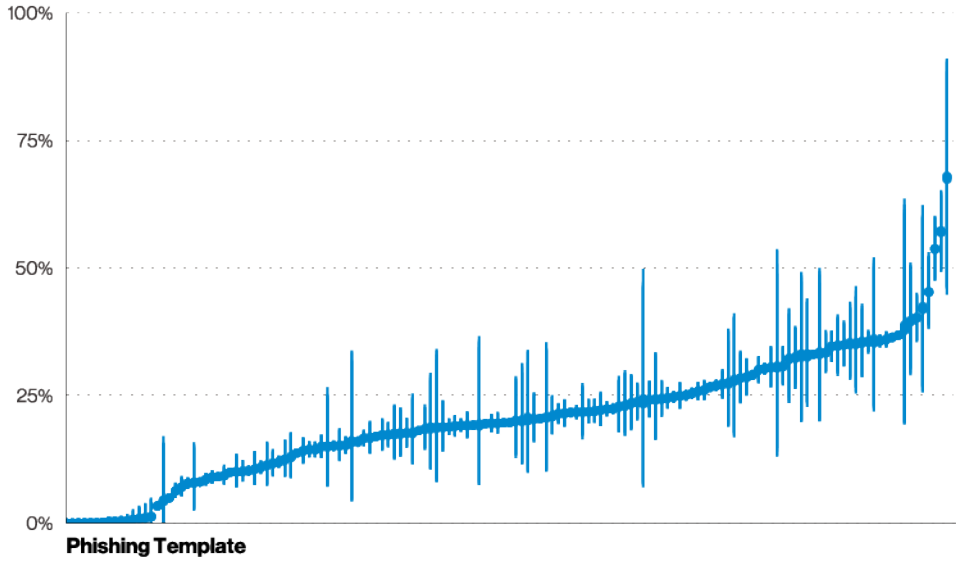



**Figure 76.** Percent of people likely to click various phishing simulation templates (n=1,186,766) The bars are our confidence. A bigger bar means less confident.

The good news about Phishing is that the click rate in phishing simulations is down to a median of 3%. But as we can see in Figure 75, it's not "most companies are around 3%." Instead, there's a long tail of companies with far larger click rates.

The phishing email itself has a lot to do with the click rate. An analysis[63] of 150 phishing templates found that the expected click rate varied significantly. In Figure 76, you can see the click rate could be anywhere from almost none to expecting over half of respondents to click. Additionally, real phishing may be even more compelling than simulations. In a sample of 1,148 people who received real and simulated phishes, none of them clicked the simulated phish, but 2.5% clicked the real phishing email. Finally, phishing volumes are very unequal. As you can see in Figure 77, no organizations experienced consistent malware by email. On the other hand, most experienced just a few days with extremely high malicious email volumes.

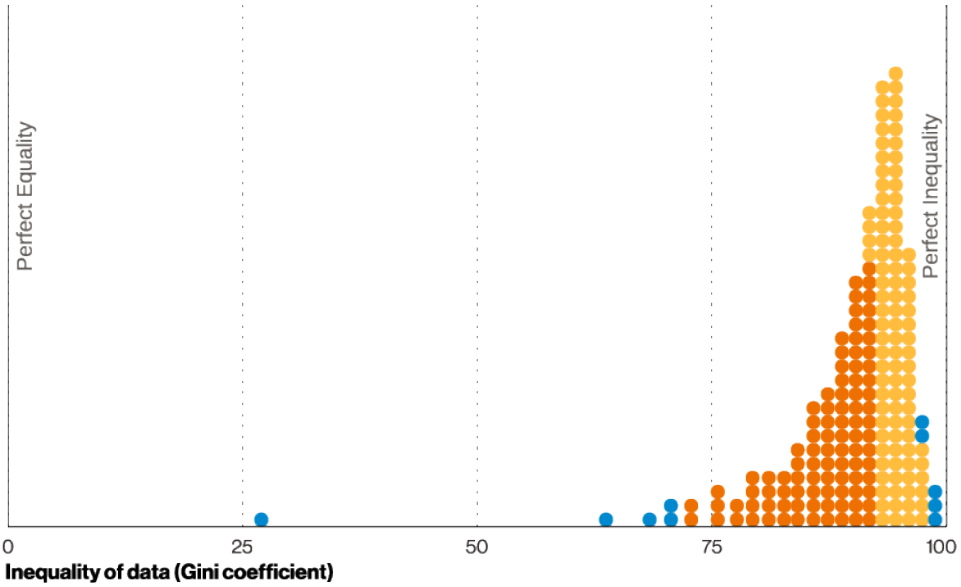



**Figure 77.** Inequality of Malware phishes per day (n=1,767) Each dot represents 2% of organizations

63 Pretty vague huh. We figured it sounded better than "a Markov Chain Monte Carlo Mixture Model." That's just downright scary. (Though we totally did it.)

# Engineering Incidents

Figure 78 reveals another concerning stat: The majority of Social Engineering incidents were discovered externally. Out of the top varieties in Figure 79, only one (Reported by employee) is internal. This means that when employees are falling for the bait, they don't realize they've been hooked. Either that, or they don't have an easy way to raise a red flag and let someone know they might have become a victim. The former is difficult to address, but the latter is simple and should be implemented—something as basic as a well-publicized email of cert@yourorganizationhere.com (which, of course, is monitored) can give you a heads-up that something is amiss.

Finally, we would be remiss if we let the BECs slide by. They were the second most common form of Social attacks and, as Figure 80 shows, they're continuing to take off. Misrepresentation is 15 times higher than last year in Social incidents.[64] Together with Phishing and Pretexting, Misrepresentation helps drive the BEC juggernaut. And while the impact can be hard to quantify in some kinds of incidents, with a BEC it's a lot easier.[65] As we point out in the Impact section, of the 58% of BECs that successfully stole money, the median loss was $30,000 with 95% of BECs costing between $250 and $984,855. Not bad for a day's work.

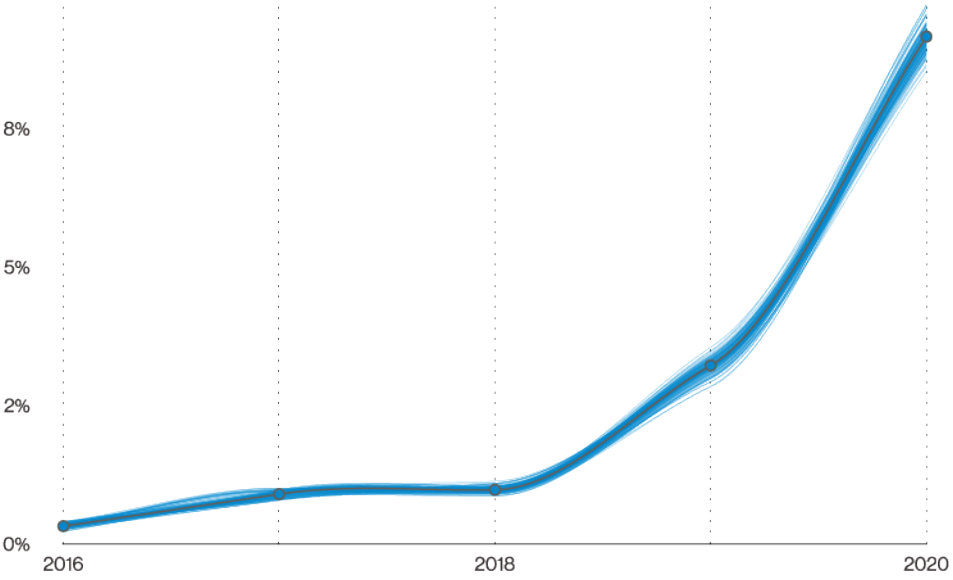



**Figure 80.** BEC over time in non-DoS incidents

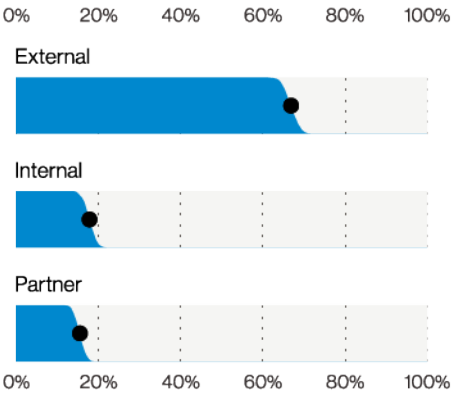



**Figure 78.** Discovery methods in Social Engineering incidents (n=691)


**Figure 79.** Top External discovery method varieties in Social Engineering incidents (n=234)

64 We mentioned that BECs don't even have to compromise an email address, but when they do, using it to send the malicious email is considered a Misrepresentation integrity compromise.

65 Readers may be familiar with the old cyber shanty regarding phishing. "Soon may the phisherman come, to bring us creds to pwn for fun, one day, when the hacking's done, we'll take our crypto and goooo....."

# Building Cybersecurity Culture

**Masha Arbisman**

**Behavioral Engineering Manager for the Paranoids, the information security team at Verizon Media**

The conversation about data leakage has flipped from "if" to "when" a company will be breached by malicious actors. The fight against cyber breaches continues to depend on an organization's ability to train and adapt its members' behaviors to protect against actions such as credential theft, social engineering, and user error.

Verizon Media believes the simulations and training offered by most security education teams do not mimic real life situations, do not parallel the behaviors that lead to breaches, and are not measured against real attacks the organization receives. This is why it is important to progress from the traditional security awareness model to that of using behavioral science to change the habits that lead to attack path breaking actions.

Huang and Pearlson's cybersecurity culture model[66] suggests that cyber secure behaviors are driven by the values, attitudes, and beliefs of an organization, which are visible at the leadership, group, and individual levels. Influencing how employees prioritize, interpret, learn about, and practice cybersecurity allows managers a way to create a cybersecurity culture within the organization.

We used the Huang and Pearlson model in combination with behavioral science techniques[67] to develop a three-step approach[68] to drive experimentation and make decisions aimed at improving the security behaviors of employees. Over two years, the approach tripled adoption of a password manager and decreased the overall phishing susceptibility of employees by half, as calculated by the results of our phishing simulation programs correlated with real company attacks measured by their Security Operations team.

There is no singular approach to minimizing the human risks that lead to breaches. Each corporation experiences different flavors of the same types of attacks and must customize their behavioral engineering and cybersecurity education programs accordingly. The Verizon Media data-driven and measurable approach can be used as a starting point to building customized programs.

66 https://scholarspace.manoa.hawaii.edu/bitstream/10125/60074/0634.pdf
67 See the Dictionary of Terms in the case study in the next footnote for a list of techniques.
68 https://cams.mit.edu/wp-content/uploads/Verizon-Media-CyberCulture-Paper.pdf

# System Intrusion

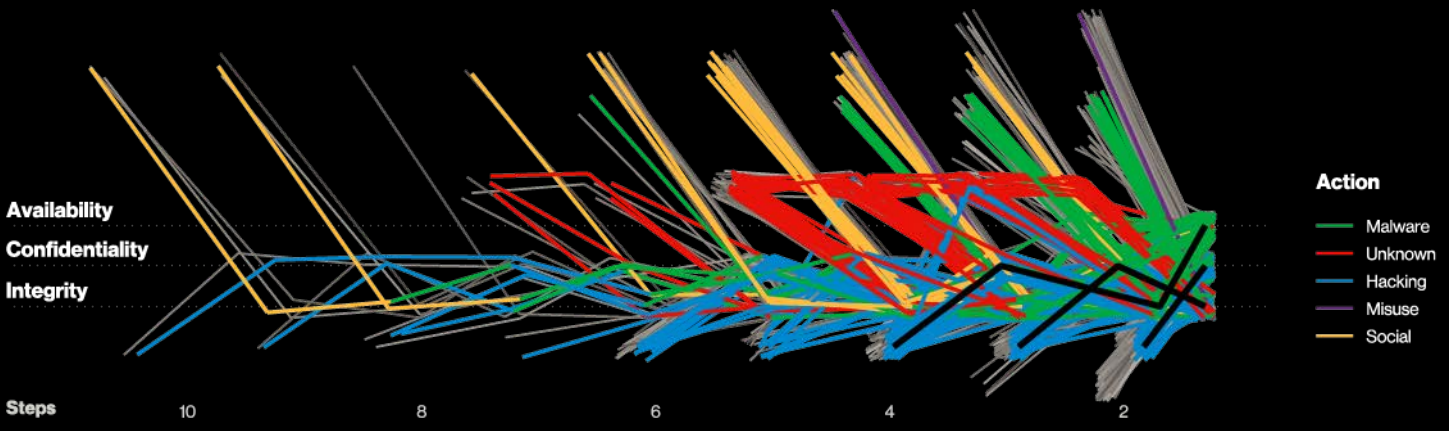



**Figure 81.** System Intrusion incident paths (n=251)


## Summary

This new pattern consists of more complex attacks, typically involving numerous steps. The majority of these attacks involve Malware (70%), usually of the Ransomware variety, but also of the Magecart attack type used to target payment card data in web applications. Hacking (40%) also appears in many attacks and most often consists of the Use of stolen credentials or Brute force attacks.


| | |
|---|---|
| **Frequency** | 3,710 incidents, 966 with confirmed data disclosure |
| **Threat Actors** | External (93%), Internal (8%), Multiple (1%) (breaches) |
| **Actor Motives** | Financial (95%), Espionage (6%) (breaches) |
| **Data Compromised** | Personal (48%), Other (35%), Credentials (33%), Payment (24%) (breaches) |

Not only is this one of the "newer" patterns, it certainly is one of the more interesting ones to talk about, as you'll see in a few. This pattern consists of the more complex attacks, often involving multiple steps as the attackers move through the environment to find the hidden stash of wealth.

In previous years, some of the incidents we discuss in this section would have fallen under the Cyber Espionage pattern, which would have captured most of the hijinks of Nation-states and their affiliated actors looking for Secrets. Still others would have been found in the Crimeware pattern, and lastly, the often-forgotten POS server attacks that target servers processing credit cards. Our new System Intrusion pattern is intended to capture those (sometimes only slightly) more elaborate "human-operated" attacks regardless of the motive the actors present. Without further ado, let's get into the details.

## Actors in chains

As "trained" data scientists, when we're presented with complex data and detailed charts like Figure 81, representing the event chains associated, we'll go through and quickly triage potential key findings. We pull out gems like "there sure are a lot of colors" and "those lines definitely seem long" to see if they are indeed relevant or statistically significant. In this case, the lines are indeed long, indicating that a lot of the attacks within this pattern involve a variety of different actions done by actors until they finally achieve their goal. Only the Social Engineering pattern has a similar number of steps

involved in both data breaches and incidents. In terms of colors, this pattern has a good combination of mostly Malware events, with some Hacking and a very small smattering of other Action types as a garnish.

Figure 82 describes this differently, and shows Malware being involved in over 70% of the cases and Hacking in over 40%. Lastly, at a very high level, we can tell that the vast majority of the incidents in this pattern are from Financially motivated External actors. The further we dig, the more interesting this pattern becomes.

When we did a deep dive into the data, we found that there are three main "components" that make up this pattern. The first is Ransomware, with 99% of the Ransomware cases falling into this one pattern. The second is Malware in general, and the third is Magecart attacks in which Web applications are compromised with a script to export data as it is processed. Let's go over them.

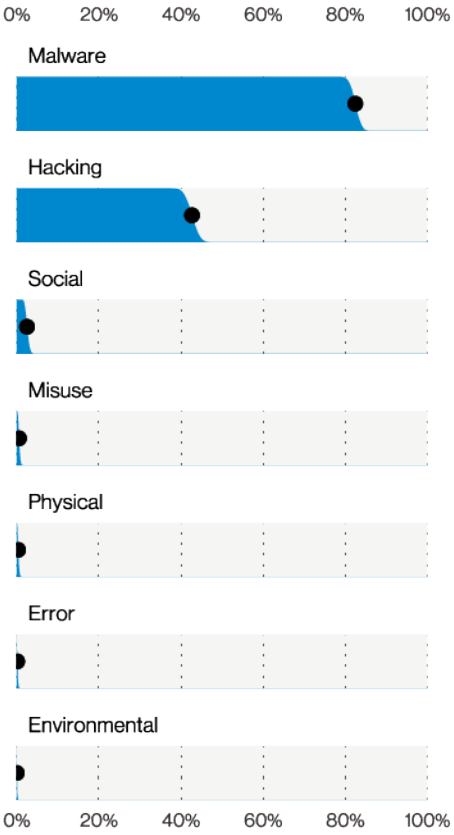



**Figure 82.** Actions in System Intrusion breaches (n=966)

# We're still writing about ransomware?

Unfortunately, this is a section that we've had to write consistently over the last few years and odds are that we'll probably continue to write about this in subsequent reports. This year, we're displeased to report that we've seen yet another increase in Ransomware cases, which has been continuing on an upward trend since 2016 and now accounts for 5% of our total incidents. The novel fact is that 10% of all breaches now involve Ransomware. This is because Actors have adopted the new tactic of stealing the data and publishing it instead of just encrypting it. These attacks have some variety in terms of how the Ransomware gets on the system, with Actors having strong preferences that can be broken into several vectors. The first vector is through the Use of stolen credentials or Brute force. We've seen 60% of the Ransomware cases involving direct install or installation through desktop sharing apps. The rest of the vectors that we saw were split between Email, Network propagation and Downloaded by other malware, which isn't surprising as we found in our web proxy detections dataset that 7.8% of organizations attempted to download at least one piece of known Ransomware last year (Figure 83). For these types of incidents and breaches, we largely see servers being targeted, which makes sense considering that's where the data is located.

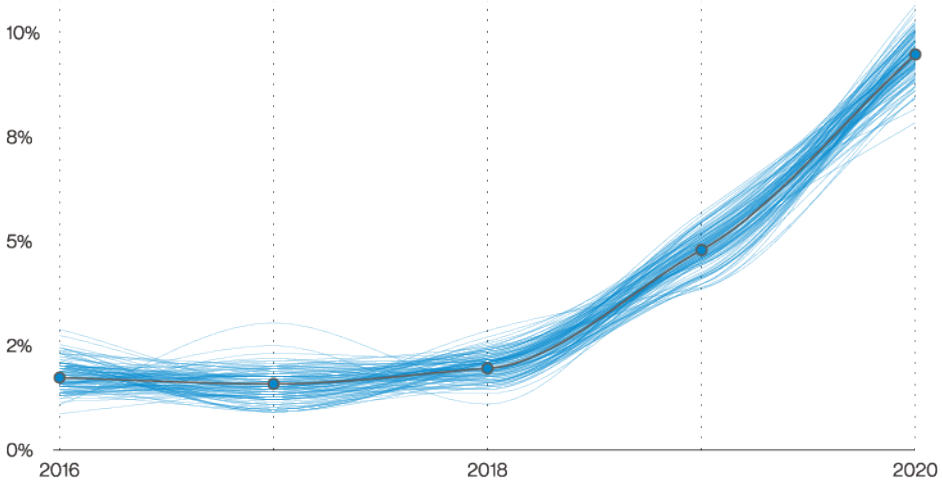



**Figure 83.** Ransomware in breaches over time

# Magecart attacks

The second attack type that we found in this pattern involved the targeting of Web applications processing Payment cards. Now before you interrupt us and ask "but DBIR team, isn't there a whole pattern dedicated to attacks against Web applications?" let us state that the incidents we discuss here are slightly different than those attacks based on a few key components. The biggest differentiator is the subsequent use of Malware to capture Payment card data. In the System Intrusion pattern, we found that of the web servers targeted in this pattern, 60% had malware installed to capture app data and 65% of incidents involved payment cards. These types of attacks follow the trends of attack that we in the biz[69] have been calling Magecart-style attacks based on their original targets. For those who aren't familiar with this attack archetype, attackers will exploit some vulnerability, then use stolen credentials or some other means to access the code of an e-commerce website that processes credit card data. By using that access to the code base or server, they will insert additional code that will ship off the payment data not only to the correct endpoint, but also to their own servers, thereby quietly siphoning off valuable data.

**30% of the malware was directly installed by the actor, 23% was sent there by email and 20% was dropped from a web application. While this probably doesn't surprise many people, it does highlight the importance of having a robust defense to cover these three major entry paths for Malware.**

# General malware

The final breakdown of this pattern involves the general use of Malware that is found on a system. In many of these situations, we may not necessarily know if that Malware would have been used to cause further damage down the road or if it was just there for the sake of being there, doing the kind of things Malware enjoys doing.[70] When we removed the Ransomware cases, we found that 40% of the Malware cases we had left involved the use of C2/Trojans/Downloaders. There was also an interesting split in terms of how the Malware arrived on the system. We found 30% of the malware was directly installed by the actor, 23% was sent there by email and 20% was dropped from a web application. While this probably doesn't surprise many people, it does highlight the importance of having a robust defense to cover these three major entry paths for Malware.

When it comes down to the daily amount of malware incidents, Figure 84 shows that for the majority of organizations, this data has a whole lot of spikiness, which means some days it's probably relatively quiet—until it's not.

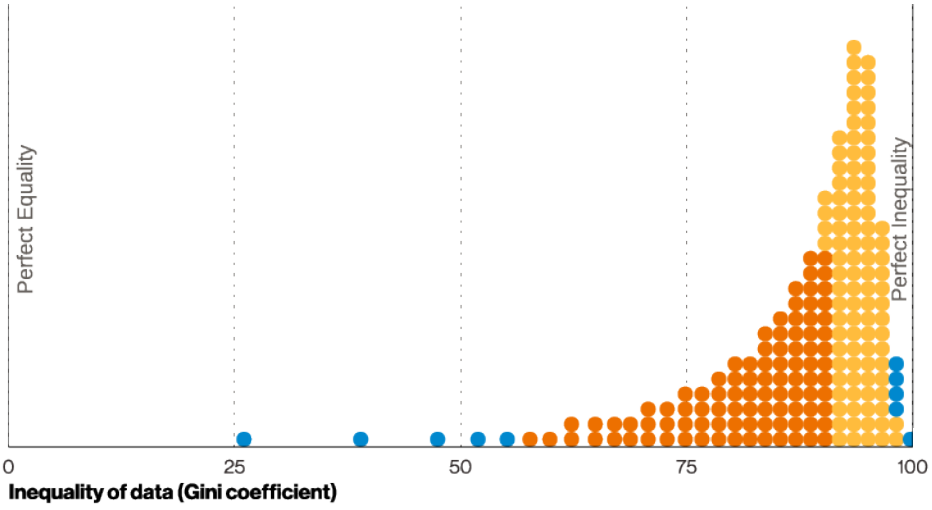



**Figure 84.** Inequality of Malware per day (n=16,524)
Each dot represents 0.5% of organizations

69 There is no biz like Cyberbiz.
70 Even Malware wants to live its best life.

While we don't necessarily know the severity of these malware events, we do know that data from botnet incidents we reviewed indicates that the majority of botnet infections only compromised three or fewer credentials. So, having malware in your environment, if properly cleaned and handled, probably isn't the end of the world, but it's best to not let it fester.

# The big picture shifts.

In the last few iterations of this report, we have mentioned the decrease in the targeting of Payment data. We have continued to see this trend in this pattern. As Figure 85 demonstrates, attackers are less likely to purely target Payment data and are more likely to broadly target any data that will impact the victim organization's operations. This will increase the likelihood that the organization will pay up in a Ransomware incident. As we have often repeated, the monetization through Ransomware seems to have become the preferred method, and the targeting of data will shift to reflect that. The attacks that come out of this pattern impact all of the industries we track at some level, which shows the wide net that these Actors cast to turn a profit.

