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 that 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. In order to continue to increase the transparency of our work, we introduced a couple of new features we are including in the report this year.

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: https://enterprise.verizon.com/resources/reports/dbir/2020/report-corrections/

Second, we check our work. The same way the data behind the DBIR figures can be found in our GitHub repository, for the first time 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.

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.

While we may not be perfect, we believe we provide the best obtainable version of the truth and a useful one at that. Please review the “Acknowledgement and analysis of bias” section below for more details on how we do that.

The DBIR process

Our overall process remains intact and largely unchanged from previous years. All incidents included in this report were individually reviewed and converted (if necessary) into the VERIS framework to create a common, anonymous aggregate dataset. 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 that 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 contributors using VERIS
3. Converting contributors’ existing schema into VERIS

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

Reviewed spreadsheets and VERIS WebApp JavaScript Object Notation (JSON) are ingested by an automated workflow that converts the incidents and breaches 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 contributors providing the data, the data is cleaned and reanalyzed. This process runs nightly for roughly three months as data is collected and analyzed.

56 Interested in how we test them? Check out Chapter 9, Hypothesis Testing, of ModernDive: https://moderndive.com/9-hypothesis-testing.html
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) contain 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 no information. Because they are “unmeasured,” they are not counted in sample sizes. The enumeration “Other” is, however, counted as it means the value was known but not part of VERIS or not included, as is the case with “top” figures. Finally, “Not Applicable,” (normally “NA”), may be counted or not counted depending on the hypothesis.

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. This year, we have made liberal use of confidence intervals to allow us to analyze smaller sample sizes. We have 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 talk about count or percentage for small samples. This goes for figures too and is why some figures lack the dot for the median frequency

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 hypothesis testing and confidence interval approaches listed above

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, et al.) 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 time frame of analysis (November 1, 2018, to October 31, 2019, for this report). The 2019 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 into the DBIR, we have to know about it, which brings us to several potential biases we will discuss below.

57 Our line figures use the calendar year the incident occurred in as they are continuous, while our dumbbell charts use the year of the DBIR report, as they are ordinal.
Acknowledgement and analysis of bias

Many breaches go unreported (though not in our sample). 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. 58 Unfortunately, this process introduces bias.

The first type of bias is random bias introduced by sampling. This year, our maximum confidence is +/-1.5% 59 for breaches and +/-0.5% for incidents, 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) that we have been using since the 2019 report.

The second source of bias is sampling bias. We strive for “the best obtainable version of the truth” 60 by collecting breaches from a wide variety of contributors. 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.

---

58 Interested in sampling? Check out Chapter 7, Sampling, of ModernDive: https://moderndive.com/7-sampling.html
59 This and all confidence intervals are 95% confidence intervals determined through bootstrap simulation. Read more in Chapter 8, Bootstrapping and Confidence Intervals, of ModernDive: https://moderndive.com/8-confidence-intervals.html
60 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)
As stated above, 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. When we hear hooves, we think horse, not zebra. We also try to review findings with subject matter experts in the specific areas ahead of release.

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

Both subsets were separately analyzed the last three years as well.

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 earlier 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 partners with similar data to conduct the analysis wherever possible. Once analysis is complete, we try to discuss our findings with the relevant partner or partners so as to validate it against their knowledge of the data.

---

61 The DBIR is a pre-Crisis on Infinite Earths work environment.
62 A unique finding is more likely to be something mundane (such as a data collection issue) than an unexpected result.
Appendix B: VERIS Common Attack Framework (VCAF)

VERIS was developed as a solution to the need for consistent definitions of incident and breach data for analysis.

With its close ties to the DBIR and data analysis, it was created to remove the ambiguity inherent in terms surrounding breaches and provide a data-driven structure capable of quantifying the majority of breaches. While VERIS covers a lot of different detailed information about an incident, including things such as Victim demographics and Timeline, the core of VERIS is captured in what we call the four “As” of an incident: Actor, Action, Asset, Attribute.