**Attackers are less likely to purely target Payment data and are more likely to broadly target any data that will impact the victim organization's operations. This will increase the likelihood that the organization will pay up in a Ransomware incident.**

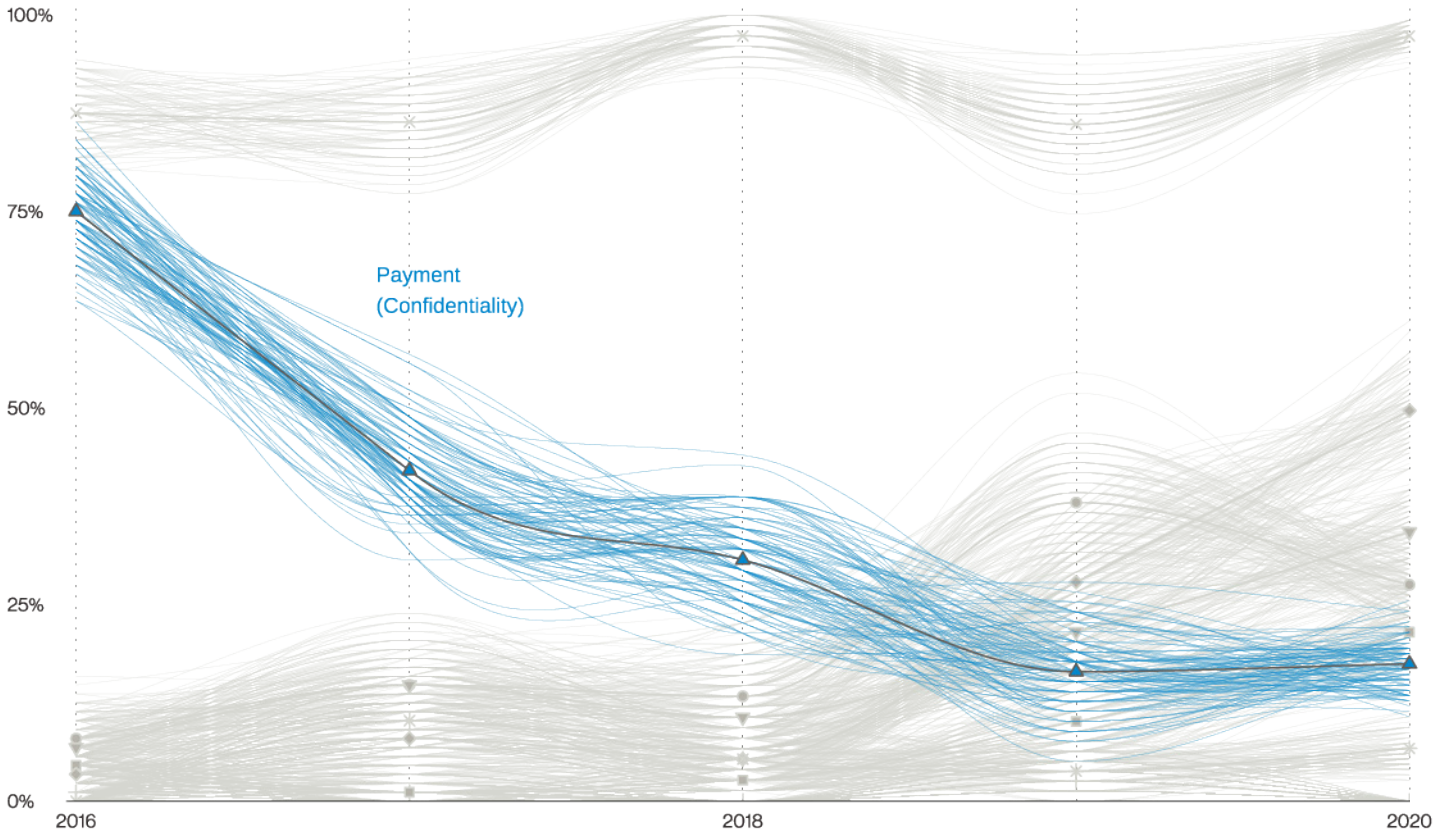



**Figure 85.** Attribute varieties in breaches over time

# Basic Web Application Attacks

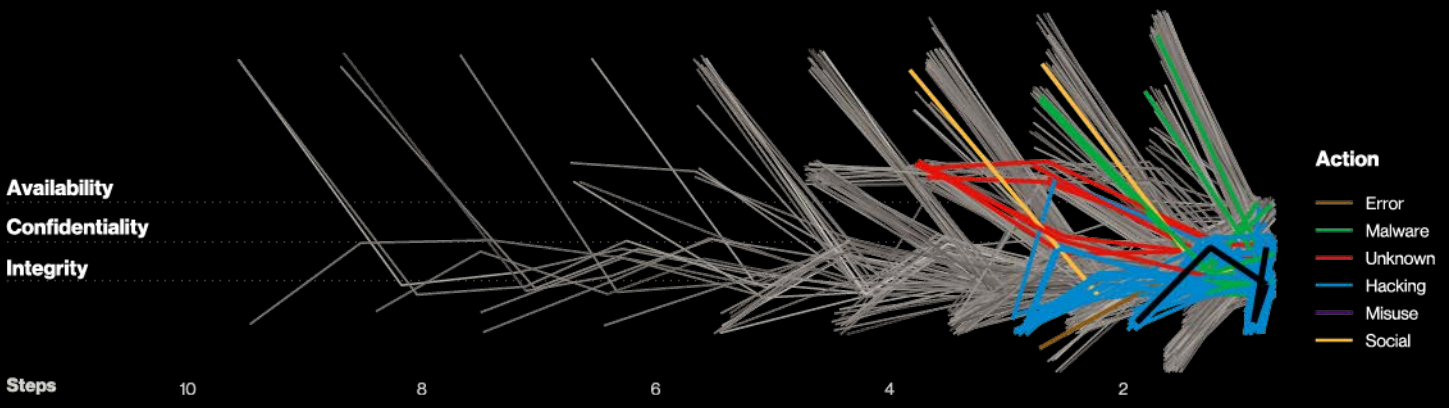



**Figure 86.** Basic Web Application Attacks incident paths (n=130)

## Summary

Basic Web Application Attacks are those with a small number of steps or additional actions after the initial Web application compromise. They are very focused on direct objectives, which range from getting access to email and web application data to repurposing the web app for malware distribution, defacement or future DDoS attacks.

| | |
|---|---|
| **Frequency** | 4,862 incidents, 1,384 with confirmed data disclosure |
| **Threat Actors** | External (100%), Internal (1%), Multiple (1%) (breaches) |
| **Actor Motives** | Financial (89%), Espionage (7%), Grudge (2%), Fun (1%) (breaches) |
| **Data Compromised** | Credentials (80%), Personal (53%), Other (25%), Internal (12%) (breaches) |

Basic Web Application Attacks (or BWAA)—we wanted BWAHA but we couldn't justify the H—is the new and improved version of our trusty Web Applications pattern. We do realize the name is a mouthful, but it better captures the nature of these short and to-the-point attacks that target open web and web-adjacent interfaces (it also freshens breath and whitens teeth). Our other name option was almost as long: Simple Web Attack Group (or SWAG), and perhaps that would have been better, since those attacks are looking for some low-hanging, easily available knickknacks to grab.

While the Assets present in this pattern according to Figure 88 are overwhelmingly represented by the Hacking of Servers, there are a few different sub-patterns encapsulated here, and they are all easy to explain and visualize.

The first sub-pattern covers the Use of stolen credentials and Brute force through a Web application vector to compromise either actual Web apps or Mail servers, as you can see on Figure 86. Almost all (96%) of those Mail servers compromised were cloud-based, resulting in the compromise of Personal, Internal or Medical data.

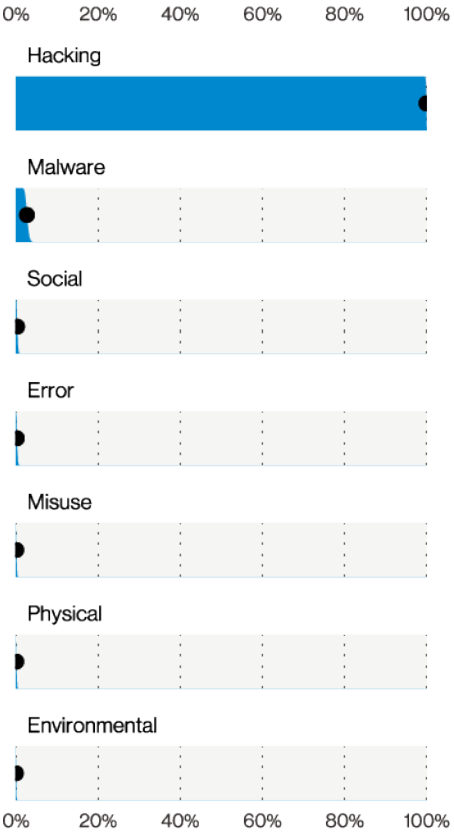



**Figure 87.** Actions in Basic Web Application Attacks breaches (n=1,384)

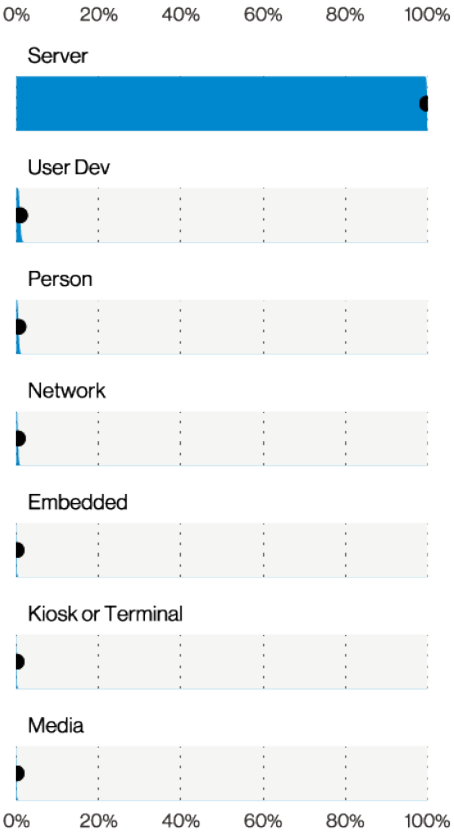



**Figure 88.** Assets in Basic Web Application Attacks breaches (n=1,369)

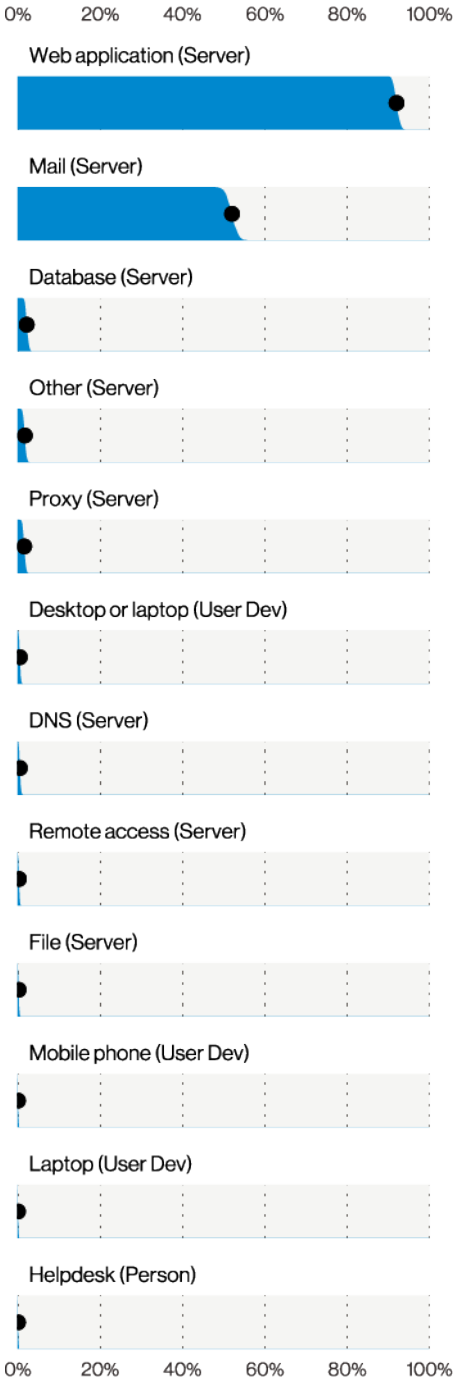



**Figure 89.** Asset varieties in Basic Web Application Attacks breaches (n=1,324)

## All of those Brute force attempts do not happen all at the same time, or even with any predictable regularity.

Astute readers will point out that if using stolen credentials is the leading characteristic of this part of BWAA, how is it differentiated from other threat actor favorites such as Social Engineering and System Intrusion? Glad you asked! It turns out that the credential abuse actions in this pattern were not preceded by any kind of Social attacks as far as the victims were aware. This could mean that either they didn't notice it, or that they were victims of a credential stuffing attack, where the credentials were actually compromised elsewhere and were, sadly, the same on the affected system.

Brute force and credential stuffing attacks are extremely prevalent according to SIEM data analyzed in our dataset. We found that 23% of the organizations monitored had security events related to those types of attacks, with 95% of them getting between 637 and 3.3 billion(!) attempts against them, as Figure 90 demonstrates. This is a very large number at face value, but when you consider the sheer volume of automated bots and worms looking for vulnerable services out there, it feels par for the course.

However, as you may suspect if you have been reading up on the other patterns, all of those Brute force attempts do not happen all at the same time, or even with any predictable regularity. Figure 91 demonstrates that more often than not for the organizations we reviewed, those attacks happened in very uneven intervals. It seems the cost of keeping up with potential credential dumps can't be simplified as something you should do every month or so.

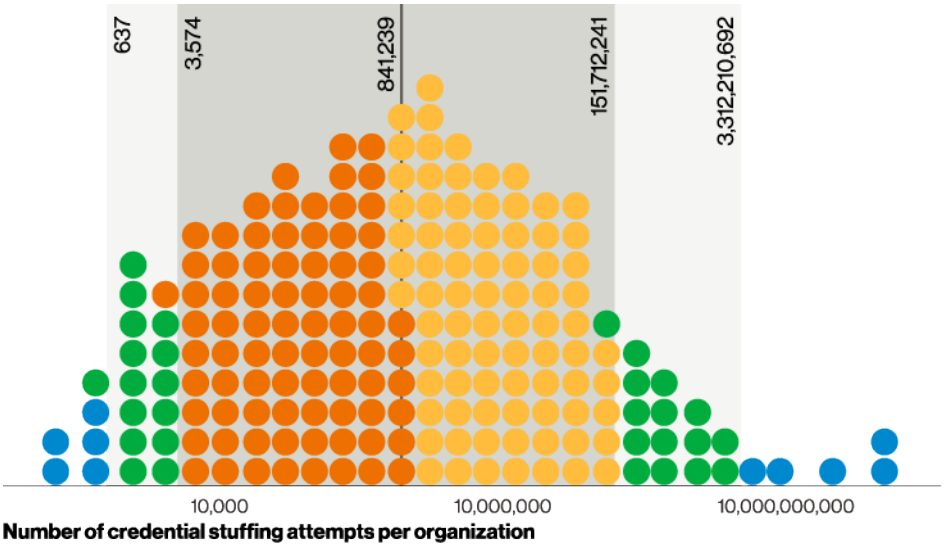



**Figure 90.** Credential stuffing attempts per organization (n=821)
Each dot represents 0.5% of organizations.

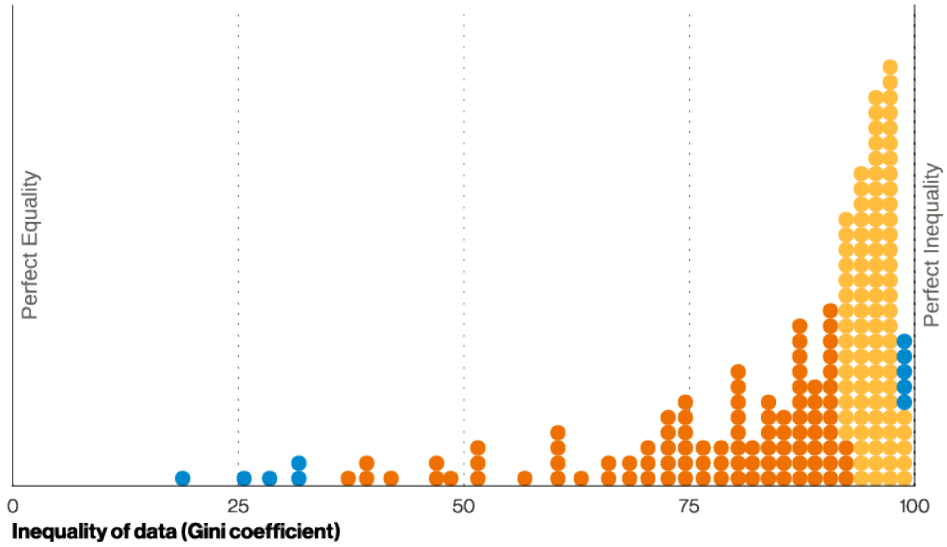



**Figure 91.** Inequality of login attempts per day (n=328)
Each dot represents 0.5% of organizations

The other sub-pattern covers the exploitation of vulnerabilities in Web applications. They are not as common as the credential-related ones, as Figure 92 shows, but they are significant. Vulnerability exploitation is also the territory of a sister pattern, System Intrusion, but those present here in BWAA are not only focused on Web applications. They are also attacking with a small number of steps or additional actions after the initial Web application compromise.

In those incidents, the Actor will be focused on repurposing the web app for malware distribution, defacement[71] or installing malware for future DDoS attacks and calling it a day. Needless to say, a lot of the motive here is Secondary, more precisely in 78% of incidents. Threat actors are clearly not wasting the opportunity to shout "It's free real estate!" and expand their nefarious domains. Figure 93 shows this distribution in incidents, as in defacement, cases we often cannot get confirmation of a fully realized breach.

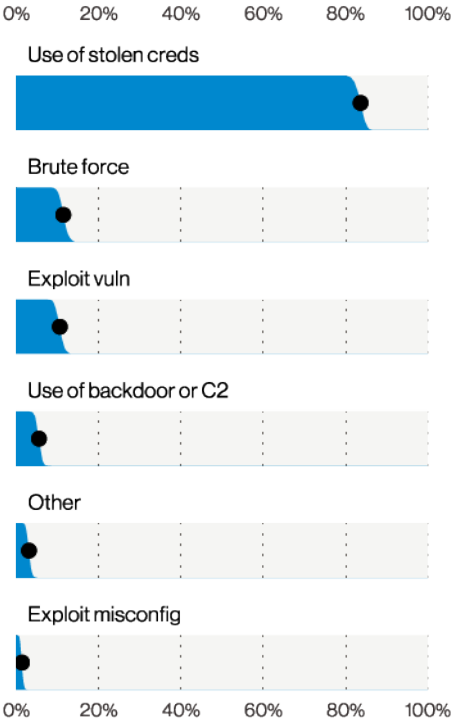



**Figure 92.** Top Hacking varieties in Basic Web Application Attacks incidents (n=947)

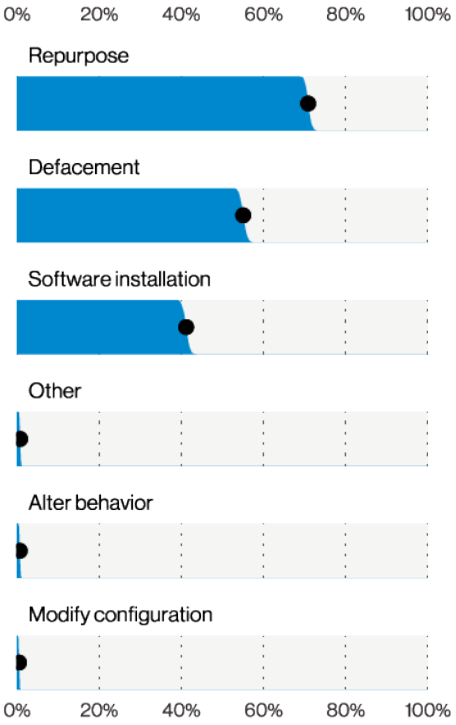



**Figure 93.** Top Integrity varieties in Basic Web Application Attacks breaches (n=3,653)

71 It's the '90s! Join our DBIR webring in Geocities!

# Everything Else


**Figure 94.** Everything Else incident paths (n=3)

## Summary

This pattern was recalibrated and now consists primarily of Physical tampering cases, in addition to three shiny new Environmental cases, which still have that new incident smell. It does not feature prominently in any of the industries this year and has been relegated to the "stuff leftover that didn't fit in anywhere else" status it formerly occupied prior to the astronomical rise of Social Engineering.

| | |
|---|---|
| **Frequency** | 129 incidents, 38 with confirmed data disclosure |
| **Threat Actors** | External (95%), Internal (5%) (breaches) |
| **Actor Motives** | Financial (100%) (breaches) |
| **Data Compromised** | Payment (61%-96%) (breaches) |

The fairway plot (Figure 94) provides a good illustration of the two main types of incidents that ended up in the Everything Else pattern. As you may recall from last year, this pattern was quite popular and could be found in the top three patterns in several industries. It was clearly time for us to recalibrate when our catch-all bucket was full to overflowing with incidents that didn't fit the other patterns.

Now that we've sifted through the data and completed our recalibration (which is covered at length in the Introduction to Patterns section), there are still a few incidents and breaches that fit into the Everything Else pattern. They are Physical tampering cases (think ATM and gas pump skimmers) and the so-rare-we-are-excited-to-be-able-to-talk-about-it-FINALLY Environmental cases.

Yes, you read that correctly, we actually had three cases from the Environmental action that made it into the dataset this year. It does our geeky VERIS hearts proud to finally be able to talk about them. We considered creating "Ask me about my Environmental breaches" bumper stickers, but bumper stickers are bad for the environment.

We used to have (back in the murky depths of antiquity) an entire pattern devoted to Payment Card Skimmers, but they have been decreasing dramatically in our dataset over the years. This year we saw an even sharper drop-off than ever before. There were only 20 skimming incidents (all confirmed breaches) in the dataset this year. We attribute this decrease, at least in part, to the travel restrictions related to COVID-19.

In prior years, particularly in the public dataset (VCDB),[72] we saw evidence of skimming groups from abroad coming into the U.S. and installing skimming devices on their infrastructure of choice (some favor ATMs, some focus on Gas terminals). In fact, one could almost plot their progress along the major routes before they would presumably return to their place of origin along with their stolen data. Given the travel restrictions that began in March 2020, the freedom to carry out this type of concentrated raid has significantly diminished. And while it is possible that this kind of breach is no longer being tracked at the national level, we like to think there is at least one positive outcome from what has been a very difficult year for most of the world.

Now, on to our Environmental breaches. As mentioned, we only have three of them, which is admittedly a very small number. However, they are separate and distinct events. We saw incidents that arose from one fire, one hurricane and one tornado (Table 3). All three affected paper documents strewn to the winds (in the classic Wizard of Oz fashion) from the violence of their encounters with the forces of nature. The actor in these cases is considered External of type Force majeure. We hope nature will now retire from the data breach stage and leave the loss of records to the normally scheduled actors.

| # breaches | Environmental variety |
|---|---|
| 1 | Fire |
| 1 | Hurricane |
| 1 | Tornado |

**Table 3.** Environmental breaches

72 https://github.com/vz-risk/VCDB

04

# Industries

# Introduction to industries

This year we looked at 29,207 incidents, which boiled down to 5,258 confirmed data breaches (Table 4). Once again, we break these incidents and breaches into their respective industries to illustrate that all industries are not created equal in terms of attack surfaces and threats. The kinds of attacks suffered by a particular industry will have a lot to do with what kind of infrastructure they rely on, what kind of data they handle, and how people (customers, employees and everyone else) interact with them.

A large organization whose business model focuses entirely on mobile devices, where customers use an app on their phone, will have different risks than a small mom and pop shop with no internet presence, but who uses a Point of Sale vendor to manage their systems. The infrastructure, and conversely the attack surface, largely drives the risk.

While keeping that in mind, we caution our readers not to make inferences about the security posture (or lack thereof) of a particular sector based on how many breaches or incidents that industry reports. These numbers are heavily influenced by several factors, including data breach reporting laws and partner visibility. Because of this, some of the industries have very low numbers, and as with any small sample, we must caution you that our confidence in any statistics derived from that small number must also be less.

As in past years, we have broken down the breaches and incidents by industry in a heat map that categorizes the data into Patterns, Actions and Assets (Figures 95 and 96 respectively). These figures help to answer the "so what?" question in our data, and are useful as indications of what attack patterns an organization is most likely to encounter, given their industry. This, paired with the CIS Controls in each industry section, can be a guide for determining how best to mitigate risk.

| Incidents | Total | Small (1-1,000) | Large (1,000+) | Unknown | Breaches | Total | Small (1-1,000) | Large (1,000+) | Unknown |
|---|---|---|---|---|---|---|---|---|---|
| Total | 29,207 | 1,037 | 819 | 27,351 | | 5,258 | 263 | 307 | 4,688 |
| Accommodation (72) | 69 | 4 | 7 | 58 | | 40 | 4 | 7 | 29 |
| Administrative (56) | 353 | 8 | 10 | 335 | | 19 | 6 | 7 | 6 |
| Agriculture (11) | 31 | 1 | 0 | 30 | | 16 | 1 | 0 | 15 |
| Construction (23) | 57 | 3 | 3 | 51 | | 30 | 3 | 2 | 25 |
| Education (61) | 1,332 | 22 | 19 | 1,291 | | 344 | 17 | 13 | 314 |
| Entertainment (71) | 7,065 | 6 | 1 | 7,058 | | 109 | 6 | 1 | 102 |
| Finance (52) | 721 | 32 | 34 | 655 | | 467 | 26 | 14 | 427 |
| Healthcare (62) | 655 | 45 | 31 | 579 | | 472 | 32 | 19 | 421 |
| Information (51) | 2,935 | 44 | 27 | 2,864 | | 381 | 35 | 21 | 325 |
| Management (55) | 8 | 0 | 0 | 8 | | 1 | 0 | 0 | 1 |
| Manufacturing (31-33) | 585 | 20 | 35 | 530 | | 270 | 13 | 27 | 230 |
| Mining (21) | 498 | 3 | 5 | 490 | | 335 | 2 | 3 | 330 |
| Other Services (81) | 194 | 3 | 2 | 189 | | 67 | 3 | 0 | 64 |
| Professional (54) | 1,892 | 793 | 516 | 583 | | 630 | 76 | 121 | 433 |
| Public (92) | 3,236 | 22 | 65 | 3,149 | | 885 | 13 | 30 | 842 |
| Real Estate (53) | 100 | 5 | 3 | 92 | | 44 | 5 | 3 | 36 |
| Retail (44-45) | 725 | 12 | 27 | 686 | | 165 | 10 | 19 | 136 |
| Wholesale Trade (42) | 80 | 4 | 10 | 66 | | 28 | 4 | 7 | 17 |
| Transportation (48-49) | 212 | 4 | 17 | 191 | | 67 | 3 | 8 | 56 |
| Utilities (22) | 48 | 1 | 2 | 45 | | 20 | 1 | 2 | 17 |
| Unknown | 8,411 | 5 | 5 | 8,401 | | 868 | 3 | 3 | 862 |
| Total | 29,207 | 1,037 | 819 | 27,351 | | 5,258 | 263 | 307 | 4,688 |

**Table 4.** Number of security incidents and breaches by victim industry and organization size

## Breaches

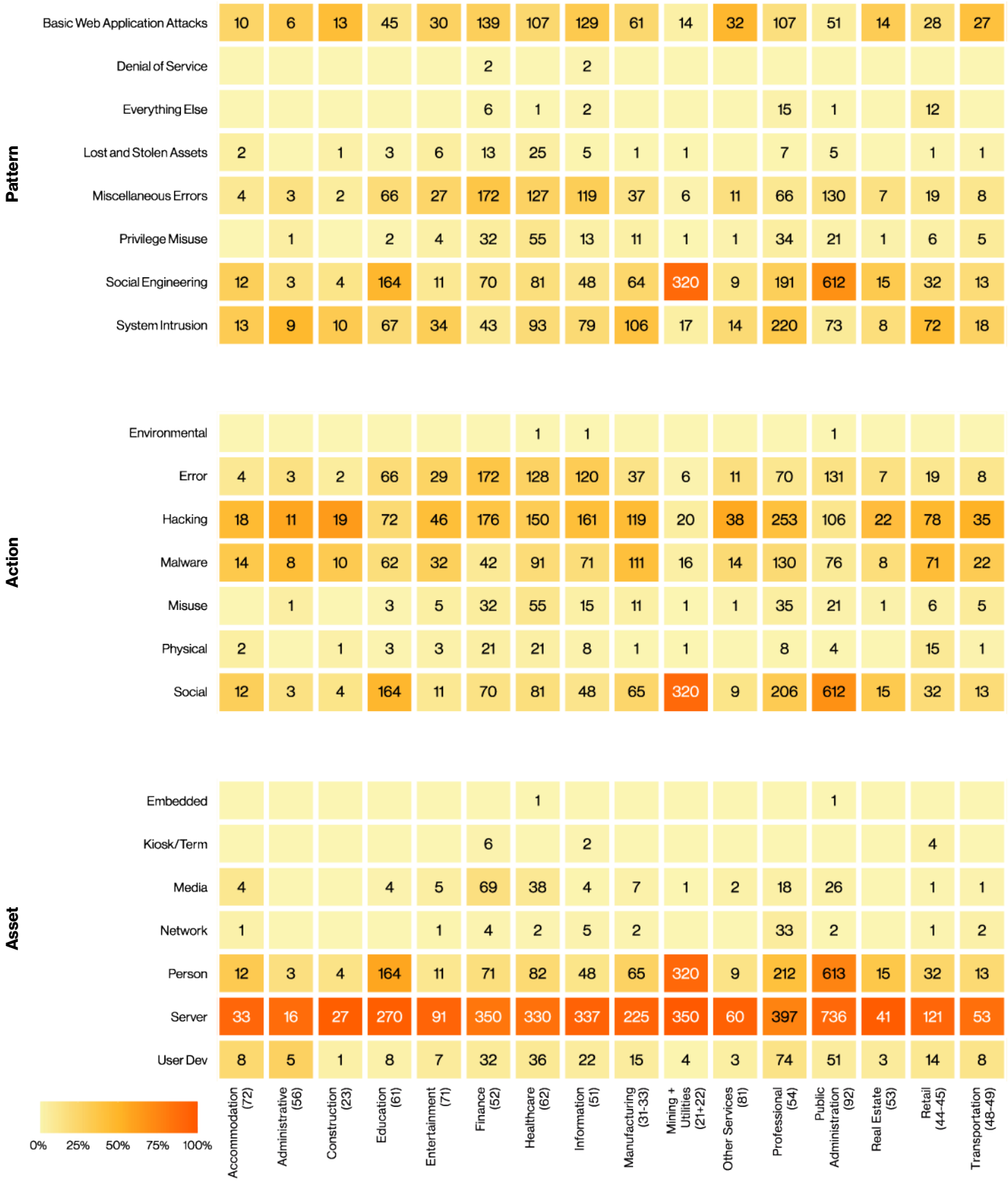



**Figure 95.** Breaches by industry

## Incidents

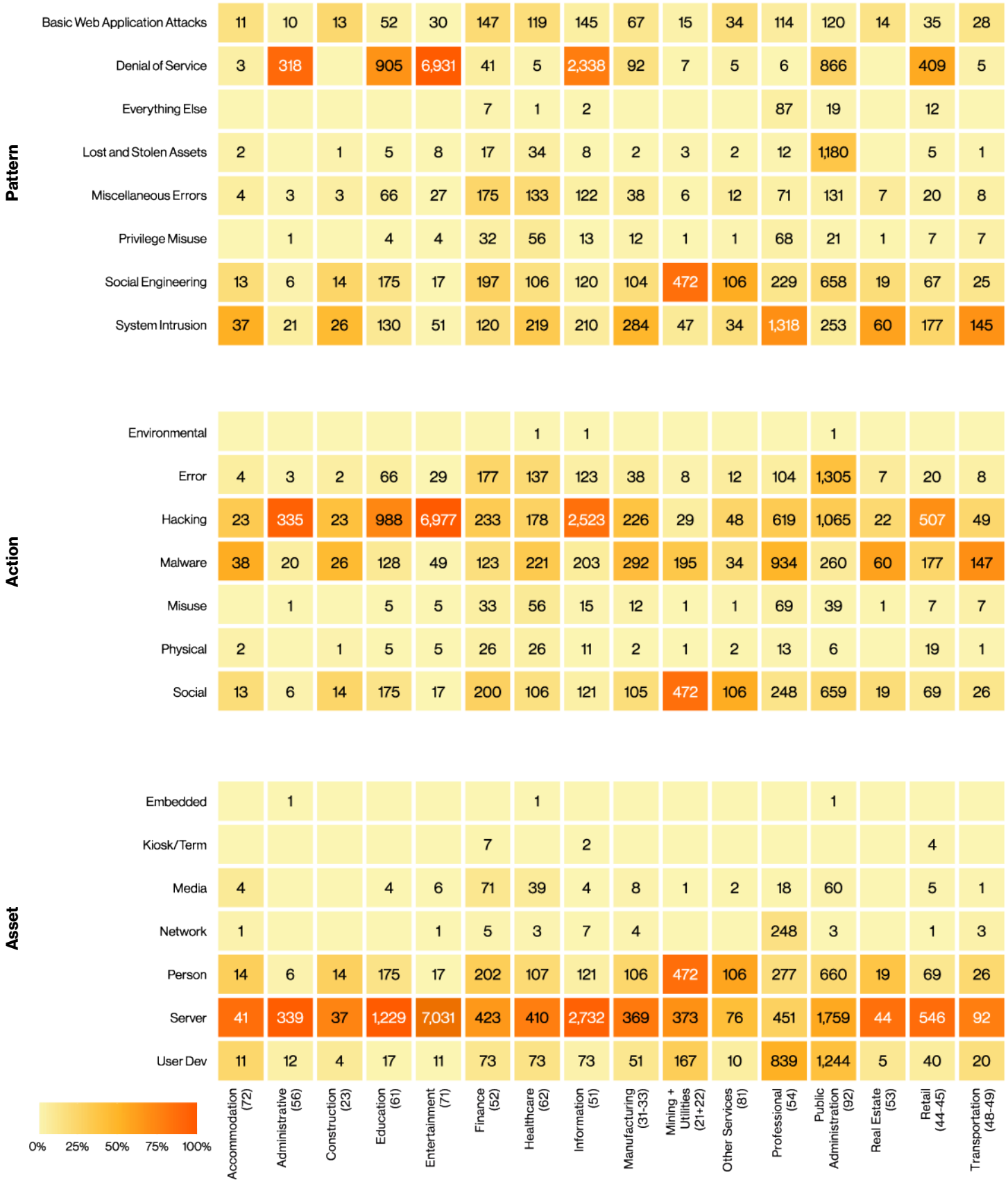



**Figure 96.** Incidents by industry

When discussing the industries with a small sample, we will provide ranges within which the actual value may reside. This allows us to maintain our confidence interval while still providing you with an idea of what the actual number might be, had we been given a large enough sample. For example, instead of saying "In the Accommodation industry, 92% of attacks were Financially motivated," we show that Financially motivated attacks ranged between 86 and 100%. Check out our riveting Methodology section for more information about the statistical confidence background used throughout this report.

It is worth noting that some of the industry sections this year may look smaller than usual. This is because we did not want to steal the thunder from the deep-dive analysis we did on the new Patterns. If you are just here for a glimpse of your industry,[73] our recommendation is to verify what the Top Patterns are in the At-a-Glance table accompanying each industry and then spend some time with those pattern sections.

We also provide a description of which CIS Controls from Implementation Group 1 (IG1) to prioritize in each industry section for ease of reading in case you want to get straight to strategizing your security moves.

**Check out our riveting Methodology section for more information about the statistical confidence background used throughout this report.**

73 We can't blame you. Sometimes we eat the dessert first, too.

# Accommodation and Food Services NAICS 72

## Summary

The Accommodation and Food Services industry is experiencing Hacking, Social and Malware attacks with close to equal frequency.

| | |
|---|---|
| **Frequency** | 69 incidents, 40 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Social Engineering and Basic Web Application Attacks represent 85% of breaches |
| **Threat Actors** | External (90%), Internal (10%) (breaches) |
| **Actor Motives** | Financial (86%-100%), Espionage (0%-14%) (breaches) |
| **Data Compromised** | Personal (51%), Credentials (49%), Payment (33%), Other (15%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Access Control Management (6), Secure Configuration of Enterprise Assets and Software (4) |

The Accommodation and Food Services Industry (NAICS 72) shows fewer breaches this year than in the past (92 last year). A logical explanation for this would be that due to the global conditions during the greater part of 2020, travel and dining out were significantly curtailed. That would result in fewer transactions, and by extension, less breaches. Nevertheless, 40 incidents are a statistically sufficient number for us to derive some conclusions. The most prevalent patterns in this industry were System Intrusion, Social Engineering and Basic Web Application Attacks, although there was almost nothing to tell them apart (Figure 97).

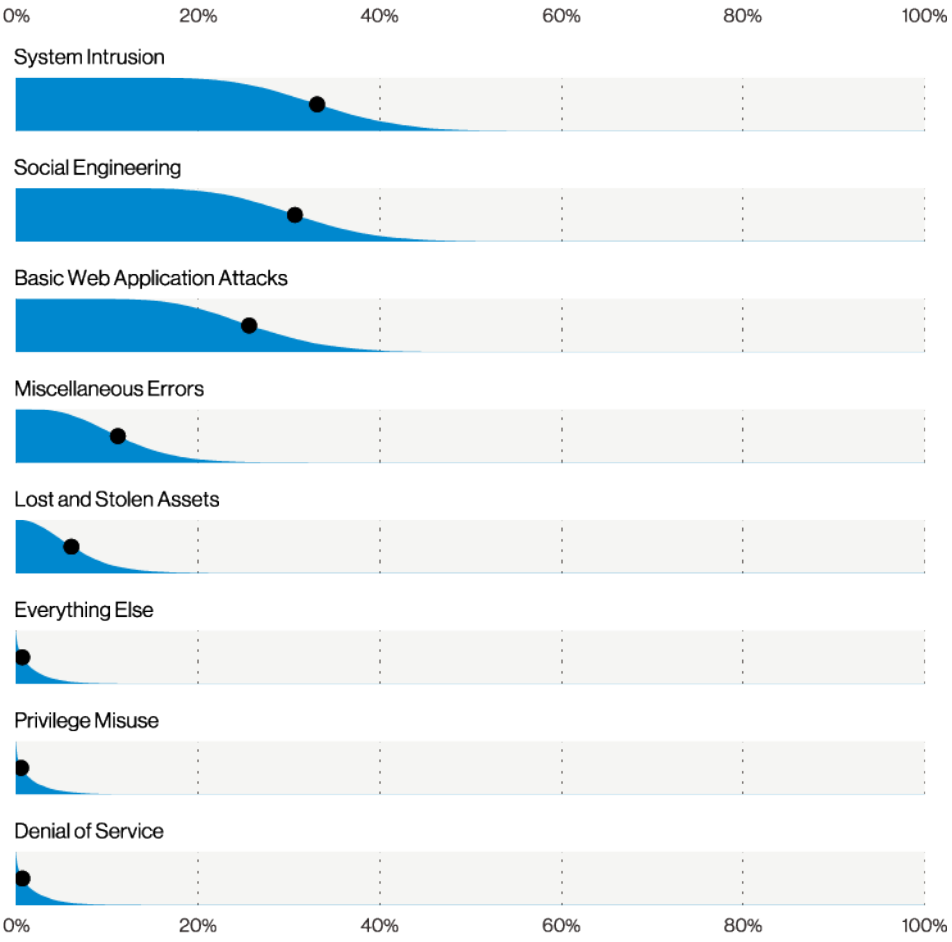



**Figure 97.** Patterns in Accommodation and Food Services breaches (n=40)

As pointed out elsewhere in this report, certain Action types have been clustered together to form the System Intrusion pattern. This includes Malware actions that would have previously been found in the Crimeware pattern. However, while the patterns may have changed, as you can see in Figure 98, the malware prevalent in this industry is of the Backdoor, C2 and Trojan varieties that we have witnessed in previous years.

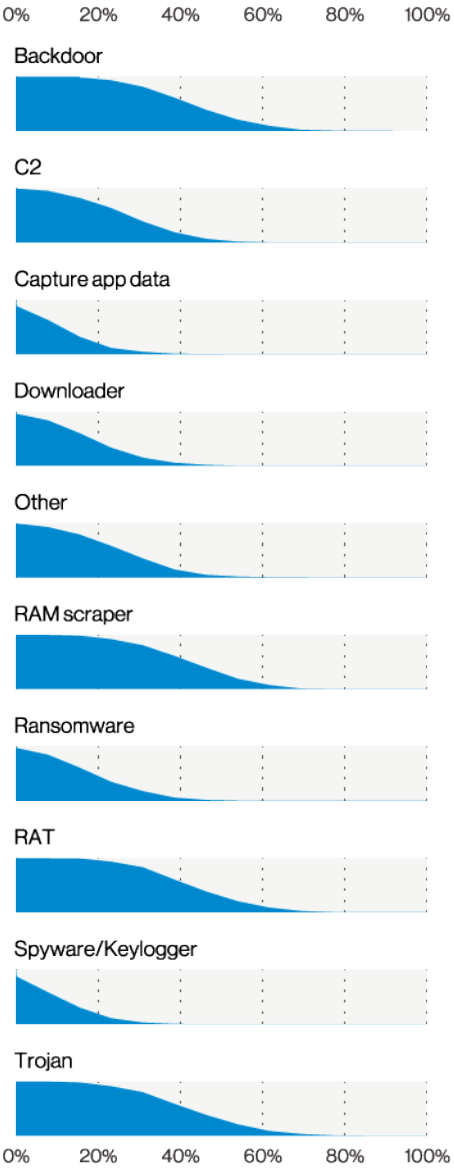



**Figure 98.** Top Malware varieties in Accommodation and Food Services breaches (n=13)

Direct installation by the attacker is by far the most common vector for the malware seen in this vertical.

With regard to data type, Credentials (49%), Personal (51%) and Payment (33%) all come in at or near the same number, and are again what one might expect as a result of the attack types mentioned above. Finally, while we must admit that our sample size is very small (n=18), the Discovery method, when known, is (as it has been for many years) via a third party, 39%-75%. Usually via notification by law enforcement or from a Common Point of Purchase audit, but in some cases by the threat actors themselves. We would love to see some positive change in Discovery methods for this industry, as it only stands to reason that the impact of a breach will likely be greater if you have to wait for someone outside of your organization to inform you.

# Arts, Entertainment and Recreation NAICS 71

## Summary

The Use of stolen credentials, Phishing and Ransomware continue to play big roles in this industry. Compromised Medical information was seen at an unexpectedly high level as well.

| | |
|---|---|
| **Frequency** | 7,065 incidents, 109 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Basic Web Application Attacks and Miscellaneous Errors represent 83% of breaches |
| **Threat Actors** | External (70%), Internal (31%), Multiple (1%) (breaches) |
| **Actor Motives** | Financial (100%) (breaches) |
| **Data Compromised** | Personal (83%), Credentials (32%), Medical (26%), Other (18%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Secure Configuration of Enterprise Assets and Software (4), Access Control Management (6) |

While the way in which we consumed entertainment changed this year, hopefully temporarily, attackers continued to follow the same winning combination that they've been using for the last few years in this industry. Namely, targeting web applications and utilizing malware to its fullest extent. And of course, there was the occasional human blunder that serves to keep life interesting.

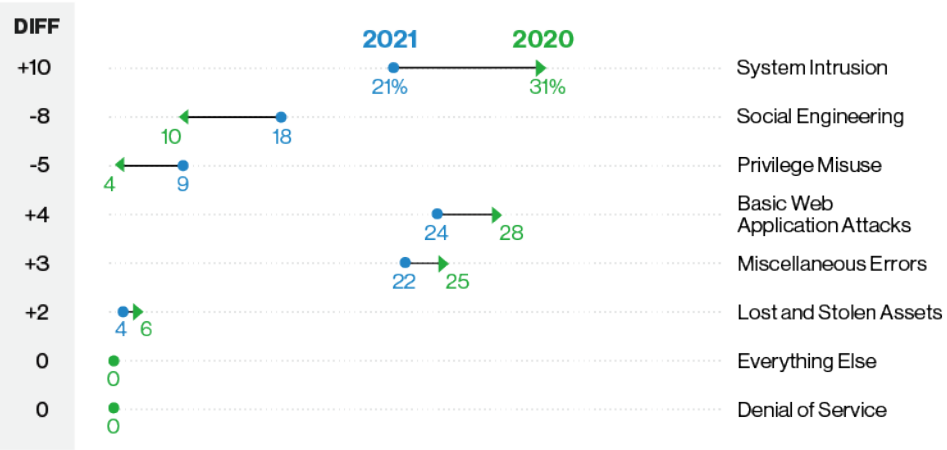



**Figure 99.** Patterns in Arts and Entertainment breaches

System Intrusion, Web Applications and Errors are more or less tied for the top ranking. Their combined weight accounts for 83% of the breaches in this sector. This is in line with the trend set in previous years, and what we saw in last year's report (Figure 99). With that in mind, it is perhaps only to be expected that action types such as the Use of stolen credentials, Ransomware, Phishing and Misconfiguration were responsible for most breaches (Figure 100).

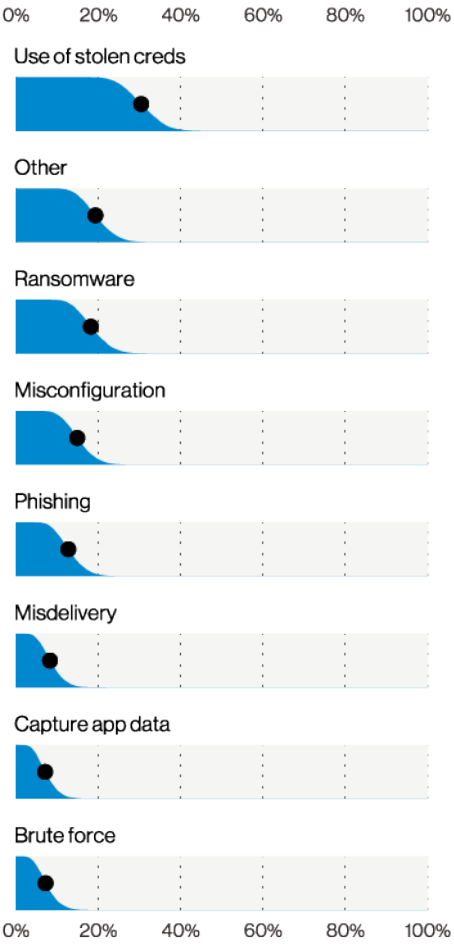



**Figure 100.** Top Actions in Arts and Entertainment breaches (n=90)

What was a bit surprising was the high level of Medical information breached in this sector. One would typically associate medical record loss with the Healthcare industry. However, upon digging into the data a bit more, the Personal Health Information (PHI) was related to athletic programs, which fall under this vertical. It is possible the medical nature of the data was unclear, and so the worst case (medical rather than just personal) data was reported. Still, this reveals an important lesson: Don't assume that because your organization is not in the medical field that you don't possess medical data (or that you don't have a duty to ensure that it is protected appropriately).

From an incident point of view, DDoS attacks were once again quite high this year. This is potentially due to the gambling websites that also reside in this sector. Therefore, if you are operating an online gambling platform the safe bet is to plan for DDoS, because the house always needs to win.

# Educational Services NAICS 61

## Summary

The Education vertical has an unusually large percentage of Social Engineering attacks in which Pretexting is the variety. These are typically with a view toward instigating a fraudulent transfer of funds. Miscellaneous Errors and System Intrusion are both still enrolled as well, and are taking a full load.

| | |
|---|---|
| **Frequency** | 1,332 incidents, 344 with confirmed data disclosure |
| **Top Patterns** | Social Engineering, Miscellaneous Errors and System Intrusion represent 86% of breaches |
| **Threat Actors** | External (80%), Internal (20%), Multiple (1%) (breaches) |
| **Actor Motives** | Financial (96%), Espionage (3%), Fun (1%), Convenience (1%), Grudge (1%) (breaches) |
| **Data Compromised** | Personal (61%), Credentials (51%), Other (12%), Medical (7%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Access Control Management (6), Secure Configuration of Enterprise Assets and Software (4) |

The Education sector has certainly had a challenging year, with the pandemic mandating that classes be held online, in a hybrid form and sometimes not at all. With those challenges comes opportunity—mostly for criminals. This sector is assailed by Financially motivated actors looking to gain access to the data and systems of the people who are just trying to get through the school day.

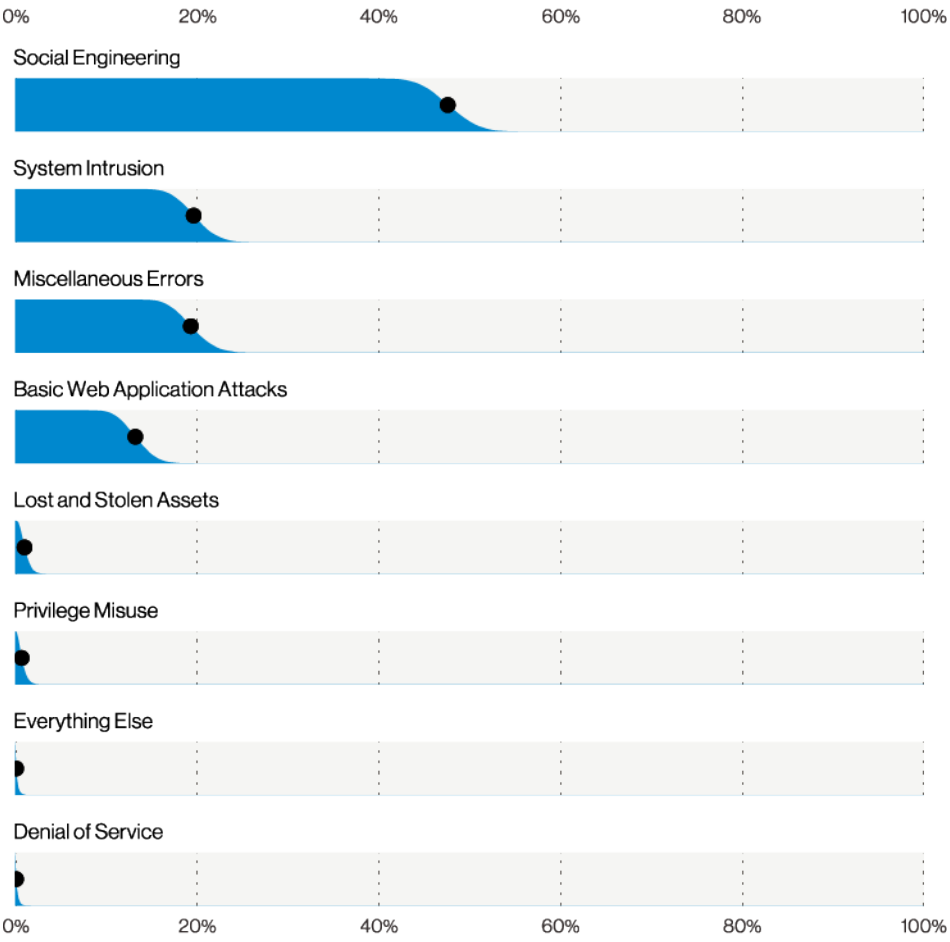



**Figure 101.** Patterns in Education breaches (n=344)

One of the top patterns in this industry is Social Engineering (Figure 101), and in looking at these cases, we find a larger than usual amount of Pretexting. Frequently, Social Engineering aficionados will craft a simple phishing email and wait for their victims to reach out to them. In the Education sector, they seem to be harkening back to their creative writing courses, and are putting forth the effort to invent a convincing scenario to get their victim to respond (Figure 102).

Are they getting good grades for their efforts? Yes, they get an A for "appropriation" of funds that do not belong to them. Considering their continued success at causing money to be transferred to them, they have clearly mastered the art of believability in their prose.

It stands to reason that people with access to wire transfers and other kinds of payments should be targeted for special training to help combat this kind of attack. Other controls to prevent wire transfers to new bank accounts should also be put in place.

Miscellaneous Errors and System Intrusion were almost tied in their bid for second place in the patterns for this sector. We see Misconfiguration (largely of databases that are spun up without the benefit of access controls, open for the world to see because knowledge wants to be free, right?) as the most common variety (Figure 103).

The System Intrusion pattern tells a tale of two actions—namely Hacking and Malware. Credential attacks are the most common starting point, with the credentials frequently coming from the result of other breaches and/or credential re-use. The attacker moves on to installing malware once they have their foothold established. Ransomware is a favorite malware flavor, and we've seen some groups taking copies of the data prior to triggering the encryption and then using it as further pressure against the victim.