However, VERIS was not designed to represent precise and detailed tactical and technical minutiae around attackers’ techniques, chosen methods of persistence or methodology for executing malicious code on a compromised asset. Thankfully, it doesn’t need to because there is something else that has come along to help address that need.

Massive (adoption of) ATT&CK

MITRE privately developed the original Adversarial Tactics, Techniques and Common Knowledge (ATT&CK) framework starting in 2013 as a means of codifying adversarial behavior and released it publicly in 2015. ATT&CK has become a well-established way for describing the tactical actions used by attackers (including a heavy focus on advanced threats). Much like VERIS, ATT&CK is subdivided into a handful of key components, but the core of the framework are the “Techniques,” which describe the atomic means of how an attacker achieves an objective called a “Tactic.” The 260+ Techniques in ATT&CK for Enterprise are logically grouped with their corresponding 11 Tactics, which describe the different objectives an adversary might take as part of their intrusion.

VCAF serves as a bridge to ATT&CK, covering the portions of VERIS not in ATT&CK with the aim of creating a holistic framework. At its very core, VCAF is made of two components: one is the conceptual mapping between VERIS and ATT&CK, and another is the extension of ATT&CK with techniques that cover all possible Threat Actions present in VERIS. As much as we would have liked to leverage a default “meteor falling from the sky” technique in ATT&CK, those events are definitely quite rare.

This approach should be flexible enough to accommodate both general categories found in VERIS (such as Ransomware) and some of the more specific attack types found either in VERIS or ATT&CK. Aside from expanding the scope of what is covered and can be tracked, using VCAF can help provide essential context to these incidents. Below is a list that includes a variety of the different benefits of leveraging this powerful combination:

- Understand the technical details associated with an incident
- Prioritize mitigations based on previous all incident types (not just the malware or hacking kind)
- Better understand the junction of targeting and capabilities
- Capture incident context that goes beyond technical artifacts
- Ease communication of cybersecurity concepts with non-cybersecurity experts

What is this, a crossover episode?

Our solution to bridge the gap and help operationally connect the relationships between ATT&CK and VERIS is through the creation of an extension that we call the VERIS Common Attack Framework (VCAF).

We’re better when we’re together.

While both VERIS and ATT&CK grew out of different needs and different objectives, VERIS to codify incidents and ATT&CK to codify adversary technique, there is without a doubt an overlap between the two that could be leveraged to improve the value of both standards. To get a better understanding of the relationships between these two frameworks, the team spent some time researching to see if they could map the VERIS framework to the ATT&CK techniques and vice-versa, the results of which you can see in Figure 139.

64 But they sure have a large impact!
In this issue of the DBIR, we used VCAF to map simulated breach data, SIEM data and malware features to VERIS action categories to compare and draw conclusions in conjunction with our incident corpus.

### The beginning of something great

Clearly, VCAF is not the end-all be-all of cybersecurity frameworks. It is a modest step toward having an integrated way for the community to discuss security incidents and attackers. As the number of cybersecurity frameworks grows and the field of knowledge surrounding cybersecurity topics deepens, there is a need for us as a community to integrate our own languages and understanding in an effort to help us communicate to the larger community of non-cybersecurity experts. Keep your eyes peeled for future developments and information on VCAF by visiting our VERIS GitHub page at https://github.com/vz-risk/veris.

---

65 And don’t forget to smash that like and subscribe button!
Appendix C

Michael D’Ambrosio
Assistant Director
U.S. Secret Service

Jonah Force Hill
Senior Cyber Policy Advisor
U.S. Secret Service

Following the money—the key to nabbing the cybercriminal

This year’s DBIR has once again highlighted the principal motive for the vast majority of malicious data breaches: the pursuit of profit. This is surprising to some, given the extensive media coverage of national security-related breaches. However, it should not be. Most malicious cyber actors are not motivated by national security or geopolitical objectives, but rather by simple greed. Cybercriminals primarily profit through fraud and extortion. They target financial and payment systems, steal information to use in various fraud schemes, and hold IT systems hostage through ransomware and other means. Whatever their criminal scheme, they then depend upon a money movement and laundering apparatus to transfer and liquidate their proceeds.