**Ransomware is a favorite malware flavor, and we've seen some groups taking copies of the data prior to triggering the encryption and then using it as further pressure against the victim.**

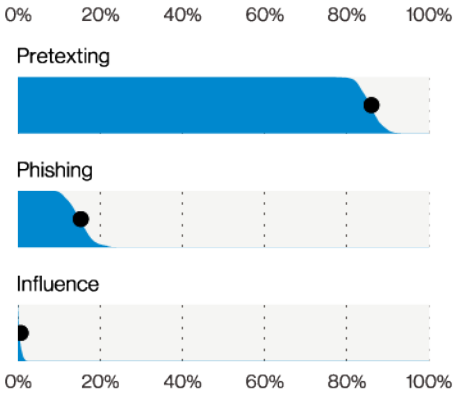



**Figure 102.** Social varieties in Education breaches (n=164)

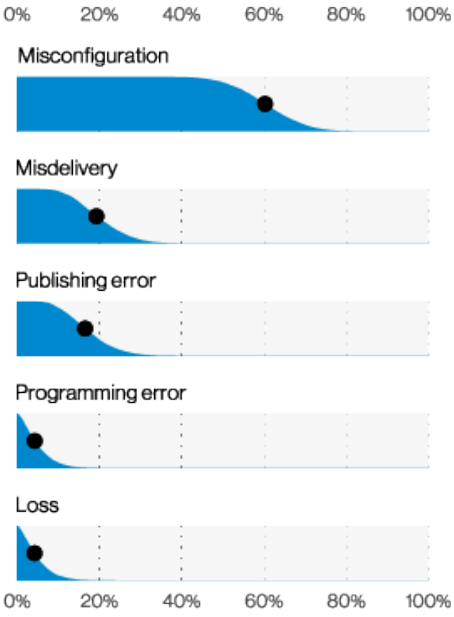



**Figure 103.** Error varieties in Education breaches (n=33)

# Financial and Insurance NAICS 52

## Summary

Misdelivery represents 55% of Financial sector errors. The Financial sector frequently faces Credential and Ransomware attacks from External actors.

| | |
|---|---|
| **Frequency** | 721 incidents, 467 with confirmed data disclosure |
| **Top Patterns** | Miscellaneous Errors, Basic Web Application Attacks and Social Engineering represent 81% of breaches |
| **Threat Actors** | External (56%), Internal (44%), Multiple (1%), Partner (1%) (breaches) |
| **Actor Motives** | Financial (96%), Espionage (3%), Grudge (2%), Fun (1%), Ideology (1%) (breaches) |
| **Data Compromised** | Personal (83%), Bank (33%), Credentials (32%), Other (21%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Secure Configuration of Enterprise Assets and Software (4), Access Control Management (6) |

The Financial Services industry has long been known for rapid changes, including sudden dips, dizzying highs and unforeseen fluctuations (thanks, Reddit users). This vertical has seen quite a diverse set of changes when it comes to the cybersecurity landscape as well. One that we have seen over the last few years has been a convergence of Internal actors and their associated actions with the more famous and nefarious External varieties.

This year, 44% of the breaches in this vertical were caused by Internal actors (having seen a slow but steady increase since 2017) (Figure 104). The majority of actions performed by these folks are the accidental ones, specifically the sending of emails to the wrong people, which represents a whopping 55% of all Error-based breaches (and 13% of all breaches for the year).

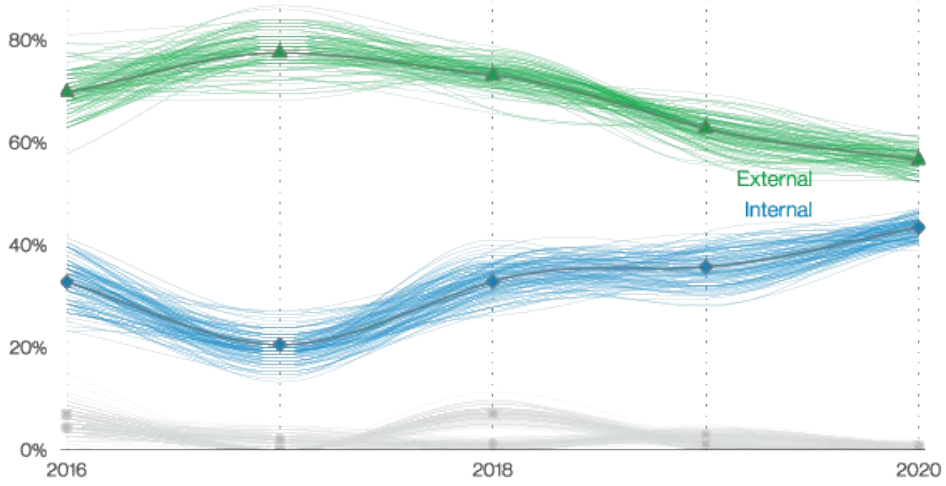



**Figure 104.** Actors in Finance breaches over time

When we turn our attention to malicious External actors, the Financial industry faces a similar onslaught of Credential attacks, Phishing and Ransomware attacks that we see topping the charts in other industries. With regard to data type, Personal comes in first, followed by Credentials and Bank data, hardly surprising given the focus of the industry.

Finally, this industry continues to be heavily reliant upon external parties for breach discovery. Typically via bad actors making themselves known (38% of the incidents) or notification from monitoring services (36% of incidents).

# Healthcare NAICS 62

## Summary

Basic human error continues to beset this industry as it has for the past several years. The most common Error continues to be Misdelivery (36%), whether electronic or of paper documents. Malicious Internal actions, however, have dropped from the top three for the second year in a row. Financially motivated organized criminal groups continue to target this sector, with the deployment of Ransomware being a favored tactic.

| | |
|---|---|
| **Frequency** | 655 incidents, 472 with confirmed data disclosure |
| **Top Patterns** | Miscellaneous Errors, Basic Web Application Attacks and System Intrusion represent 86% of breaches |
| **Threat Actors** | External (61%), Internal (39%) (breaches) |
| **Actor Motives** | Financial (91%), Fun (5%), Espionage (4%), Grudge (1%) (breaches) |
| **Data Compromised** | Personal (66%), Medical (55%), Credentials (32%), Other (20%), (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Secure Configuration of Enterprise Assets and Software (4), Access Control Management (6) |

Since 2019, the Healthcare sector has seen a shift from breaches caused by Internal actors to primarily External actors. This brings this vertical in line with the long-term trend seen by the other industries. This is good news actually, as no industry wants their employees to be their primary threat actor. While one of the top patterns for Healthcare continues to be Miscellaneous Errors, with Misdelivery being most common, at least errors are not malicious in nature (Figure 105). The insider breaches that were maliciously motivated have not shown up in the top three patterns in Healthcare for the past several years. But does this mean they are no longer occurring, or are they still around but we just aren't catching them (like Bigfoot)? Only time will tell.

For the second year in a row, we have seen Personal data compromised more often than Medical in this sector. That strikes us as strange, given the fact that this is the one sector where you would expect to see Medical information held most commonly. However, with the increase of External actor breaches, it may simply be that the data taken is more opportunistic in nature. If controls, for instance, are more stringent on Medical data, an attacker may only be able to access Personal data, which is still useful for financial fraud. Simply put, they may take what they can get and run.

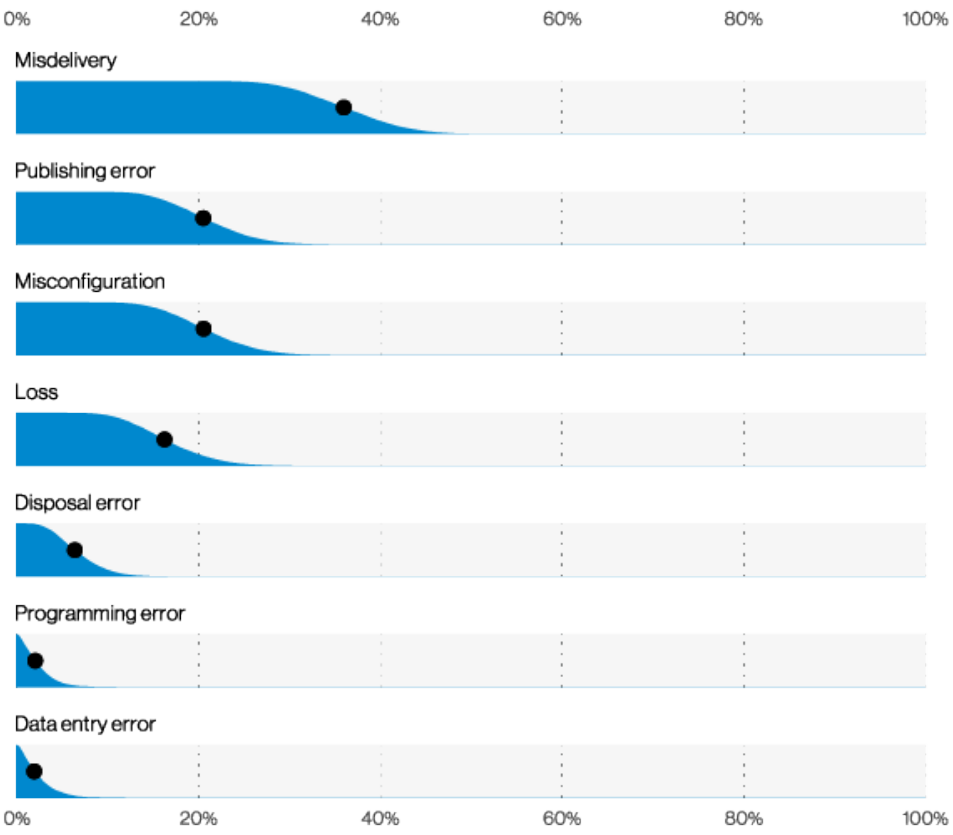



**Figure 105.** Error varieties in Healthcare breaches (n=70)

# Information NAICS 51

## Summary

This industry struggles with credential stealing botnets. Errors are also very common with Misconfiguration leading the way. From an incident perspective, DoS attacks accounted for the vast majority of attacks.

| | |
|---|---|
| **Frequency** | 2,935 incidents, 381 with confirmed data disclosure |
| **Top Patterns** | Basic Web Application Attacks, Miscellaneous Errors and System Intrusion represent 83% of breaches |
| **Threat Actors** | External (66%), Internal (37%), Multiple (4%), Partner (1%) (breaches) |
| **Actor Motives** | Financial (88%), Espionage (9%), Grudge (2%), Convenience (1%), Fun (1%) (breaches) |
| **Data Compromised** | Personal (70%), Credentials (32%), Other (27%), Internal (12%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Secure Configuration of Enterprise Assets and Software (4), Access Control Management (6) |

Errors and accidents, depending on your worldview, are either natural occurrences of complex systems or the fault of an intern who overcame your organization's robust and well-crafted safeguards. Regardless of your opinions on errors, they certainly are not uncommon in the Information sector. The pattern of Miscellaneous Errors, along with Basic Web Application Attacks and System Intrusion, accounted for 83% of breaches in this vertical.

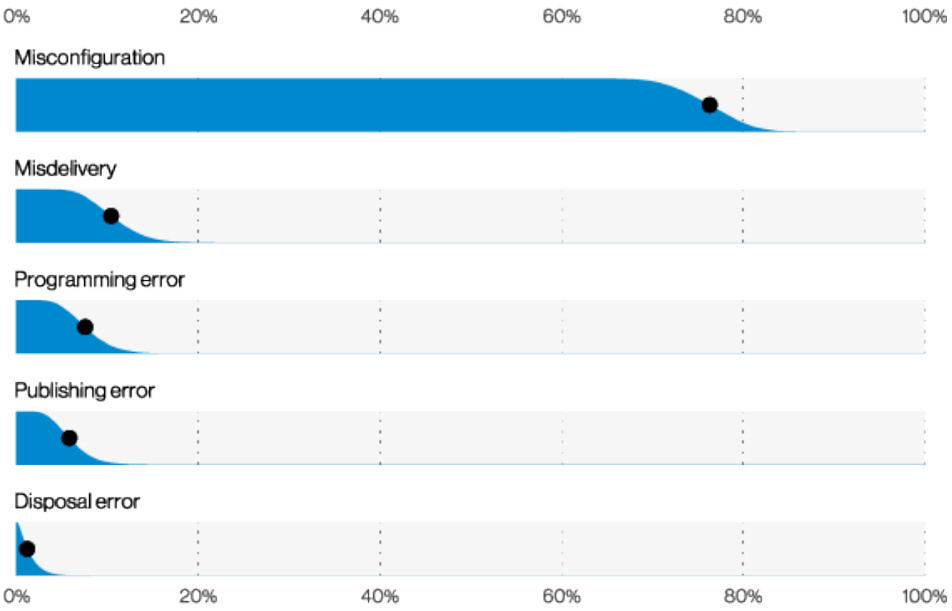



**Figure 106.** Error varieties in Information breaches (n=111)

In terms of the types of Errors seen, Misconfigurations accounted for over 70% of all Errors in this industry (Figure 106). This was followed by a three-way tie of Misdeliveries, Programming and Publishing Errors. With this combination, it shouldn't be a surprise that System Engineers (or are they called DevOps 24/7 Super Engineers?) had a strong showing in terms of the Internal actors responsible for those breaches. While the overall percentage of Error breaches hasn't increased over the last few years, it remains a persistent issue facing organizations in this sector.

When organizations discover that something unpleasant has occurred, External actors typically delivered the news (Figure 107). We found that 50% of the breaches were disclosed by the bad actor themselves, which sounds helpful of them, but really isn't. This is usually done either when a ransom note politely informs you that you're going to have a really bad day, or when actors openly share or sell your data on forums that are monitored by researchers and advisories alike—who then make the notification. Speaking of Security researchers, they accounted for 30% of these data breach discoveries.

If we look at only incidents, we find that this industry tends to be bombarded with DoS attacks, a trend that has been occurring ever since computers were networked, or at least since we've been doing this report (Figure 108). Of the incidents, DoS alone accounts for over 90% of the Hacking actions we observed, with the rest being credential-based attacks such as Brute force or the Use of stolen credentials.

We identified another interesting finding in the Information industry when we analyzed botnet-related breaches. This year, the amount of credential stealing botnet breaches targeting Information organizations overtook the Finance sector (Figure 109). Data is really the new oil, it seems.

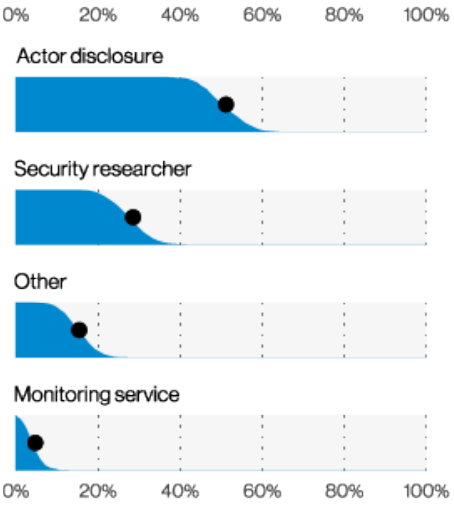



**Figure 107.** Top Discovery method varieties in Information breaches (n=84)

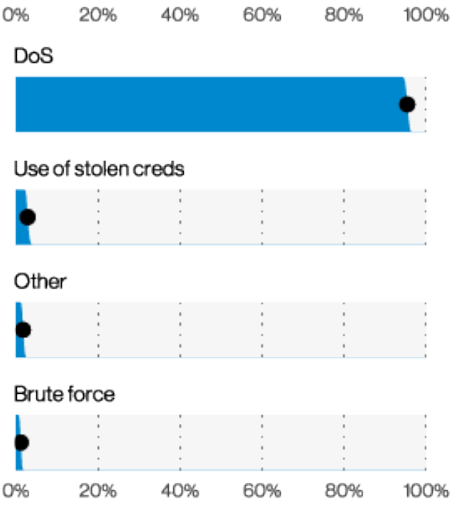



**Figure 108.** Top Hacking varieties in Information incidents (n=2,452)

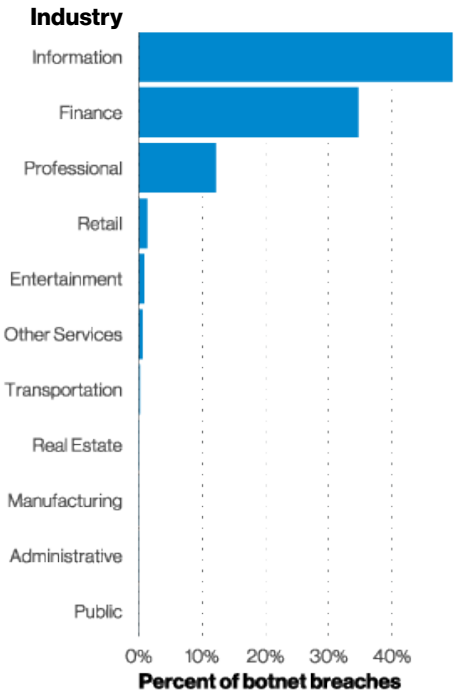



**Figure 109.** Industries in botnet breaches (n=222,162)

# Manufacturing NAICS 31-33

## Summary

This industry, like many others, is beset by Social Engineering attacks. Manufacturing also saw a marked rise in Ransomware related breaches.

| | |
|---|---|
| **Frequency** | 585 incidents, 270 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Social Engineering and Basic Web Application Attacks represent 82% of breaches |
| **Threat Actors** | External (82%), Internal (19%), Multiple (1%) (breaches) |
| **Actor Motives** | Financial (92%), Espionage (6%), Convenience (1%), Grudge (1%), Secondary (1%) (breaches) |
| **Data Compromised** | Personal (66%), Credentials (42%), Other (36%), Payment (19%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Access Control Management (6), Secure Configuration of Enterprise Assets and Software (4) |

As we confronted our organic almond milk and toilet paper shortages this past year, we were reminded of the real implications of continuous strain on factories and the manufacturing supply chain. Certain areas in this vertical faced some very unique and difficult challenges in 2020 due to the demand created by the pandemic. Even so, the Manufacturing sector was still not given a free pass by the threat actors who are not known for their magnanimity.

However, the challenges faced from a cybercrime perspective were not unique. In fact, Manufacturing suffered most from the same devious trio of System Intrusion, Social Engineering and Basic Web Application Attacks as did our overall breach dataset.

**Breaches**

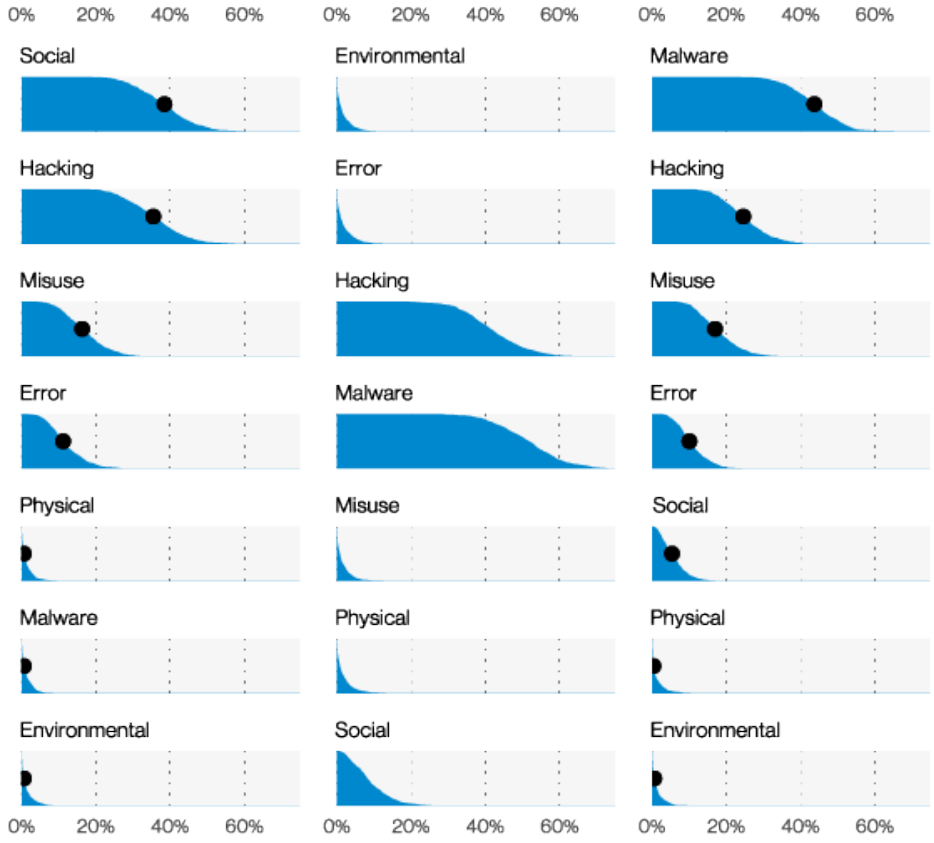



**Figure 110.** Actions at the beginning, middle and end of Manufacturing breaches

The scenarios play out in Figure 110, which illustrates the top Actions taken in each step of the breach. Threat actors were more likely to use a Social attack (75.4% were Phishing) or a Hacking attack (79.5% were Use of stolen credentials) to gain the initial foothold. From there, either additional Credentials would be compromised and utilized, or Malware would be installed.

On that note, Ransomware played a significantly increased role in Malware associated breaches (61.2%) in relation to previous years. This is likely attributable to the continued rise of "name and shame" tactics of Ransomware actors. In those cases, we can be sure the data has been compromised as well as rendered inaccessible in place.

Personal data was the most compromised data type in this sector, possibly also related to increased automation and the ease of attack. This data type (mostly consisting of customer PII) overtook Credentials, thus breaking the statistical tie we saw between them last year. This suggests more Actors are achieving their final goals, since Credentials breaches happen naturally as an attacker moves within an environment.

The number of ransomware related Malware incidents (as opposed to breaches discussed above) also saw a sharp increase from last year, overtaking both DoS and Phishing as the most common varieties of attacks shown in Figure 111.

If you are asking yourself the question "who would win in a fight: massive factories or one "encrypt-y boi?" the result may surprise you. This is definitely a great area to focus improvement with regard to this sector's defense strategy.

**The number of ransomware related Malware incidents (as opposed to breaches) also saw a sharp increase from last year, overtaking both DoS and Phishing as the most common varieties of attacks.**

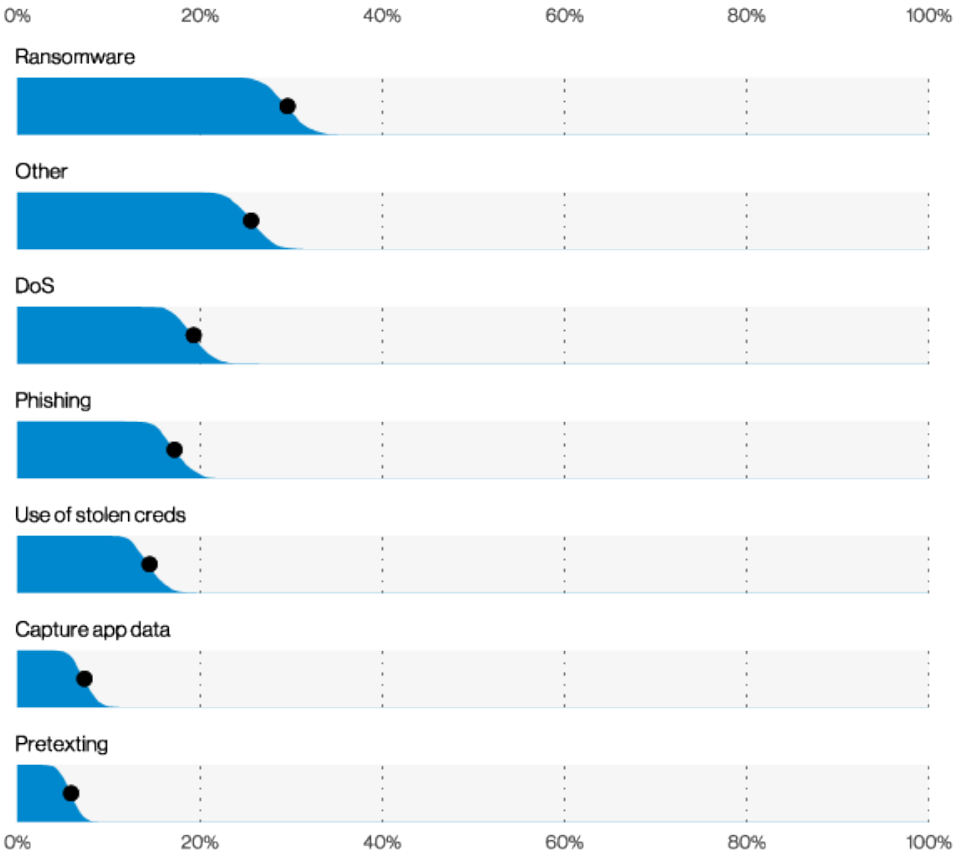



**Figure 111.** Top Action varieties in Manufacturing incidents (n=476)

# Mining, Quarrying, and Oil & Gas Extraction + Utilities

NAICS 21+22

## Summary

These industries suffered from Social Engineering attacks this year. Credentials, Personal and Internal data are the most commonly lost data varieties. Ransomware is also a major threat for these verticals.

| | |
|---|---|
| **Frequency** | 546 incidents, 355 with confirmed data disclosure |
| **Top Patterns** | Social Engineering, System Intrusion and Basic Web Application Attacks represent 98% of breaches |
| **Threat Actors** | External (98%), Internal (2%) (breaches) |
| **Actor Motives** | Financial (78%-100%), Espionage (0%-33%) (breaches) |
| **Data Compromised** | Credentials (94%), Personal (7%), Internal (3%), Other (3%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Access Control Management (6), Account Management (5) |

While most of us do not have to think about how to extract precious metals and minerals, or how to generate electricity and manage the complex infrastructure required to power up your PlayStation 5 (if you could find one), the folks in these industries have to do all those things on a daily basis. Not only must they combat various environmental threats, like thunderstorms, broken pipes and squirrels, but they also face threats from the cyber world. Let us dig into the industries that have made our modern connected world possible, despite how that modern connected world tries to bite the hands that feed them.

These industries do not differ vastly from other industries in regard to the top three patterns. However, the breakdown of these patterns does vary. In this sector, Social Engineering seems to be dominating both breaches and incidents this year, with sustained phishing campaigns occurring against some organizations (Figure 112). Social Engineering accounts for 86% of the breaches in this vertical, followed by System Intrusions and Basic Web Application Attacks.

The next most common type of attack is Ransomware, which accounts for 44% of non-Social Engineering attacks in this industry.

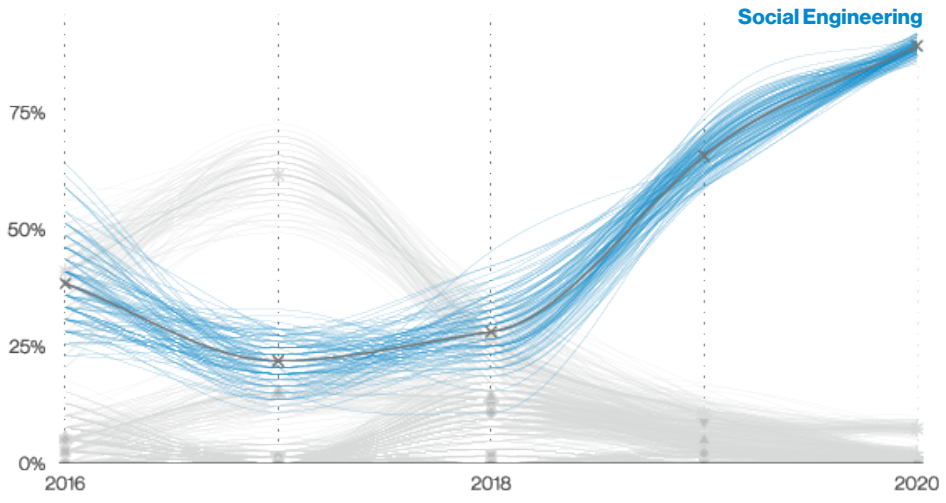



**Figure 112.** Patterns in Mining and Utilities incidents over time

# Professional, Scientific and Technical Services NAICS 54

## Summary

The combination of the System Intrusion and Social Engineering patterns account for the majority of cases in this sector. The Use of stolen credentials is widespread and employees have a definite tendency to fall for Social tactics.

| | |
|---|---|
| **Frequency** | 1,892 Incidents, 630 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Social Engineering and Basic Web Application Attacks represent 81% of breaches |
| **Threat Actors** | External (74%), Internal (26%) (breaches) |
| **Actor Motives** | Financial (97%), Espionage (2%), Grudge (1%) (breaches) |
| **Data Compromised** | Credentials (63%), Personal (49%), Other (21%), Bank (9%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Access Control Management (6), Secure Configuration of Enterprise Assets and Software (4) |

If Professional Services is your sector, you know that it is at best an eclectic NAICS code, with members that have wildly different footprints in terms of attack surfaces. One thing they seem to have in common is their reliance on internet connected infrastructure, and the risk inherent in that architecture. The System Intrusion and Social Engineering patterns competing for the top slots illustrates not only the vulnerability of that infrastructure, but also of the employees of these organizations (Figure 113).

The actors behind the System Intrusion pattern have some powerful tools at their disposal to gain access to their targets. Some of these cases began with the Use of stolen credentials or Exploiting a vulnerability, and ended with Malware being dropped on their victims. Frequently that malware was Ransomware, leading to extortion demands and downtime. The overall rise of Ransomware is something we've talked about in prior DBIRs, and the trend shows no signs of slowing. The growing tactic of the adversaries taking a copy of the data as a prod to help encourage their victims to pay up (which we saw begin just after the data collection period had ended for last year's report) has become increasingly popular as well. Thus we see a rise of Ransomware cases where there is also a confirmed data breach, as these actors post copies of their victim's data on the internet.


**Figure 113.** Patterns in Professional Services breaches (n=630)

Combine this with the Social Engineering pattern, and you have to worry about not only your infrastructure, but your people's ability to withstand Social tactics as well. Phishing was the leading Social action, but we also saw a good representation of Pretexting via email (Figure 114).

When you have the use of an invented scenario, the follow-on action is frequently an attempt to get money. This shows up in our data as a Fraudulent transaction and is represented along with the Integrity violation of Alter behavior when someone falls for the Social action (Figure 115).


**Figure 114.** Social varieties in Professional Services breaches (n=191)

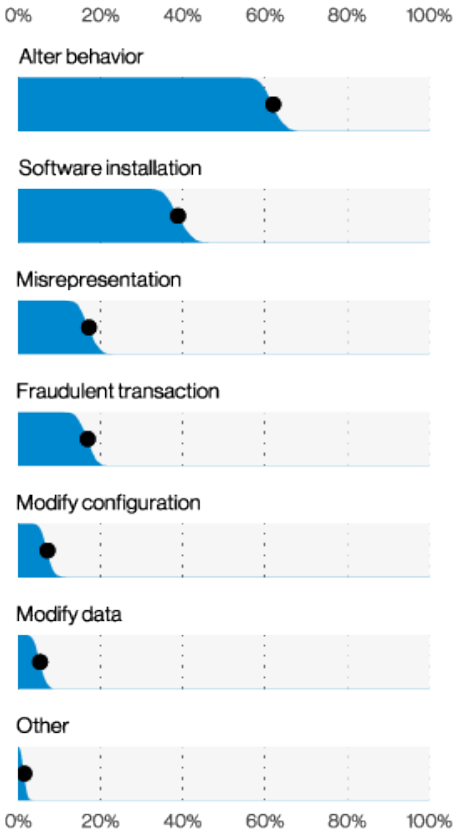



**Figure 115.** Top Integrity varieties in Professional Services breaches (n=337)

**Phishing was the leading Social action, but we also saw a good representation of Pretexting via email.**

# Public Administration NAICS 92

## Summary

By far the biggest threat in this industry is the social engineer. Actors who can craft a credible phishing email are absconding with Credentials at an alarming rate in this sector.

| | |
|---|---|
| **Frequency** | 3,236 incidents, 885 with confirmed data disclosure |
| **Top Patterns** | Social Engineering, Miscellaneous Errors and System Intrusion represent 92% of breaches |
| **Threat Actors** | External (83%), Internal (17%) (breaches) |
| **Actor Motives** | Financial (96%), Espionage (4%) (breaches) |
| **Data Compromised** | Credentials (80%), Personal (18%), Other (6%), Medical (4%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Access Control Management (6), Account Management (5) |

The Social Engineering pattern was responsible for over 69% of breaches in this vertical (Figure 116). Clearly, this industry is a favorite honey hole among the phishing fiends. The Social actions were almost exclusively Phishing with email as the vector (Figure 117). Pretexting was rarely leveraged at all, and why should they go to all the work of inventing a scenario when a straight up phish gets the job done?

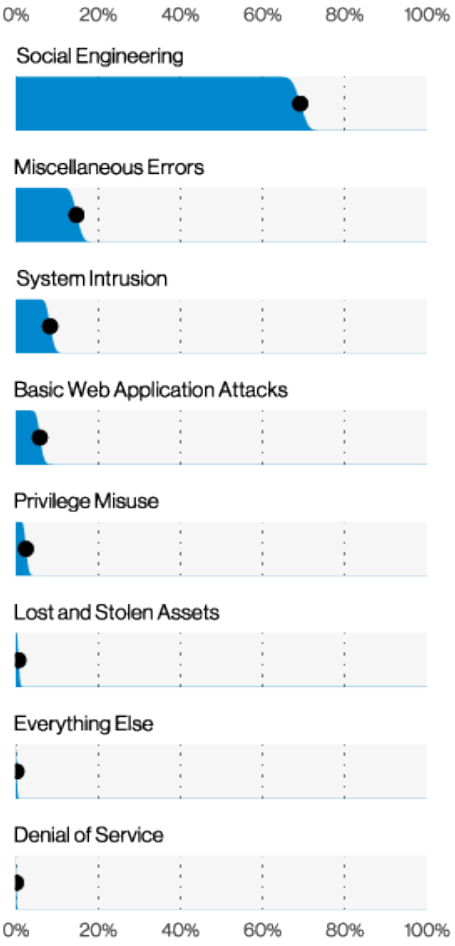



**Figure 116.** Patterns in Public Administration breaches (n=885)

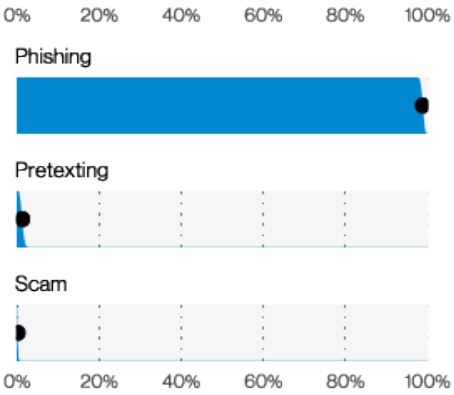



**Figure 117.** Social varieties in Public Administration breaches (n=611)

The Miscellaneous Errors pattern was a far distant second and consisted of Misconfiguration (although not usually found by Security researchers—which was a surprise, as that is the most common pairing) and Misdelivery (Figure 118). Certainly, government entities are responsible for a lot of mass mailings, and paper documents were the second most common assets that were delivered to the wrong recipient, with good old-fashioned emails taking first place.

The System Intrusion pattern rounds out our top three and is a combination of Hacking and Malware actions. We found the Use of stolen credentials, followed by dropping Malware with either C2 or ransomware capabilities to be the most common story in this pattern.

The most frequently stolen data type is Credentials, which are then used to further the attacker's presence in the victim's network and systems (Figure 119). After Credentials Personal information is the top data type compromised where breaches were confirmed in this sector.

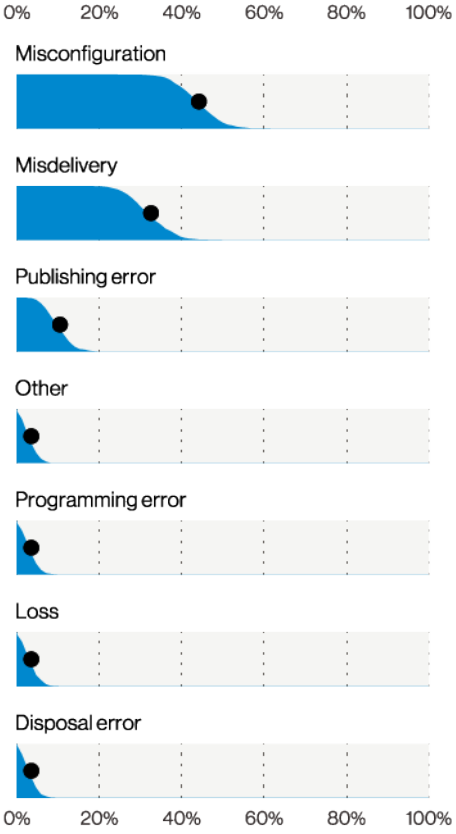



**Figure 118.** Top Error varieties in Public Administration breaches (n=86)

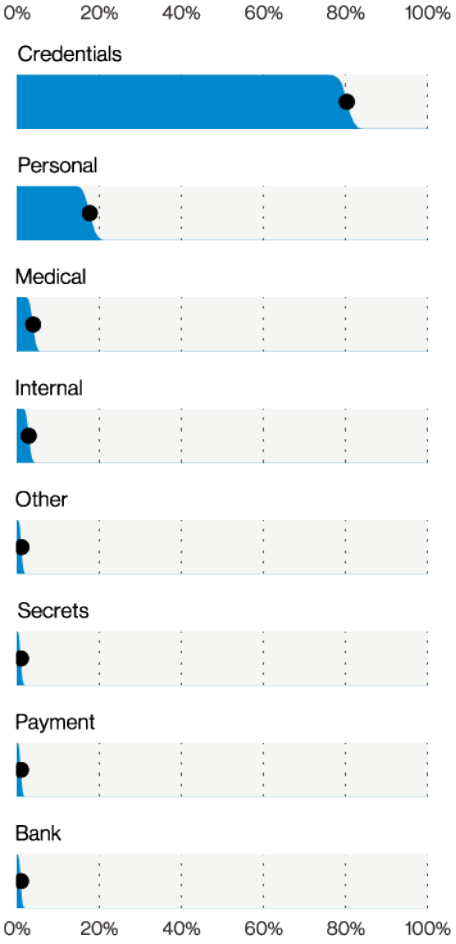



**Figure 119.** Top Data varieties in Public Administration breaches (n=841)

# Retail NAICS 44-45

## Summary

The Retail industry continues to be a target for Financially motivated criminals looking to cash in on the combination of Payment cards and Personal information this sector is known for. Social tactics include Pretexting and Phishing, with the former commonly resulting in fraudulent money transfers.

| | |
|---|---|
| **Frequency** | 725 incidents, 165 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Social Engineering and Basic Web Application Attacks represent 77% of breaches |
| **Threat Actors** | External (84%), Internal (17%), Multiple (2%), Partner (1%) (breaches) |
| **Actor Motives** | Financial (99%), Espionage (1%) (breaches) |
| **Data Compromised** | Payment (42%), Personal (41%), Credentials (33%), Other (16%) (breaches) |
| **Top IG1 Protective Controls** | Security Awareness and Skills Training (14), Secure Configuration of Enterprise Assets and Software (4), Access Control Management (6) |

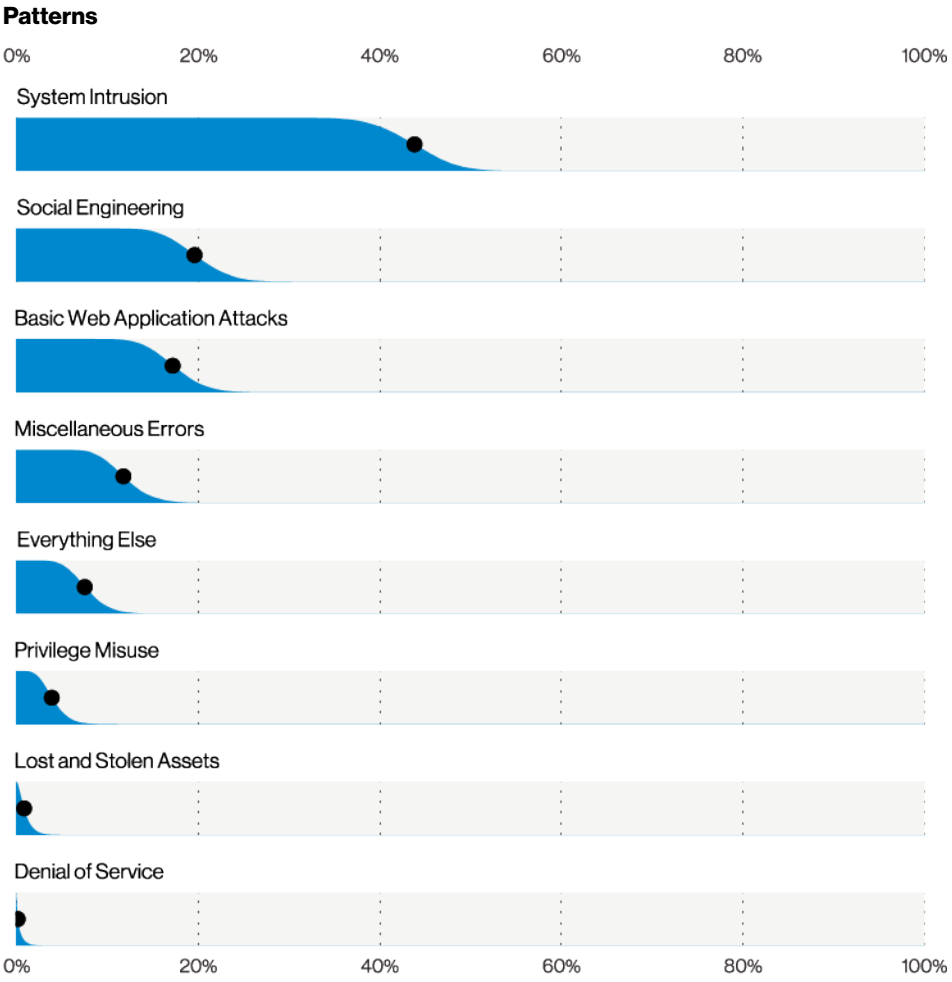



**Figure 120.** Patterns in Retail breaches (n=165)

The first noteworthy item in the At-a-Glance table is the difference in the number of incidents vs. the number of confirmed data breaches. The main cause of this was a large number of DoS attacks (409) that were launched against this sector. And while System Intrusion was the top pattern for breaches (Figure 120), it came in second place for incidents where no breach could be confirmed (177 incidents in this pattern, 69 of which were confirmed breaches).

Our main point here is: Don't let the low number of breaches fool you—this sector remains a target.

The System Intrusion pattern was prevalent, and tells the story of the common coupling of the Use of stolen creds with dropping Malware to capture application data. The Social Engineering pattern is a close runner up in this race, with

**We've said it before, and we'll say it again—everyone loves credentials. Credentials are the glazed donut of data types.**

Pretexting—where the adversary develops an invented scenario to get their target to take the bait (usually followed by a money transfer of some type)—being more common than we usually see in other industries (Figure 121). Don't get us wrong, the Phishing lure is still effective here. It is difficult to determine if the targeting of employees via Pretexting is a sign that criminals are having to work harder for the money, or if it is just simpler for the attackers to dupe employees into committing fraud on their behalf.

Unsurprisingly, the top data types compromised include Payment card data (which is largely what makes this industry so very attractive to Financially motivated criminals), Personal data (also useful for various kinds of financial fraud) and Credentials (Figure 122).

We've said it before, and we'll say it again—everyone loves credentials. Credentials are the glazed donut of data types.

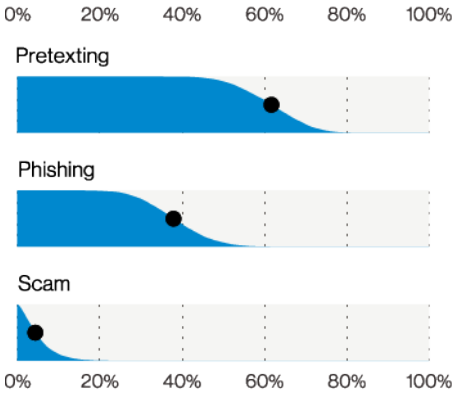



**Figure 121.** Social varieties in Retail breaches (n=32)

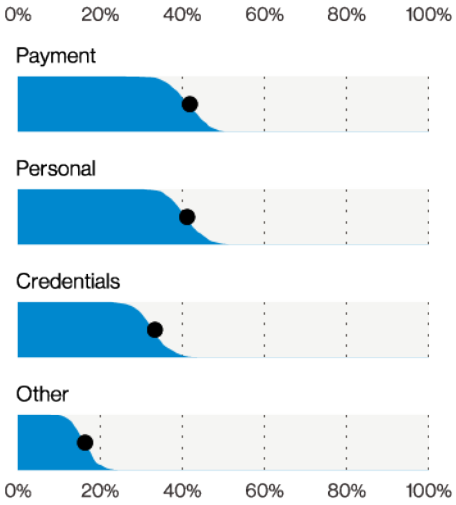
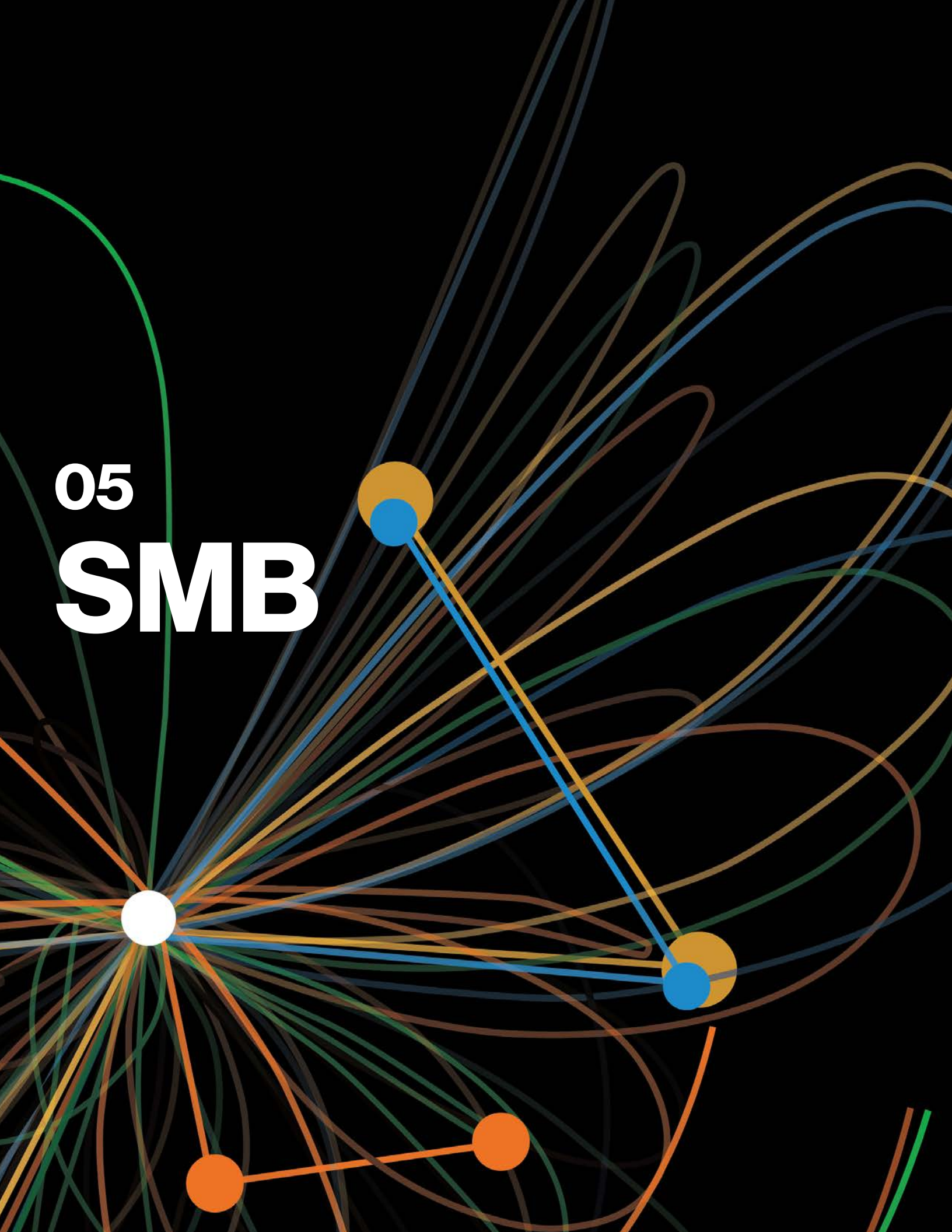



**Figure 122.** Top Data varieties in Retail breaches (n=153)



# 05 SMB

# Diving back into SMB breaches

| Small (Less than 1,000 employees) | |
|---|---|
| **Frequency** | 1,037 incidents, 263 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Miscellaneous Errors and Basic Web Application Attacks represent 80% of breaches |
| **Threat Actors** | External (57%), Internal (44%), Multiple (1%), Partner (0%) (breaches) |
| **Actor Motives** | Financial (93%), Espionage (3%), Fun (2%), Grudge (1%), Other (1%) (breaches) |
| **Data Compromised** | Credentials (44%), Personal (39%), Other (34%), Medical (17%) (breaches) |

## One size fits all-most

The first thing we noticed while analyzing the data by organizational size this year was that the gap between the two with regard to the number of breaches, has become much less pronounced. Last year, small organizations accounted for less than half the number of breaches that large organizations showed. Unlike most political parties, this year these two are less far apart with 307 breaches in large and 263 breaches in small organizations.

Another interesting finding was that the top patterns have aligned across both org sizes. For the first time since we began to look at this from an organizational size perspective, the two groups are very similar to each other and, at least pattern-wise, this seems like a "one size fits all" situation.

Last year, small organizations were greatly troubled by Web Applications, Everything Else and Miscellaneous Errors. The changes in our patterns account for a good bit of what we see this year in small organizations, since the Everything Else pattern was recalibrated, and the attacks that remain are largely Hacking and Malware, thus fitting into the System Intrusion pattern. In contrast, large organizations saw a fair amount of actual change. The top three last year were Everything Else, Crimeware and Privilege Misuse. The pattern recalibration means that most of the Crimeware type events went into System Intrusion and Basic Web Application Attacks, but Privilege Misuse is not a pattern that saw any substantial degree of change. Therefore, this is an indication that we saw fewer Internal actors doing naughty things with their employer's data.

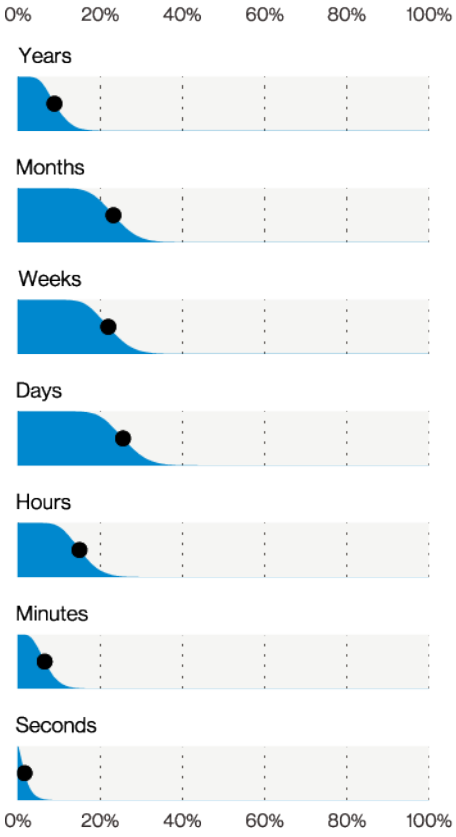



**Figure 123.** Discovery timeline in Small and Medium Business breaches (n=83)

## Large (More than 1,000 employees)

| | |
|---|---|
| **Frequency** | 819 incidents, 307 with confirmed data disclosure |
| **Top Patterns** | System Intrusion, Miscellaneous Errors and Basic Web Application Attacks represent 74% of breaches |
| **Threat Actors** | External (64%), Internal (36%), Partner (1%), Multiple (1%) (breaches) |
| **Actor Motives** | Financial (87%), Fun (7%), Espionage (5%), Convenience (2%), Grudge (2%), Secondary (1%) (breaches) |
| **Data Compromised** | Credentials (42%), Personal (38%), Other (34%), Internal (17%) (breaches) |

Since the patterns have now largely aligned between the two organizational sizes, we can talk a little about what that means for both. First, both are being targeted by financially motivated organized crime actors. This isn't a news flash to anyone (or shouldn't be) because professional criminals do tend to be motivated by money. For that matter, we'd wager most amateur criminals are as well (if we were the wagering type, which, of course, we aren't. As far as you know).

Concerning the common patterns of System Intrusion and Basic Web Application Attacks, those run the gamut of simple to complex attacks, frequently focused on web infrastructure. The Hacking action of Use of stolen creds followed by Malware installation is the playbook these actors prefer to follow. Increasingly, we see ransomware deployed by the actor after access; sometimes after they have taken a copy of the data to incentivize their victims to part with their hard-earned Bitcoin.

When we turn to Discovery timelines, we see a difference between the organizational sizes (Figures 123 and 124 respectively). Last year we reported that smaller organizations seemed to be doing better in terms of discovering breaches more quickly than their larger counterparts.

This year's data shows that large organizations have made a shift to finding breaches within "Days or less" in over half of the cases (55%), while small organizations fared less positively at 47%.

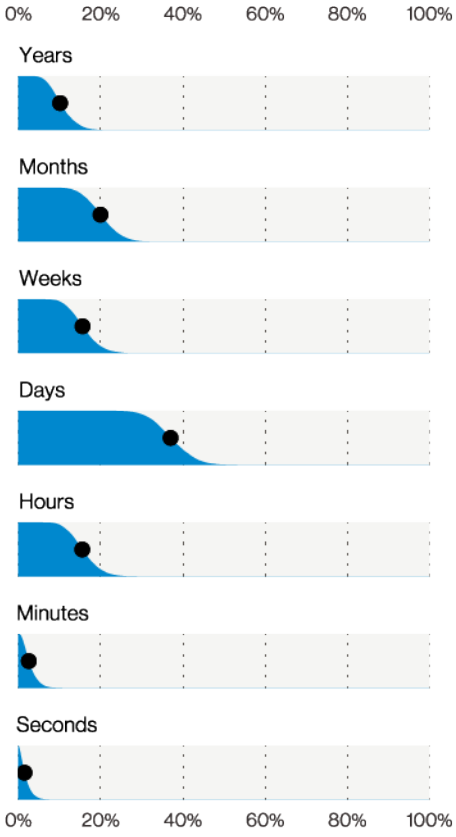
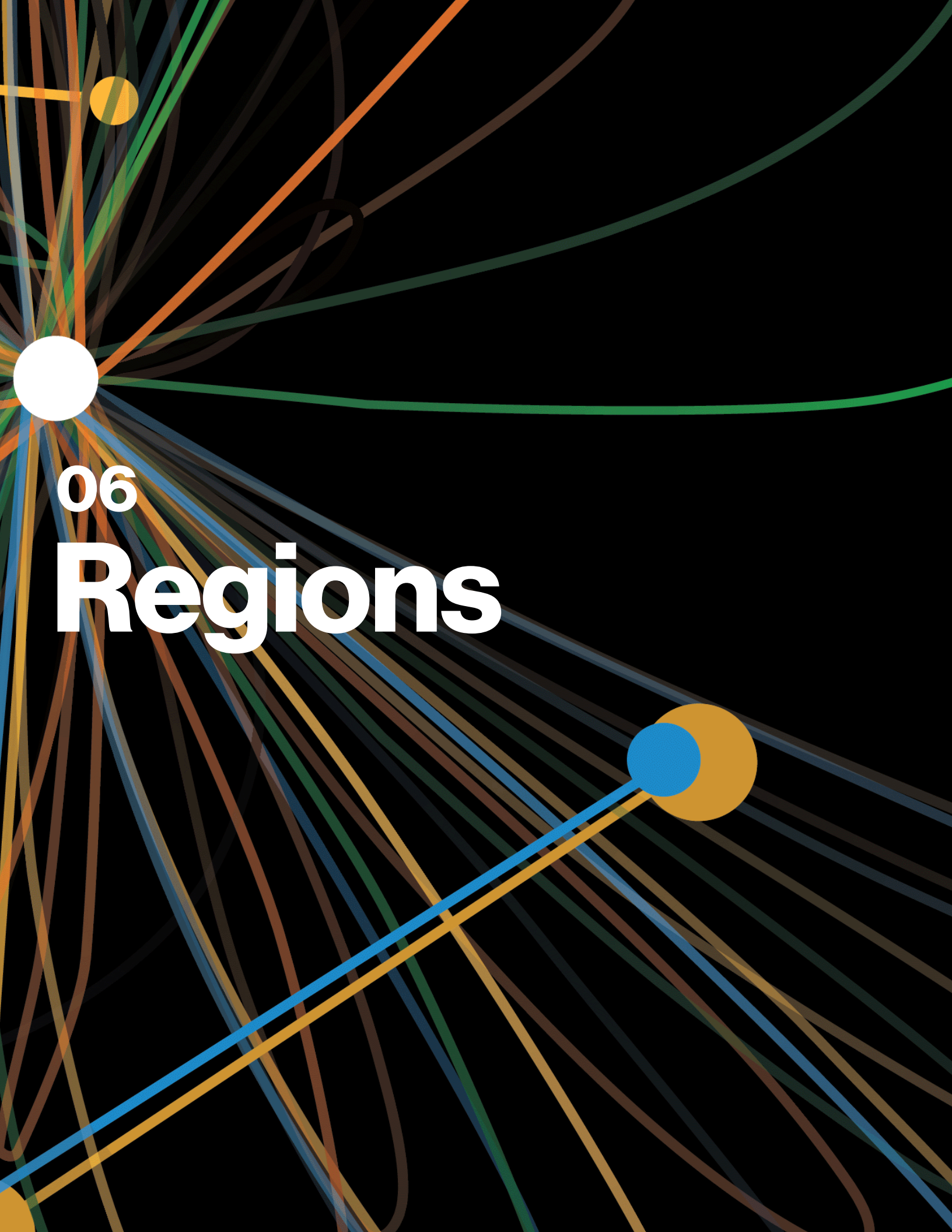



**Figure 124.** Discovery timeline in Large Business breaches (n=92)



# 06 Regions

# Introduction to Regions

Last year we analyzed incidents and presented them from a macro-region perspective for the first time. This year we once again visit (where possible) the various regions of the world in an attempt to provide readers with a more global view of cybercrime. As one might expect, we have greater or lesser visibility into a given region based on several factors such as contributor presence, regional disclosure regulations, our own caseload and so on.

Do you live and work in an area of the world that is not mentioned below? Do you feel that more focus should be given to your sphere of operations? Then contact us about becoming a data contributor and/or encourage other organizations in your area and industry to share their data so that we can continue to expand and refine our coverage each year. It is important to keep in mind that if you do not see your region represented here, it does not necessarily mean that we have no visibility at all into the region, but simply that we do not have enough incidents in that geographic location to be statistically relevant.[74]

We define the regions of the world in accordance with the United Nations M49 standards, which combine the super-region and sub-region of a country together. By so doing, the regions we will examine are as follows:

**APAC:** Asia and the Pacific, including Southern Asia (034), South-eastern Asia (143), Central Asia (143), Eastern Asia (030) and, last but certainly not least, Oceania (009).

**EMEA:** Europe, Middle East and Africa, including North Africa (002), Europe and Northern Asia (150) and Western Asia (145).

**NA:** Northern America (021), which primarily consists of breaches in the U.S. and Canada.

74 Don't blame the messenger. We don't make the rules here (they are made by the Illuminati and Intergalactic Aliens, or something like that).

# Asia Pacific (APAC)

## Summary

The most common type of breaches that took place in APAC were caused by Financially motivated attackers Phishing employees for creds, and then using those stolen creds to gain access to mail accounts and web application servers.

| | |
|---|---|
| **Frequency** | 5,255 incidents, 1,495 with confirmed data disclosure |
| **Top Patterns** | Social Engineering, Basic Web Application Attacks and Miscellaneous Errors represent 98% of breaches |
| **Threat Actors** | External (95%), Internal (6%) (breaches) |
| **Actor Motives** | Financial (96%), Espionage (3%), Fun (1%) (breaches) |
| **Data Compromised** | Credentials (96%), Personal (3%), Other (2%), Secrets (1%) (breaches) |


The APAC region covers an immense portion of the globe, and includes a multitude of nations, languages and diverse cultures, along with a fair share of venomous reptiles. In keeping with that diversity, the APAC region shows a relatively wide range of industries that were breached over the last year. All of the main verticals you might expect to see are present to some degree. Finance, Healthcare, Retail, Manufacturing and Education all make an appearance. In fact, for the first time ever we saw more breaches in APAC last year than in any other region.

One industry in particular that posted impressive numbers this year was NAICS 21: Mining, Quarrying, and Oil and Gas Extraction (Figure 125). This was due to the fact that organizations in that vertical fell prey to sophisticated Social Engineering attacks.

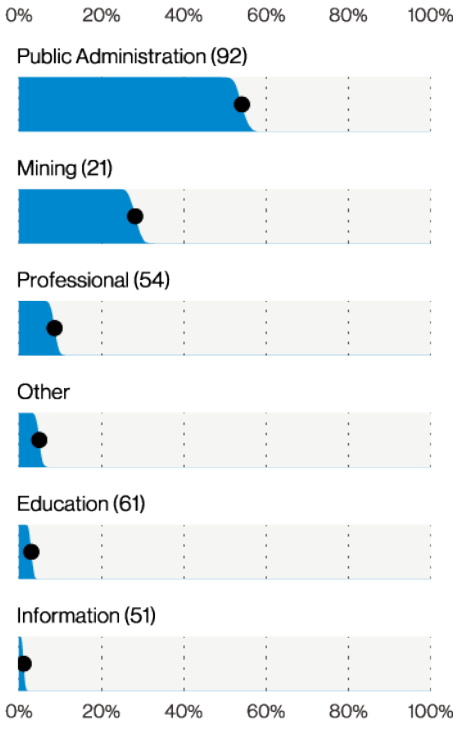



**Figure 125.** Top industries in APAC breaches (n=1,130)

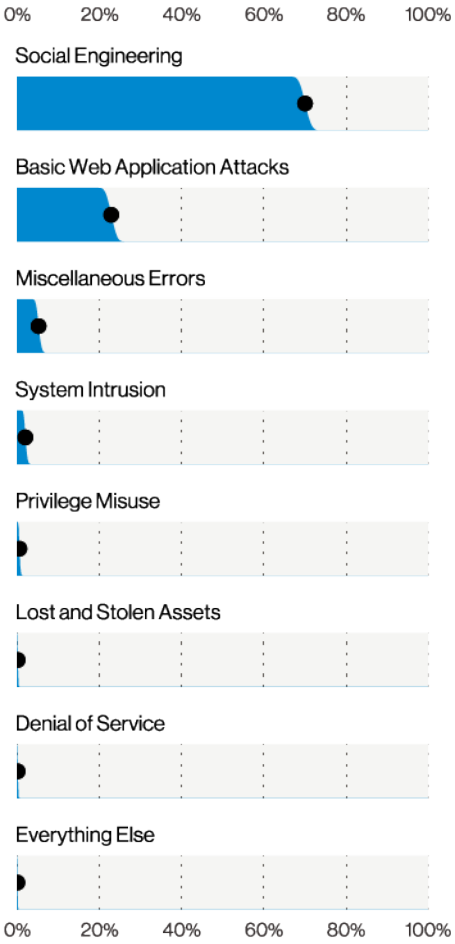



**Figure 126.** Patterns in APAC breaches (n=1,495)

As Figure 126 illustrates, 70% of attacks in APAC contained a Social Engineering action, typically Phishing. What those attacks harvested were almost exclusively Credentials (98%). Those creds were then either used to escalate or laterally expand the Social attack or were otherwise utilized to hack into web applications operated by the organization (23%).

If you have perused the other sections of this report, you might be asking yourself certain questions at this point. Who perpetrated these crimes? Were they in a dark room wearing a hoodie? Why am I not seeing ransomware in this region? All good questions, and as far as we can tell, they were mostly committed by Financially motivated organized criminals. While we have only anecdotal data on this topic, we feel certain that hoodies and dark rooms were involved to some degree. But regarding the last and most interesting of those questions (where is ransomware?), it holds the number 10 spot in Malware variety for APAC, which is in relatively stark contrast to what we see elsewhere. However, this is almost certainly a byproduct of our contributors and their caseload rather than an actual dearth of this type of malware. We expect the "stand-and-deliver, your money or your data" attacks are flourishing in APAC as they most certainly are in other regions.

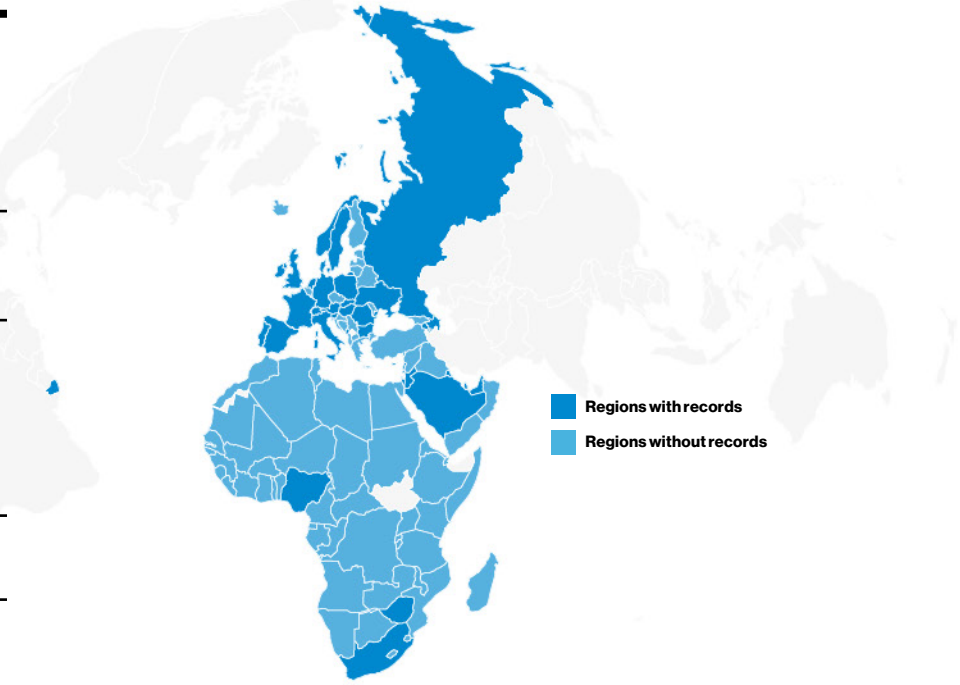
# Europe, Middle East and Africa (EMEA)

## Summary

EMEA continues to be beset by Basic Web Application Attacks, System Intrusion and Social Engineering.

| | |
|---|---|
| **Frequency** | 5,379 incidents, 293 with confirmed data disclosure |
| **Top Patterns** | Basic Web Application Attacks, System Intrusion and Social Engineering patterns represent 83% of breaches |
| **Threat Actors** | External (83%), Internal (18%) (breaches) |
| **Actor Motives** | Financial (89%), Espionage (8%), Fun (1%), Grudge (1%) (breaches) |
| **Data Compromised** | Credentials (70%), Internal (52%), Personal (22%), Other (16%) (breaches) |

**For the second year in a row, Basic Web Application Attacks are the most commonly seen pattern in this region, accounting for approximately 54% of breaches.**




EMEA is made up of Europe, the Middle East and Africa. For the second year in a row, Basic Web Application Attacks are the most commonly seen pattern in this region, accounting for approximately 54% of breaches.

Sometimes these attacks are aimed at obtaining the data within the application itself, but in other cases it is simply a means to an end in order to perpetrate other forms of badness.

The System Intrusion, Social Engineering and Miscellaneous Errors patterns are all closely grouped for second place in this region (Figure 127). By far the most often breached data type in EMEA is Credentials, and this goes some way toward explaining the placement of the patterns. While in many cases we know that stolen Credentials were used, we do not always have visibility into how they were initially acquired. However, we do know that Social Engineering in the form of Phishing is very often the means attackers use to obtain them.

Regardless of how they originally got their grubby little hands on them, using stolen Credentials is the primary means by which the actor hacks into the organization, and in many cases, it is via a Web application.

Finally, 17% of actors in EMEA are Internal (most often system administrators), which explains the presence of Miscellaneous Errors in the top four patterns. In the majority of cases (67%), these are unintentional Misconfiguration errors.

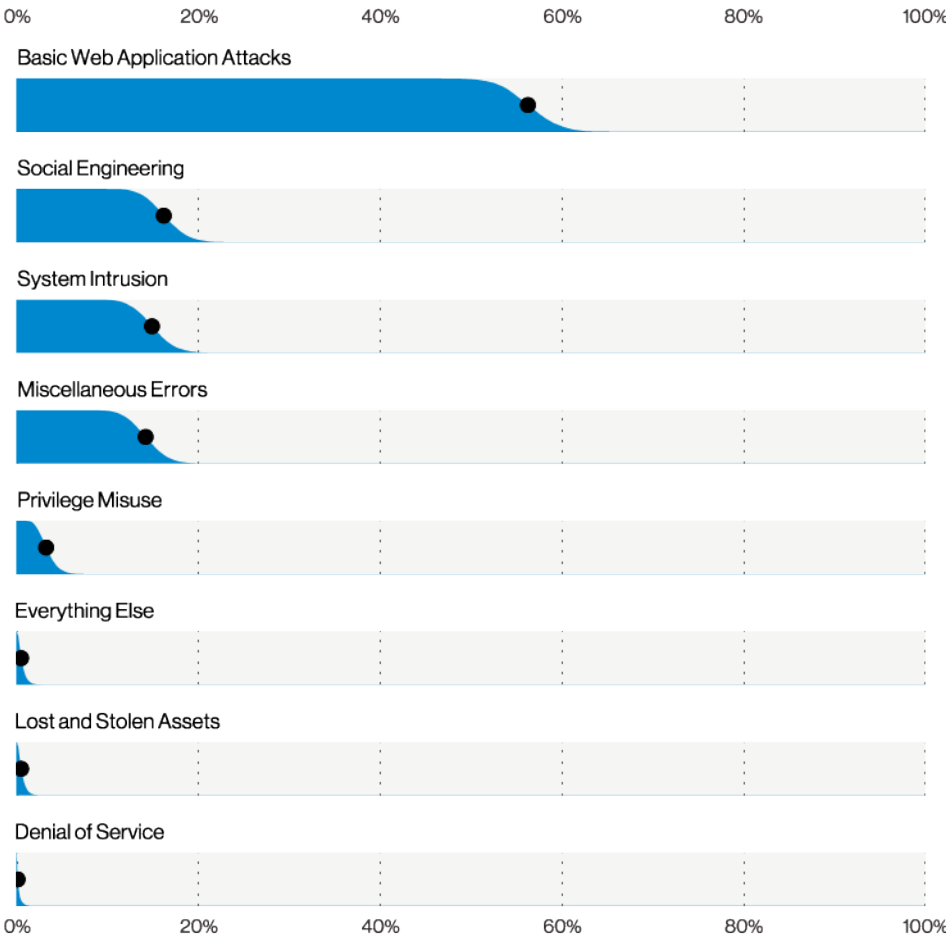



**Figure 127.** Patterns in EMEA breaches (n=293)

# Northern America (NA)

## Summary

Northern American organizations continue to be the target of Financially motivated actors searching for money or easily monetizable data. Social Engineering, Hacking and Malware continue to be the favored tools utilized by these actors.

| | |
|---|---|
| **Frequency** | 13,256 incidents, 1,080 with confirmed data disclosure |
| **Top Patterns** | Social Engineering, System Intrusion and Basic Web Application Attacks represent 92% of breaches |
| **Threat Actors** | External (82%), Internal (19%), Multiple (2%), Partner (1%) (breaches) |
| **Actor Motives** | Financial (96%), Espionage (3%), Grudge (2%), Fun (1%) (breaches) |
| **Data Compromised** | Credentials (58%), Personal (34%), Other (27%), Internal (11%) (breaches) |


When viewing data regarding incidents and breaches in Northern America, it is important to realize the influence of the regulatory environment on the numbers shown.

Data breach disclosure laws in this region are prevalent and far reaching with the result that our visibility into cybercrime is better than in areas where such laws are not in place. Healthcare and Public Administration are among the more strongly regulated industries; therefore, we see a corresponding prevalence in these industries. In addition to the aforementioned laws, one must keep in mind that we also have more contributors in this geographical area than in others.

There seem to be two very distinct competitions with regard to Northern America's data (Figure 128). The first of these is a tight race between Social Engineering and System Intrusion (approximately 35% each). The second struggle is between Basic Web Application Attacks and Miscellaneous Errors for a smaller piece of the action. The confidence intervals overlap to such a degree between those groups that it is very difficult to call a clear winner. Therefore, when looking at the statistics from these patterns, keep in mind what we are really seeing are two sets of partners dancing together.

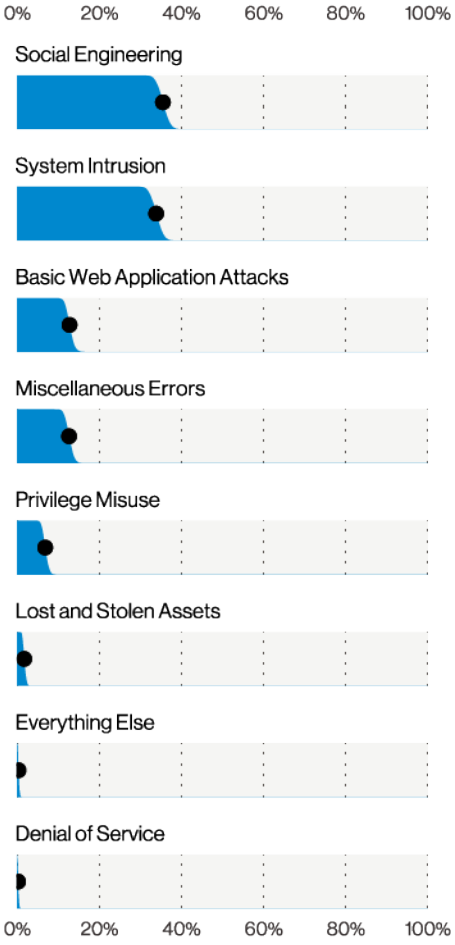



**Figure 128.** Patterns in Northern American breaches (n=1,080)

Our brand-new Social Engineering pattern is largely comprised of Pretexting and Phishing actions (Figure 129). Usually, we see more of the simple type of phishing activities than we do people going to the trouble of inventing a scenario. As a rule, criminals tend to be efficient in their efforts and the basics usually bring success, so why put in more work than necessary? One possible answer is that the end goal of the Pretexter is not the same as that of the standard Phisher. Pretext attacks are frequently an attempt to get a direct route to the money: The most common goal is to influence the target to send them money (under false pretenses, of course). These invented scenarios vary somewhat, but examples include the substitution of banking information, or the payment of fictitious invoices. A phisher, in contrast, may be going for data rather than cash, and their aim may ultimately be either to monetize the data stolen in the phish (Credentials), or to gain a foothold into the organization. The System Intrusion pattern (also newly minted) most often tells the story of a Hacking action paired with a Malware action. We typically see the Use of stolen creds to gain access, followed by the actor dropping Malware to further their aims in the organization. In Northern America, this most commonly means the deployment of Ransomware. As mentioned in last year's report, we saw Ransomware groups begin pivoting to take a copy of the data for use as leverage against their victims prior to triggering the encryption. This began with the Maze Group, and as they enjoyed success, other groups jumped onto the bandwagon. Now it has become commonplace, with many of the Ransomware groups having developed infrastructure specifically to host these data dumps.

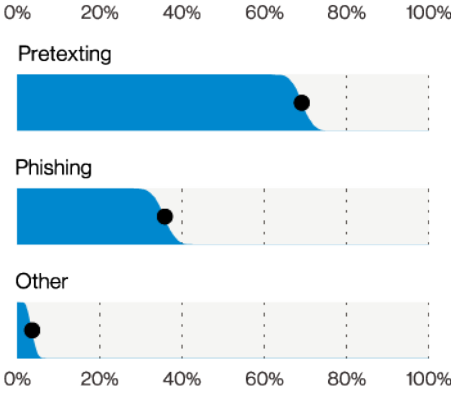



**Figure 129.** Social varieties in Northern American breaches (n=385)

All of these Social and Malware actions share one characteristic—they cause Integrity violations in the CIA triad. For the Social attacks, Alter behavior shows up to account for the change in the behavior of the victim affected by the Social action. For the Pretexting attacks that were successful, you can see the Fraudulent transaction Integrity attribute when the criminal managed to get someone to send them cash. Malware, of course, results in Software installation as a violation, and Misrepresentation is another side effect of Phred the Phisherman and Patti the Pretexter, both pretending to be someone they aren't (like most everyone else), and attempting to gain more victims in the organization (more followers, if you will).

Given the prevalence of the Phishing attacks, this is where the Credentials frequently come into play (Figure 130). Personal data is a prime target as well, since that includes such data elements as Social Security/Insurance numbers paired with other bits of information that allow criminals to commit further financial fraud.

Looking at our Discovery timeline, you can see a significant percentage are discovered in Days or less (Figures 131 and 132 respectively). However, over half of these cases were discovered by the threat actor disclosing the breach—this is typically the way Ransomware is discovered, when the ransom note flashes up on the screen.

We would expect to see that happen soon after the encryption is triggered. While we would rather see internal detective controls be responsible for finding the majority of the breaches, at least when that ransom note appears, organizations can start to contain the breach and get the actors out of their network.

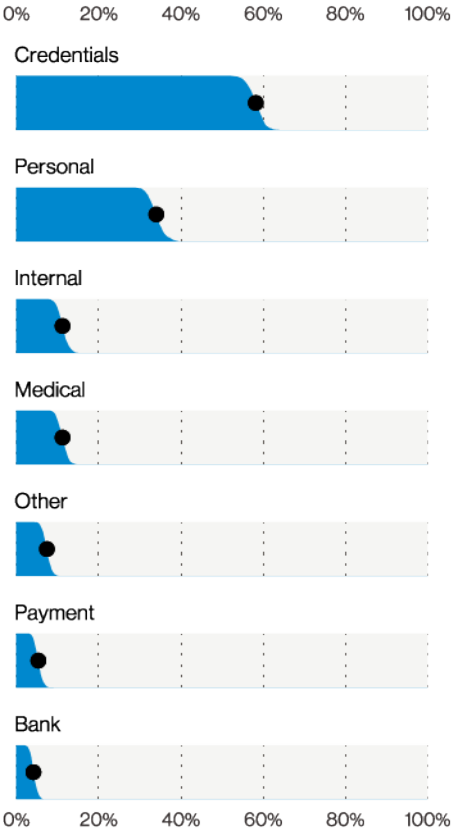



**Figure 130.** Top Data varieties in Northern America breaches (n=579)

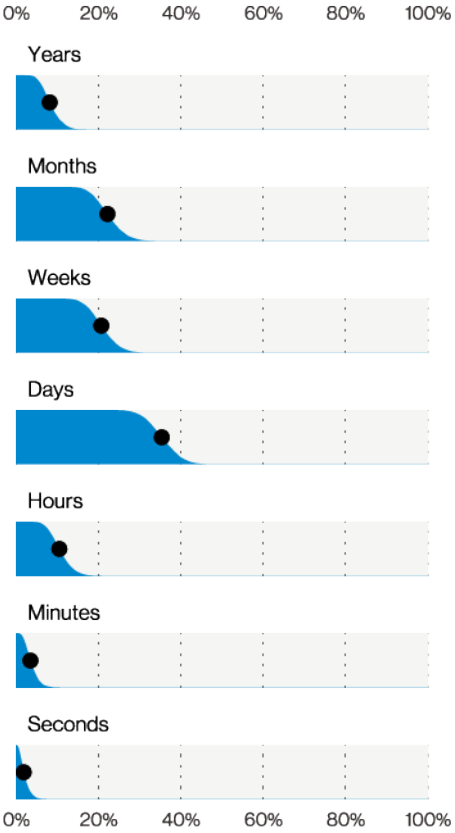



**Figure 131.** Discovery timeline in Northern America breaches (n=128)

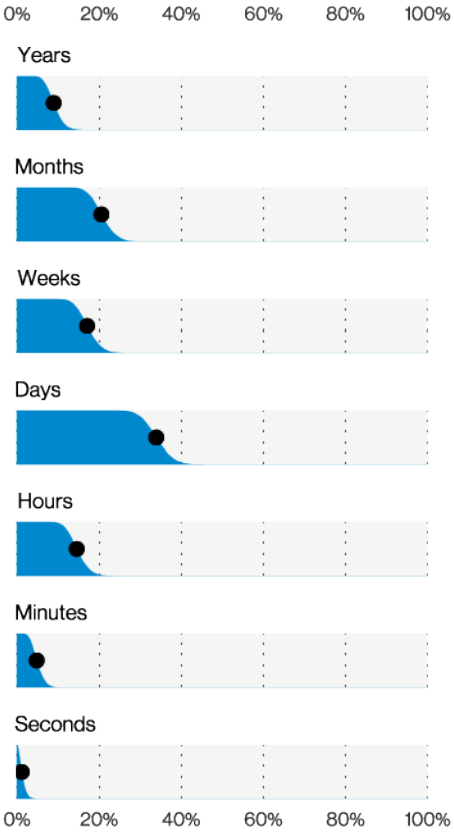



**Figure 132.** Discovery timeline in breaches (n=195)

# 07
# Wrap-up

# Here we are at last, at the conclusion of the 14th installment of the Verizon Data Breach Investigations Report.

**Give yourselves, and each other, a pat on the back, or even better, a big virtual hug.[75] All will be well. Thank you, readers, for spending time here with us yet again. We hope that the information contained in these pages has been of assistance to you and that you found it both informative and easy to ingest. As we mentioned at different points in this year's report, it is not always easy to see what is coming at us around the next bend. But one thing we do know is that if we meet whatever it may be with reason, with compassion and caring,[76] most importantly, with each other, we can handle it.**

**Of course, we can't close out a report without thanking our contributors who freely give their time, their expertise and, most importantly, their data to make this report a reality each year. On behalf of the DBIR Team, we thank you all. We encourage you, our readers, to reach out to us with your questions, comments and thoughts, or just to say hi. Here is hoping that we will find you all with us next year for number 15. Stay safe, and be happy!**

75 Or a real one if you have really long arms.
76 As Dan Kaminsky would do.

# Year in review[77]

## January

The Verizon Threat Research Advisory Center intelligence collections in both 2019 and 2020 began with cyber espionage targeting cloud environments by the Chinese menuPass threat actor. Among the ongoing threats were attacks on remote access. These included attacks on new vulnerabilities in Citrix products and continued password spraying attacks on Pulse Secure, FortiOS and Palo Alto VPN servers. London-based financial services company Travelex suffered a Sodinokibi ransomware infection on New Year's Eve that some sources claimed was the result of failing to patch a Pulse Secure VPN server. The U.S. Coast Guard announced a port facility had to shut down for 30 hours due to a Ryuk infection. The first zero-day attacks in 2020 exploited CVE-2020-0674 Internet Explorer use-after-free vulnerability in JScript. Qihoo 360 reported a watering hole attack by the DarkHotel actor using a cocktail of exploits: CVE-2020-0674 (Internet Explorer JScript) and CVE-2019-17026 (Firefox) and CVE-2017-11882 (Office Equation editor).

## February

The Australian Cyber Security Centre issued an advisory on ransomware known as "Mailto" or "Netwalker" after the Australian transportation and logistics company The Toll Group suffered an attack. On patch Tuesday, Microsoft released 99 patches including one for CVE-2020-0674. Another patch was for a vulnerability in Microsoft Exchange, CVE-2020-0688. Within two weeks, the VTRAC collected intelligence about mass scanning and exploitation targeting the Exchange Server vulnerability. The Cybersecurity and Infrastructure Security Agency (CISA) issued an alert with intelligence about a Ryuk ransomware attack on a natural gas pipeline facility. Industrial Control Systems (ICS) security company Dragos released an assessment with links to January's U.S. Coast Guard report. Five days after releasing a new version of their Chrome browser, Google released another to mitigate a type confusion vulnerability, CVE-2020-6418, that was being exploited in the wild (ITW).

## March

Fans of Westerns (movie genre) will recognize "ringing the chuck wagon triangle bell" at dinnertime. COVID-19 began to have the same effect for cybercriminals. Perhaps the most immediately useful collection was RiskIQ's COVID-19 Daily Update reports and domain watch or block lists. Prevailion and Proofpoint produced intelligence on TA505 attacks using COVID-19 bait. Before the end of the month, Microsoft was warning customers about limited targeted attacks exploiting a new Windows 7 vulnerability. Windows 10 was not vulnerable. CVE-2020-1020 was a security flaw in the Adobe Type Manager Library. FIN7 targeted a Trustwave customer with a malicious USB drive in conjunction with a US$50 gift card bait.

## April

BAH published a re-assessment of 200-plus cyber operations by the GRU (Russian military intelligence) concluding they conform to Russian strategic doctrine, which makes them somewhat more predictable. Recorded Future leveraged MITRE's ATT&CK for a report exploring the most common cyber-attacker TTP in 2019. Malwarebytes published "APTs and COVID-19: How advanced persistent threats use the coronavirus as a lure." Two other resources for cybersecurity during the COVID-19 pandemic were BBC cybersecurity correspondent Joe Tidy's searchable Coronavirus Phishing Scams collection, and the National Cyber Security Alliance's COVID-19 Security Resource Library. Three of the 113 vulnerabilities patched by Microsoft were being exploited ITW. Patches for CVE-2020-1020 and CVE-2020-0938 mitigated the "limited targeted Windows 7 based attacks that could leverage un-patched vulnerabilities in the Adobe Type Manager Library." The third surprise attack exploited a Windows kernel elevation of privilege vulnerability, CVE-2020-1027. But before the end of April, Microsoft released an out-of-cycle advisory for a vulnerable Autodesk DLL, CVE-2020-7085.

77 Thanks to David M. Kennedy from the VTRAC for this contribution.

**May**

Oracle reported ITW exploitation attempts on WebLogic servers without the patch for CVE-2020-2883 that was in the April Critical Patch Update. F-Secure announced two severe vulnerabilities in SaltStack Salt management framework, a configuration management and administration tool frequently used in data centers and cloud environments including Amazon Web Services and GCP. CISA published "Top 10 Routinely Exploited Vulnerabilities." New intelligence from ESET detailed Winnti attacks on video game companies in South Korea and Taiwan. Taiwan's Ministry of Justice believes Winnti was responsible for ransomware attacks on both of the countries' oil refineries. Broadcom/Symantec intelligence covered attacks on telecommunications companies in South Asia by the Greenbug threat actor. Cisco disclosed that six of its backend servers were compromised by hackers who exploited SaltStack vulnerabilities CVE-2020-11651 and CVE-2020-11652. The Australia logistics giant Toll Group was hit by a second ransomware attack in three months. Trustwave disseminated a report on "GoldenSpy," a backdoor in the tax payment software mandated by the Chinese bank of a UK-based technology company.

**June**

Cycldek, a low-profile Chinese threat actor deployed "USBCulprit" malware that Kaspersky assessed is intended to spread to and exfiltrate data from systems isolated from the internet. None of the 150-plus vulnerabilities patched in June were being exploited prior to patch release. Australian Prime Minister Morrison said Australian organizations, including governments and businesses, are currently being targeted by a sophisticated foreign "state-based" actor. The "Evil Corp'' APT-grade cybercrime threat actor began "big game hunting" with relatively new WastedLocker ransomware. NCC Group and Symantec independently released intelligence on the new Evil Corp campaign.

**July**

Enterprises with F5 BIG-IP appliances were at risk from attacks on two new vulnerabilities that U.S. Cyber Command called to be "remediated immediately." Exploit code was ITW. BIG-IP honeypots had been attacked and malware installed. FortiGuard, Palo Alto and Deep Instinct each reported intelligence about EKANS (SNAKE) ransomware that sidelined systems at Honda and Enel. Citrix released a security bulletin and patches for 11 new vulnerabilities in Citrix ADC, Gateway and SD-WAN. Within three days, the VTRAC collected reports of Citrix exploit detections by honeypots followed by de rigueur attempts to install cryptocurrency mining software. The U.K., U.S. and Canada jointly reported APT29 (Cozy Bear) (Russia) has been targeting COVID-19 vaccine research organizations. Sansec reported the Lazarus Group had been attacking U.S. and E.U. e-tailers using Magecart payment card skimming. McAfee and SentinelOne each reported different campaigns by Lazarus.

**August**

We collected security advisories about Cisco firewalls and TeamViewer, the management tool used by many managed service providers and their clients. We collected intelligence on campaigns spreading new variants of banking Trojans: IcedID, Dridex and Emotet. MITRE published, "2020 CWE Top 25 Most Dangerous Software Weaknesses." Three U.S. agencies released joint reports on a newly distinguished North Korean threat actor, "BeagleBoyz," and malware that the actor uses for ATM "jackpotting" attacks. F-Secure reported North Korean actors targeting virtual currency organizations.

**September**

Group-IB reported "UltraRank" an actor behind Magecart payment card skimming campaigns since 2015. SWIFT and BAE Systems released a report on the cybercrime economy fittingly titled, "Follow the Money." CISA released two products covering Iranian threat activity. Several vulnerabilities used by Iranian actors are also favored by ransomware actors according to SenseCy. Intel 471 assessed Lazarus has been using Russian crimeware for initial access to their targets. Microsoft Security reported ITW attacks exploiting systems without patches for the so-called "ZeroLogon" vulnerability, CVE-2020-1472.

**October**

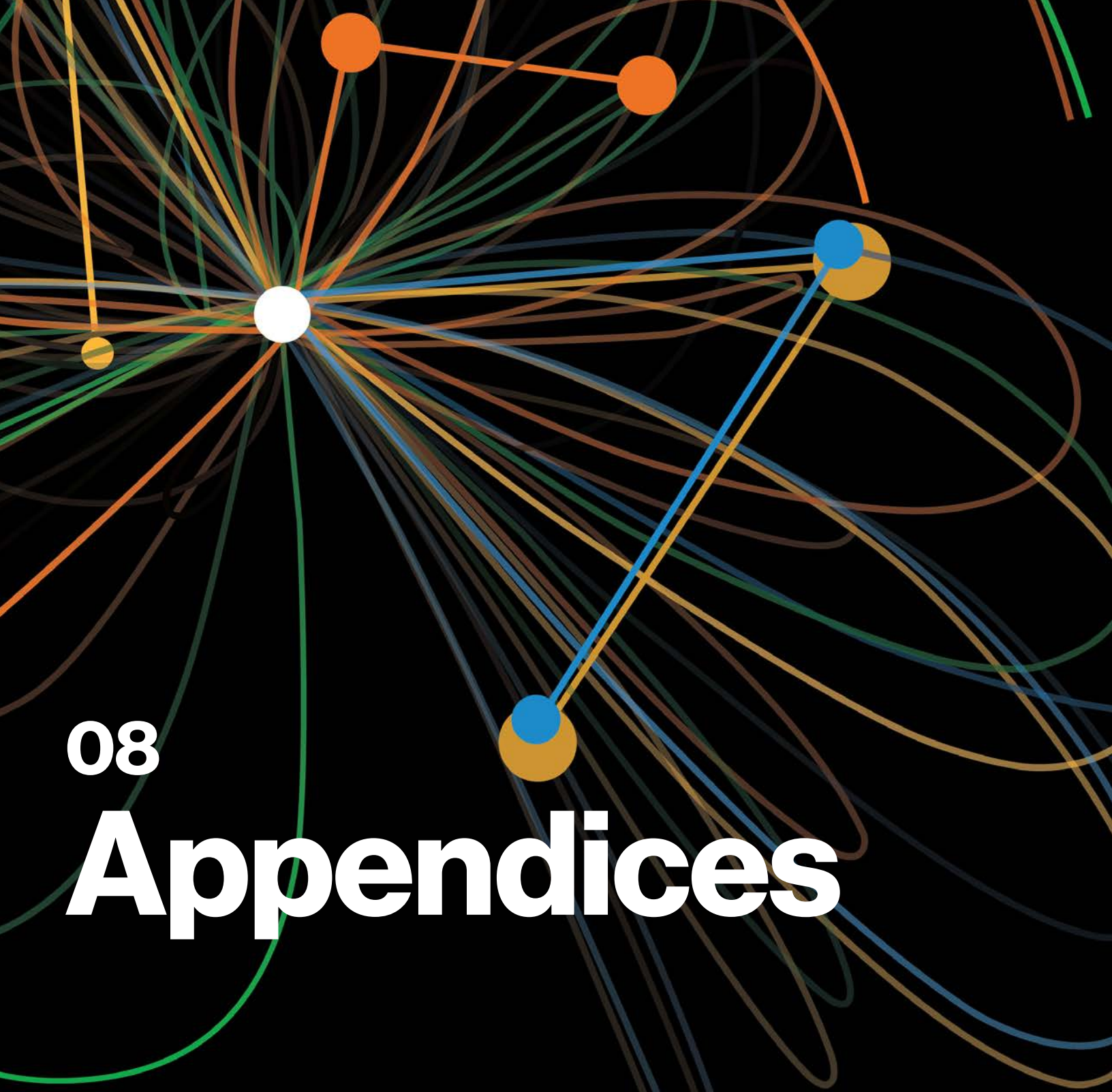
The Australian Cyber Security Centre (ACSC) issued an advisory on an "ongoing and widespread" Emotet campaign impacting Australian organizations. The VTRAC continued to collect threat intelligence about exploitation of Netlogon/ZeroLogon (CVE-2020-1472). CISA and Microsoft have observed Netlogon/ZeroLogon exploitation by APT-grade actors like MuddyWater and TA505. The MuddyWater Iranian APT actor has been targeting Israeli organizations according to ClearSky Security. Telsy attributed MuddyWater was behind another campaign targeting professionals in the aerospace and avionics sectors in Italy. Google said it mitigated a 2.54 Tbps DDoS attack, one of the largest ever recorded. The U.S. barbeque restaurant chain Dickey's suffered a point-of-sale attack between July 2019 and August 2020.

**November**

The VTRAC collected risk-relevant intelligence about eight new vulnerabilities, three of which have already been exploited and the remainder having exploit code ITW without reports of successful attacks. November's Patch Tuesday came with 114 Microsoft patches, two Adobe product updates, 12 SAP security notes (six Hot News), four Chrome browser updates and 40 Intel security advisories. Exploit code was already ITW for one Microsoft and five Chrome browser vulnerabilities. Bitdefender released a report of Chinese APT attacking South East Asian governments. Attacks by Lazarus and Kimsuky were reported by ESET and EAST Security respectively. Egregor ransomware has been establishing itself as the successor to Maze ransomware. The Australian Cyber Security Centre alerted the healthcare sector about TA505 attacks using SDBBot remote access Trojan and Clop ransomware.

**December**

Malwarebytes and CERT-Bund warned about a campaign that had been targeting users in Germany with Gootkit banking Trojans and REvil (Sodinokibi) ransomware. The milestone attack abusing the SolarWinds Orion update process will probably eclipse WannaCry as the most costly cyberattack. The 18,000 SolarWinds customers exposed to the first stage Sunburst malware will be threat hunting to determine if they were among the priority targets for the attackers. Microsoft identified more than 40 customers that were "targeted more precisely and compromised through additional and sophisticated measures." There were probably at least two different threat actors inside SolarWinds' network. One was the APT-grade actor discovered by FireEye. Another less-sophisticated actor was spreading SUPERNOVA backdoors. The APT actor prioritized a much smaller set of customers for reinforcing attacks using Teardrop dropper Trojans to deliver a Cobalt Strike Beacon. These priority victims probably number in the low hundreds and are being identified by unravelling Sunburst's network use for Command and Control and malware distribution.



# 08 Appendices

# Appendix A: Methodology

**One of the things readers value most about this report is the level of rigor and integrity we employ when collecting, analyzing and presenting data.**

Knowing our readership cares about such things and consumes this information with a keen eye helps keep us honest. Detailing our methods is an important part of that honesty.

First, we make mistakes. A column transposed here; a number not updated there. We're likely to discover a few things to fix. When we do, we'll list them on our corrections page: verizon.com/business/resources/reports/dbir/2021/report-corrections/

Second, we check our work. The same way the data behind the DBIR figures can be found in our GitHub repository,[78] as with last year, we're also publishing our fact check report there as well. It's highly technical, but for those interested, we've attempted to test every fact in the report.[79]

Third, François Jacob described "day science" and "night science."[80] Day science is hypothesis driven while night science is creative exploration. The DBIR is squarely night science. As Yanai et al. demonstrate, focusing too much on day science can cause you to miss the gorilla in the data.[81] While we may not be perfect, we believe we provide the best obtainable version of the truth[82] (to a given level of confidence and under the influence of biases acknowledged below).

However, proving causality is best left to the controlled experiments of day science. The best we can do is correlation. And while correlation is not causation, they are often related to some extent, and often useful.

## Non-committal disclaimer

We would like to reiterate that we make no claim that the findings of this report are representative of all data breaches in all organizations at all times. Even though the combined records from all our contributors more closely reflect reality than any of them in isolation, it is still a sample. And although we believe many of the findings presented in this report to be appropriate for generalization (and our confidence in this grows as we gather more data and compare it to that of others), bias undoubtedly exists.

## The DBIR process

Our overall process remains intact and largely unchanged from previous years. All incidents included in this report were reviewed and converted (if necessary) into the VERIS framework to create a common, anonymous aggregate data set. If you are unfamiliar with the VERIS framework, it is short for Vocabulary for Event Recording and Incident Sharing, it is free to use, and links to VERIS resources are at the beginning of this report.

The collection method and conversion techniques differed between contributors. In general, three basic methods (expounded below) were used to accomplish this:

1. Direct recording of paid external forensic investigations and related intelligence operations conducted by Verizon using the VERIS Webapp
2. Direct recording by partners using VERIS
3. Converting partners' existing schema into VERIS

All contributors received instruction to omit any information that might identify organizations or individuals involved.

Some source spreadsheets are converted to our standard spreadsheet formatted through automated mapping to ensure consistent conversion. Reviewed spreadsheets and VERIS Webapp JavaScript Object Notation (JSON) are ingested by an automated workflow that converts the incidents and breaches within into the VERIS JSON format as necessary, adds missing enumerations, and then validates the record against business logic and the VERIS schema. The automated workflow subsets the data and analyzes the results. Based on the results of this exploratory analysis, the validation logs from the workflow, and discussions with the partners providing the data, the data is cleaned and re-analyzed. This process runs nightly for roughly two months as data is collected and analyzed.

78 https://github.com/vz-risk/dbir/tree/gh-pages

79 Interested in how we test them? Check out Chapter 9, Hypothesis Testing, of ModernDive: https://moderndive.com/9-hypothesis-testing.html

80 Jacob F. The Statue Within: An Autobiography. CSHL Press; 1995. By way of Selective attention in hypothesis-driven data analysis, Itai Yanai, Martin Lercher, bioRxiv 2020.07.30.228916;

81 Really. They made printing the data print a gorilla and people trying to test hypotheses completely missed it

82 Eric Black, "Carl Bernstein Makes the Case for 'the Best Obtainable Version of the Truth,'" by way of Alberto Cairo, "How Charts Lie" (a good book you should probably read regardless).

## Incident data

Our data is non-exclusively multinomial, meaning a single feature, such as "Action," can have multiple values (i.e., "Social," "Malware" and "Hacking"). This means that percentages do not necessarily add up to 100%. For example, if there are five botnet breaches, the sample size is five. However, since each botnet used phishing, installed keyloggers, and used stolen credentials, there would be five Social actions, five Hacking actions, and five Malware actions, adding up to 300%. This is normal, expected and handled correctly in our analysis and tooling.

Another important point is that when looking at the findings, "Unknown" is equivalent to "Unmeasured." Which is to say that if a record (or collection of records) contains elements that have been marked as "Unknown" (whether it is something as basic as the number of records involved in the incident, or as complex as what specific capabilities a piece of malware contained), it means that we cannot make statements about that particular element as it stands in the record—we cannot measure where we have too little information. Because they are "unmeasured," they are not counted in sample sizes. The enumeration "Other," however, is counted, as it means the value was known but not part of VERIS. Finally, "Not Applicable" (normally "NA") may be counted or not counted depending on the claim being analyzed.

This year we again made use of confidence intervals to allow us to analyze smaller sample sizes. We adopted a few rules to help minimize bias in reading such data. Here we define "small sample" as less than 30 samples.

1. Sample sizes smaller than five are too small to analyze.
2. We won't discuss count or percentage for small samples. This applies to figures, too, and is why some figures lack the dot for the point estimate.
3. For small samples we may talk about the value being in some range, or values being greater/less than each other. These all follow the confidence interval approaches listed above.

## Incident eligibility

For a potential entry to be eligible for the incident/breach corpus, a couple of requirements must be met. The entry must be a confirmed security incident defined as a loss of confidentiality, integrity or availability. In addition to meeting the baseline definition of "security incident" the entry is assessed for quality. We create a subset of incidents (more on subsets later) that pass our quality filter. The details of what is a "quality" incident are:

1. The incident must have at least seven enumerations (e.g., threat actor variety, threat action category, variety of integrity loss, etc.) across 34 fields OR be a DDoS attack. Exceptions are given to confirmed data breaches with less than seven enumerations.
2. The incident must have at least one known VERIS threat action category (Hacking, Malware, etc.)

In addition to having the level of details necessary to pass the quality filter, the incident must be within the timeframe of analysis, (November 1, 2019, to October 31, 2020, for this report). The 2020 caseload is the primary analytical focus of the report, but the entire range of data is referenced throughout, notably in trending graphs. We also exclude incidents and breaches affecting individuals that cannot be tied to an organizational attribute loss. If your friend's laptop was hit with Trickbot it would not be included in this report.

Lastly, for something to be eligible for inclusion in the DBIR, we have to know about it, which brings us to several potential biases we will discuss on the next page.

Breaches


**Figure 133.** Individual contributions per action

Breaches

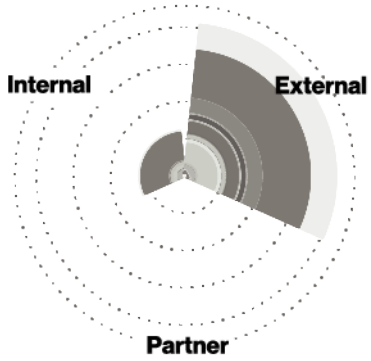



**Figure 134.** Individual contributions per actor

Breaches

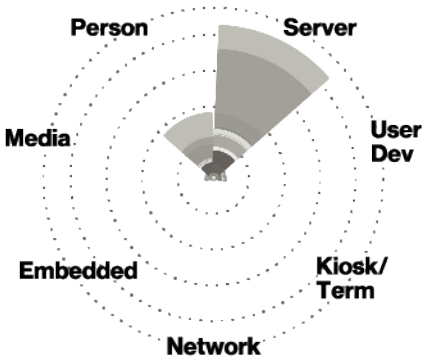



**Figure 135.** Individual contributions per asset

Breaches

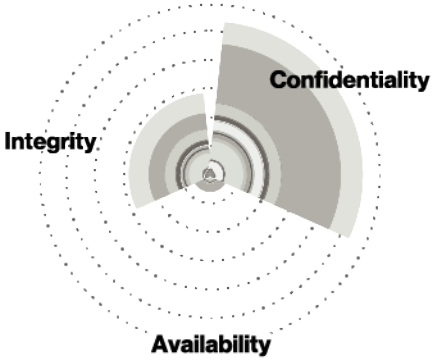



**Figure 136.** Individual contributions per attribute

## Acknowledgement and analysis of bias

Many breaches go unreported (though our sample does contain many of those). Many more are as yet unknown by the victim (and thereby unknown to us). Therefore, until we (or someone) can conduct an exhaustive census of every breach that happens in the entire world each year (our study population), we must use sampling.[83] Unfortunately, this process introduces bias.

The first type of bias is random bias introduced by sampling. This year, our maximum confidence is +/- 0.6%[84] for incidents and +/- 1.5% for breaches, which is related to our sample size. Any subset with a smaller sample size is going to have a wider confidence margin. We've expressed this confidence in the conditional probability bar charts (the "slanted" bar charts) we have been using since the 2019 report.

The second source of bias is sampling bias. Still, it is clear that we conduct biased sampling. For instance, some breaches, such as those publicly disclosed, are more likely to enter our corpus, while others, such as classified breaches, are less likely.

Figures 133, 134, 135 and 136 are an attempt to visualize potential sampling bias. Each radial axis is a VERIS enumeration, and we have ribbon charts representing our data contributors. Ideally, we want the distribution of sources to be roughly equal on the stacked bar charts along all axes. Axes only represented by a single source are more likely to be biased. However, contributions are inherently thick tailed, with a few contributors providing a lot

83 Interested in sampling? Check out Chapter 7, Sampling, of ModernDive: https://moderndive.com/7-sampling.html

84 This and all confidence intervals are 95% confidence intervals determined through bootstrap simulation or Markov Chain Monte Carlo. Read more in Chapter 8, Bootstrapping and Confidence Intervals, of ModernDive: https://moderndive.com/8-confidence-intervals.html

of data and many contributors providing a few records within a certain area. Still, we mostly see that most axes have multiple large contributors with small contributors adding appreciably to the total incidents along that axis.

You'll notice rather large contributions on many of the axes. While we'd generally be concerned about this, they represent contributions aggregating several other sources, so not actual single contributions. It also occurs along most axes, limiting the bias introduced by that grouping of indirect contributors.

The third source of bias is confirmation bias. Because we use our entire dataset for exploratory analysis (night science), we do not test specific hypotheses (day science). Until we develop a good collection method for data breaches or incidents from Earth-616 or any of the other Earths in the multiverse, this is probably the best that can be done.

As stated, we attempt to mitigate these biases by collecting data from diverse contributors. We follow a consistent multiple-review process and when we hear hooves, we think horse, not zebra.

## Data subsets

We already mentioned the subset of incidents that passed our quality requirements, but as part of our analysis there are other instances where we define subsets of data. These subsets consist of legitimate incidents that would eclipse smaller trends if left in. These are removed and analyzed separately (as called out in the relevant sections). This year we have two subsets of legitimate incidents that are not analyzed as part of the overall corpus:

1. We separately analyzed a subset of web servers that were identified as secondary targets (such as taking over a website to spread malware).
2. We separately analyzed botnet-related incidents.

Finally, we create some subsets to help further our analysis. In particular, a single subset is used for all analysis within the DBIR unless otherwise stated. It includes only quality incidents as described above and excludes the aforementioned two subsets.

## Non-incident data

Since the 2015 issue, the DBIR includes data that requires the analysis that did not fit into our usual categories of "incident" or "breach." Examples of non-incident data include malware, patching, phishing, DDoS, and other types of data. The sample sizes for non-incident data tend to be much larger than the incident data, but from fewer sources. We make every effort to normalize the data (for example weighting records by the number contributed from the organization so all organizations are represented equally). We also attempt to combine multiple contributors with similar data to conduct the analysis wherever possible. Once analysis is complete, we try to discuss our findings with the relevant contributor or contributors so as to validate it against their knowledge of the data.

# Appendix B: Controls

| | |
|---|---|
| **1** | Inventory and Control of Enterprise Assets |
| **2** | Inventory and Control of Software Assets |
| **3** | Data Protection |
| **4** | Secure Configuration of Enterprise Assets and Software |
| **5** | Account Management |
| **6** | Access Control Management |
| **7** | Continuous Vulnerability Management |
| **8** | Audit Log Management |
| **9** | Email and Web Browser Protections |
| **10** | Malware Defenses |
| **11** | Data Recovery |
| **12** | Network Infrastructure Management |
| **13** | Network Monitoring and Defense |
| **14** | Security Awareness and Skills Training |
| **15** | Service Provider Management |
| **16** | Application Software Security |
| **17** | Incident Response Management |
| **18** | Penetration Testing |

### Hopefully you didn't think we had forgotten about this important and helpful section?

Never fear, back by popular demand from auditors, CISOs and control freaks in general, we're updating our mapping with the community-built CIS Controls.[85] If you haven't heard, they have gone through a major update for their eighth iteration, much like our patterns have this year, and have been creatively named CIS Controls v8. Fortunately, there's no "should've had a V8" of the Controls mapping to VERIS, because we've got you covered.

The CIS Controls are a community-built, maintained and supported series of best practices targeted at helping organizations prioritize their defenses based on what attackers are doing—the so-called "Offense informs Defense" approach to best practices. The DBIR is but one resource of attacker knowledge at the macro level. Nevertheless, we were fortunate enough to be in a position to provide feedback and suggest input into their community process. Whether you are presenting your NIST Cybersecurity Framework (CSF) strategic roadmap at the Board level or defending an individual funding request for a new security program initiative, our goal is to allow you to easily tie our findings and data to your organization's efforts. We are thrilled to witness the evolution of the best practices due to the hard work of the individuals that donated their valuable time to help. Here is an overview of what has changed:

- Incorporating technologies such as cloud and mobile
- In recognition of "borderless" networks and tighter coordination between network/system administrators, the Controls are organized by activity, resulting in reducing the number of Controls from 20 to 18

85 https://www.cisecurity.org/controls/

- Reordering of Controls to show the importance of Data Protection (formerly 13, now 3)
- Addition of a "Service Provider Management" Control to address how organizations should manage cloud services

One of the more helpful components that the CIS community has decided to continue from version 7 are the Implementation Groups (IG), which help organizations further prioritize their implementation of Controls based on their resources, risk and other factors. The notion being that while every organization needs security, the giant, international leader on ethical pharmaceutical practices Umbrella Corp probably needs a larger and different set to protect its research facilities in Raccoon City than does the local pet hotel. The IGs build on each other, with Implementation Group 1 being the starting point where a smaller subset of the Controls are implemented (approximately 36%), and then building all the way up to Implementation Group 3, where all 153 safeguards are implemented.

Figure 137 breaks out the mapping into more granular detail and shows the relationships between the patterns and the overlap with the CIS Control for each Implementation Group.

In the report, you have hopefully noticed the addition of the Top Protective Implementation Group 1 Controls listed for each industry. By using the combination of the mappings to patterns, implementation groups and security functions of the Controls, we identified the core set of Controls that every organization should consider implementing regardless of size and budget:

**Control 4: Secure Configuration of Enterprise Assets and Software**

This control is not only a mouthful, but it also contains safeguards focused on engineering solutions that are secure from the outset, rather than tacking them on later. In this Control you will see substantial benefit toward reducing Error-based breaches like Misconfiguration and Loss of assets through enforcing remote wipe abilities on portable devices.

**Control 5: Account Management**

While this is technically a new Control in version 8, it should be extremely familiar as the safeguards are really just a centralization of the previous account management practices that were found in a few previous Controls, like Boundary Protect and Account Monitoring and Control. This control is very much targeted toward helping organizations manage the access to accounts and is useful against brute forcing and credential stuffing attacks.

**Control 6: Access Control Management**

This is Control 5's little cousin in which instead of simply looking at the user accounts and managing access to those, you're managing the rights and privileges and lastly enforcing multifactor authentication on key components of the environment, a useful tactic against Use of stolen credentials.

**Control 14: Security Awareness and Skills Training**

This control is a classic and hopefully doesn't need a whole lot of explanation. Considering the high prevalence of Errors and Social Engineering, it is obvious that awareness and technical training are probably a smart place to put some dollars to help support your team against a world full of cognitive hazards.

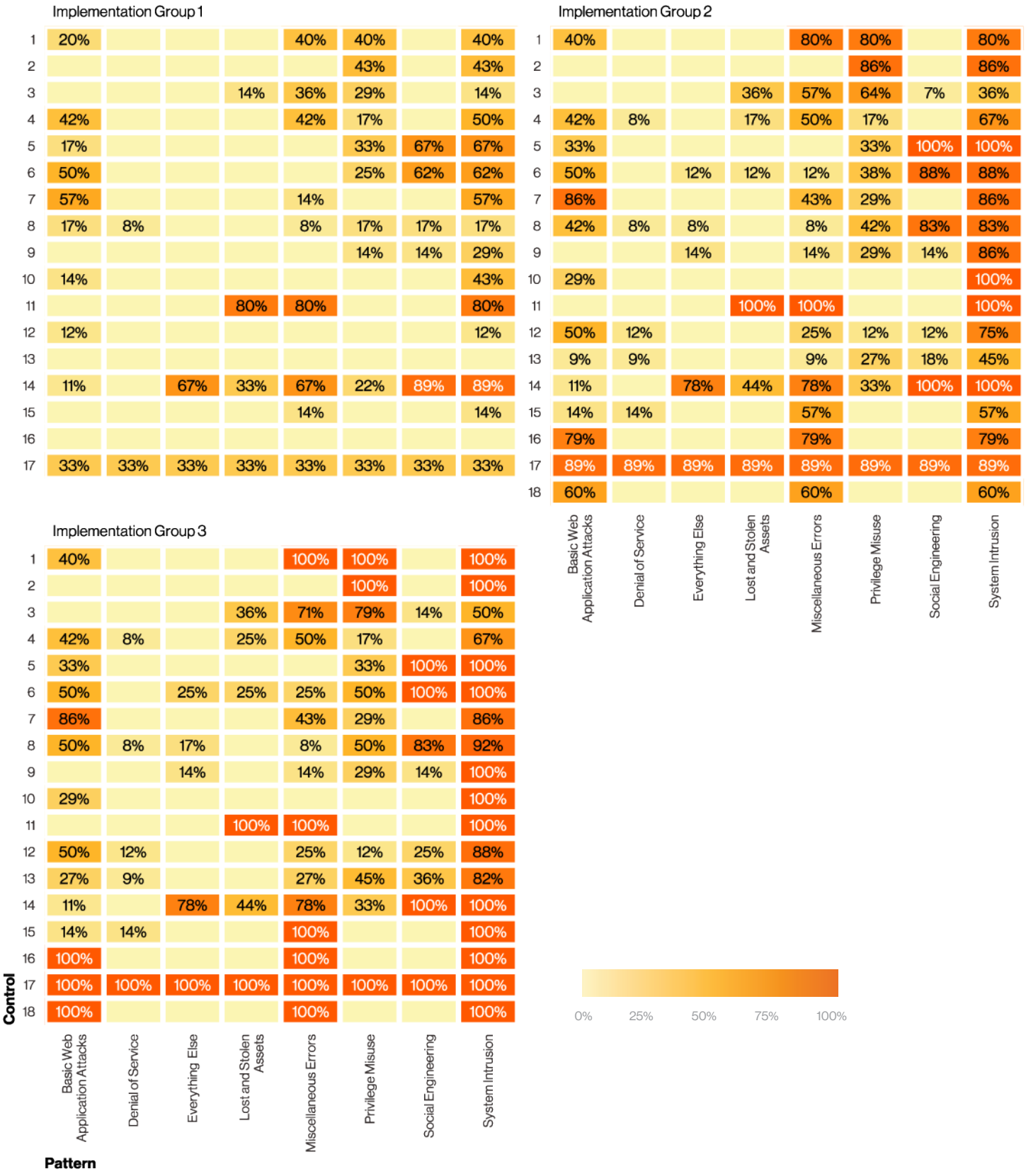



**Figure 137.** CIS to pattern mapping

# Appendix C: U.S. Secret Service

**David Smith**

**Special Agent in Charge**
**Criminal Investigative Division**
**U.S. Secret Service**

**Bernard Wilson**

**Network Intrusion Response Program Manager**
**Criminal Investigative Division**
**U.S. Secret Service**

## Protecting the Financial Infrastructure Amidst a Global Pandemic

The year 2020 will be remembered as the year of the COVID-19 global pandemic, with its short and long-term impacts. The pandemic began with lockdowns and a rapid transition to remote work, and continued with economic slowdowns and associated relief efforts. The pandemic affected all aspects of life and was particularly conducive to cybercrime.

In a matter of weeks, organizations had to transition to remote work, where possible. The reliance of a vastly expanded remote workforce resulted in a surge in the number and severity of attacks related to the weaknesses in underlying Internet and information technology infrastructure. This led to an increase in the number of incidents associated with the telework portion of the Business Continuity Plan (BCP) for many organizations. BCPs generally contain provisions for remote access to services available on an organization's network, a proliferation in email traffic for internal communications, and an increased reliance on enterprise video and audio communications. With this shift came an increase in malware and social engineering attacks, consistent with the exploitation of general communications.

Organizations that neglected to implement multi-factor authentication, along with virtual private networks (VPN), represented a significant percentage of victims targeted during the pandemic. The zero-trust model for access quickly became a fundamental security requirement rather than a future ideal. Nonrepudiation via Personal Identity Verification (PIV), Fast Identity Online (FIDO) or similar solutions became essential in zero-trust architectures. Security postures and principles, such as proper network segmentation, the prevention of lateral movement, least privilege, and "never trust, always verify" have proven to be strong indicators of an organization's ability to prevent or recover from unauthorized presence in its network environment.

In 2020, in the midst of the pandemic, cyber actors increased malware attacks against U.S. victims, including the healthcare and public health sector. The U.S. Secret Service noted a marked uptick in the number of ransomware attacks, ranging from small dollar to multi-million dollar ransom demands. While most organizations had adequate data backup solutions to mitigate these attacks, cyber actors shifted their focus to the exfiltration of sensitive data. These cyber actors, often organized criminal groups, proceeded to monetize the theft by threatening to publicize the data unless additional ransom was paid. The monetization of proceeds was typically enabled by cryptocurrency, in an attempt to obfuscate the destination of proceeds and hamper the ability of law enforcement to locate and apprehend those responsible for the crime.

**Preventing and deterring pandemic relief fraud became the focus of the Secret Service and other law enforcement agencies, particularly focused on Federal funding allocated to states for unemployment benefit programs.**

One of the primary responsibilities of the Secret Service is to protect the financial infrastructure of the United States. The pandemic required an unprecedented response from the Federal government. Legislators approved the release of $2.6 trillion of taxpayer funds to address the economic effects of the pandemic on the nation. The release of federal funding attracted the attention of organized criminal groups and individuals attempting to exploit pandemic relief programs. As a result, preventing and deterring pandemic relief fraud became the focus of the Secret Service and other law enforcement agencies, particularly focused on Federal funding allocated to states for unemployment benefit programs. The Secret Service worked with law enforcement partners at the U.S. Department of Labor to prevent criminal activity and arrest those responsible for exploiting the programs. This effort prevented more than $1.5 billion from reaching criminals and ensured that hundreds of millions of dollars intended to provide support to affected communities was returned to the states and the intended recipients.

Yet in spite of these efforts, criminals continued attempting to divert pandemic relief funds from different programs, to include $697.3 billion in loans intended to support businesses. The Secret Service and partner law enforcement agencies have expanded our efforts to prevent and mitigate these crimes, and ultimately locate and arrest those responsible.

The year 2020 demonstrated, once again, the enduring threat posed by organized cyber-criminal groups. Whether the crime involves a hospital ransomware attack, the sale of exfiltrated customer data, ATM cash-out attacks, or the theft of pandemic relief funds, the common indicator is the prevalence of organized crime. Criminals can be either formally or informally organized, at times in partnership with nation-state malicious actors, based on a common interest in illicit profit. Cyber actors quickly shift their activity based on emerging opportunities to steal and launder funds using any tactics, techniques and procedures available to them. Collaboration between domestic and foreign law enforcement partners to combat cybercriminal groups and their schemes is key to dismantling organized crime and apprehending cyber actors.

To address this continued shift of criminality, the Secret Service operates a network of Cyber Fraud Task Forces (CFTF), a partnership of federal, state, local, and foreign law enforcement agencies, prosecutors, the private sector, and academia. Outreach is at the core of the Secret Service CFTFs, as it fosters trusted relationships and information sharing, which are important tools in mitigating cybercrimes. While apprehending criminals is, and will continue to be, the ultimate goal of the Secret Service, prevention and mitigation are equally critical in the protection of the U.S. financial infrastructure.

# Appendix D: Contributing organizations

**A**

Akamai Technologies
Ankura
Apura Cybersecurity Intelligence
Arics Cooper
Atos (Paladion)
AttackIQ

**B**

Bad Packets
BeyondTrust
Bit Discovery
Bit-x-bit
BitSight
BlackBerry Cylance

**C**

Center for Internet Security
CERT European Union
CERT National Insider Threat Center
CERT Polska
Checkpoint Software Technologies Ltd.
Chubb
Cisco Talos Incident Response
Coalition
Computer Incident Response Center Luxembourg (CIRCL)
CrowdStrike
Cybersecurity and Infrastructure Security Agency (CISA)
CyberSecurity Malaysia, an agency under the Ministry of Communications and Multimedia (KKMM)
Cybir (formerly DFDR Forensics)

**D**

Dell
Digital Shadows
Dragos, Inc

**E**

Edgescan
Elevate Security
Emergence Insurance
EUROCONTROL

**F**

Farsight Security
Federal Bureau of Investigation - Internet Crime Complaint Center (FBI IC3)
F-Secure

**G**

Global Resilience Federation
Government of Telangana, ITE&C Dept.
Government of Victoria, Australia - Department of Premier and Cabinet (VIC)
Grey Noise

**H**

Hasso-Plattner Institut
Homeland Security Solutions B. V (HLSS)

**I**

ICSA Labs
Irish Reporting and Information Security Service (IRISS-CERT)

**J**

JPCERT/CC

**K**

Kaspersky
KnowBe4

**L**

Lares Consulting
Legal Services - ISAO
LMG Security

**M**

Malicious Streams
Maritime Transportation System ISAC (MTS-ISAC)
Micro Focus
Mishcon de Reya
mnemonic

**N**

National Cybersecurity & Communications Integration Center (NCCIC)
NetDiligence®
NETSCOUT

**P**

ParaFlare Pty Ltd
Proofpoint
PSafe

**Q**

Qualys

**R**

Rapid7

Recorded Future

**S**

S21sec

SecurityTrails

Shadowserver Foundation

Shodan

SISAP - Sistemas Aplicativos

Swisscom

**T**

Tetra Defense

**U**

U.S. Secret Service

**V**

VERIS Community Database

Verizon Cyber Risk Programs

Verizon DDoS Shield

Verizon Digital Media Services

Verizon Managed Security Services - Analytics (MSS-A)

Verizon Network Operations and Engineering

Verizon Professional Services

Vestige Digital Investigations

VMRay

Verizon Threat Research Advisory Center (VTRAC)

**W**

WatchGuard Technologies

**Z**

Zscaler


ankura
COLLABORATION DRIVES RESULTS

Akamai

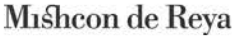
CROWDSTRIKE

emergence

apura

Malicious Streams

GOVERNMENT OF TELANGANA

CERT-EU
FOR THE EU INSTITUTIONS, BODIES AND AGENCIES

Mishcon de Reya

VICTORIA
Australia

LMG
SECURITY

digital shadows_

GREYNOISE

AricsCooper

CHUBB

kaspersky

ICSA labs

edgescan

bit-x-bit

IRISS
Irish Reporting and Information Security Service

F-Secure

CyberSecurity
MALAYSIA

KnowBe4
Human error. Conquered.

EUROCONTROL

ATTACKIQ

NASK
CERT.PL

DELL Technologies

CIRCL
Computer Incident
Response Center
Luxembourg

HOMELAND
SECURITY
SOLUTIONS BV

mnemonic

BeyondTrust

FORESIGHT
SECURITY

MICRO
FOCUS

JPCERT CC®

Coalition

FEDERAL BUREAU OF INVESTIGATION
IC3
INTERNET CRIME COMPLAINT CENTER

BIT
DISCOVERY

Check Point
SOFTWARE TECHNOLOGIES LTD

HPI
Hasso
Plattner
Institut
Digital Engineering · Universität Potsdam

CYBIR


LARES

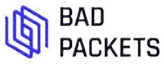

WatchGuard


SecurityTrails


SHODAN


S21
SEC


BITSIGHT


BAD
PACKETS


RAPID7


CIS
Center for
Internet
Security


Qualys.


BlackBerry.
CYLANCE


zscaler


swisscom


VESTIGE
Digital Investigations


SHADOWSERVER


TALOS
INCIDENT
RESPONSE


proofpoint.
Security Awareness Training


Elevate Security


Atos


VMRAY


V
C
D
B


TETRA
DEFENSE


NETSCOUT.


SISAP
Sistemas Aplicativos


DRAGOS


LEGAL
SERVICES
ISAO


GLOBAL
RESILIENCE
FEDERATION


NetDiligence


NETSCOUT.


Carnegie
Mellon
University
Software
Engineering
Institute


PARA
FLARE


Recorded Future

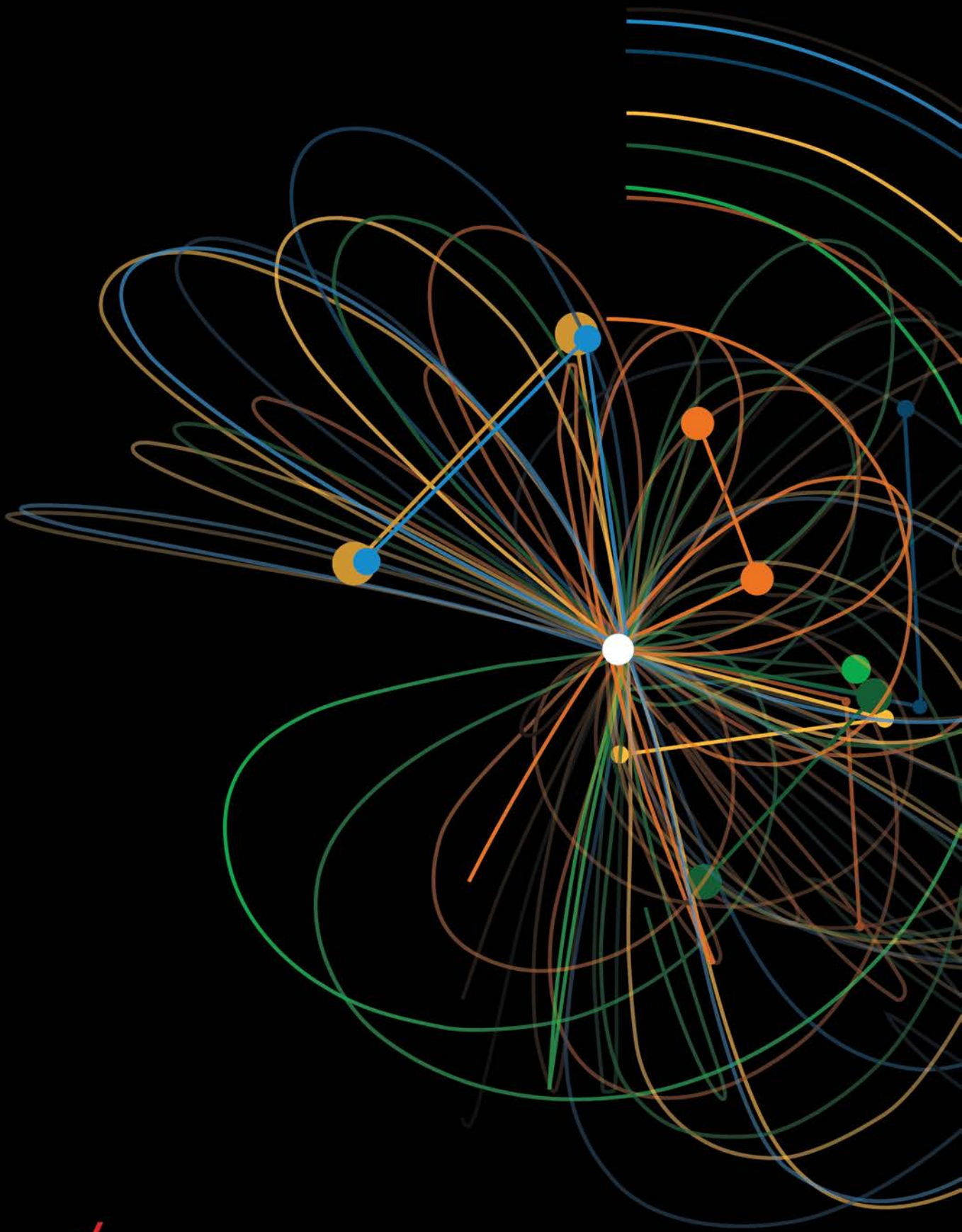

NCCIC


CISA


PSafe
dfndr


Maritime Transportation System
Information Sharing and Analysis Center


Verizon Threat
Research Advisory
Center (VTRAC)



verizon