That is why the U.S. Secret Service was first assigned responsibility for investigating cybercrimes in the early 1980s, before it was even called “cyber,” and why we continue to do so today. Secret Service agents are financial crimes investigators, skilled not only at “following the money,” but at preventing criminals from profiting from their activities and at recovering the stolen assets of victims. When investigating any criminal cyber incident, a data breach, an “unlimited ATM cash-out” conspiracy, a ransomware attack or any other diverse, financially motived crime committed via the internet, the heart of the Secret Service’s approach is following the money.

We have learned over the decades that it is through the movement of funds—from the victim to the criminal, between and among criminals, and through the process of money laundering—that investigators are able to generate the greatest insights and criminal leads. Malware samples and indicator sharing are useful, no doubt, but it is the money and where it moves that leads to arrests, asset seizures and the recovery of assets stolen from victims of fraud.

For example, in a typical business email compromise (BEC) scheme, a victim is lured into sending a payment, usually via a wire transfer, to a bank account maintained under a criminal’s control. The methods used in the deception part of the crime can range from highly sophisticated (such as deploying tailor-made malware) to shockingly simple (such as impersonating a vendor on the phone). How the fraudsters fool the victim is often insignificant; what is important is how they move and liquidate their proceeds.
Smart criminals understand this. They know that the accounts, shell companies and processes they use to move their stolen funds contain a wealth of location data and other information that can lead to their arrest. As a result, criminals try to distance themselves and their identities from all accounts and institutions that might be associated with their crimes.

There are number of ways criminals do this, but one of the principal mechanisms is the use of “mules,” outside individuals recruited to participate in the scheme. Mules can be either witting or unwitting participants. Some mules join the scheme with full knowledge of the criminal nature of their involvement; others are recruited through what appear to be legitimate job postings. Still others are victims themselves of ancillary frauds, often romance scams, in which they are conned into believing that they are sending money to a romantic partner, when in fact they are just moving money for crooks.

A similar dynamic exists in cases of ransomware and in other crimes in which cryptocurrencies play a role. When an organization pays a ransom to unlock its IT systems, for instance, the criminal generally instructs the victim to send a bitcoin payment to a cryptocurrency wallet. These wallets are hosted either on a cryptocurrency exchange, which can be either legitimate or illegitimate, or on a device operated by the criminal or an associate. Here too, the criminals seek to obscure the location of the wallets and to limit access to any other information that might tie their activities to a specific wallet or account.

Criminals engaged in ransomware attacks employ many of the same techniques as BEC scammers to cover their tracks. They may pay mules to set up crypto wallets, or con unwitting mules into thinking they have landed a legitimate job in the cryptocurrency industry. They may use cryptocurrency tumblers and mixers to swap funds from one form of cryptocurrency to another (for instance, from bitcoin to ether), to keep law enforcement from tracking their movements on the blockchain. They may set up shell companies, open overseas bank accounts and move money repeatedly from one country to the next, all with the aim of making their financial movements as difficult as possible to trace.

Yet there is always a chokepoint. If cybercriminals want to enjoy the fruit of their criminal labor, they must convert their profits into a form of money they can actually use, without being tracked by law enforcement. These chokepoints create the greatest opportunities to counter cybercriminal activity.

The Secret Service focuses on these chokepoints to disrupt these financial flows, whether they are explicitly illicit services or legitimate businesses that are exploited by criminals. Through undercover operations, confidential informants and partnerships with industry and the broader law enforcement community, the Secret Service excels at identifying and interdicting these illicit financial flows. In 2019, the Secret Service prevented $7.1 billion of cybercrime losses and returned over $31 million in stolen assets to victims of fraud.

The lessons for industry are simple: Invest in the defense of your networks and, in the event of a breach, collect as much evidence as you can. When shared with law enforcement partners, that evidence can lead not only to the arrest of the criminal, but also to the seizure of their assets. In many cases, the recovered money can be returned to the victim. This is how we prevent cybercriminals from operating with impunity. It is a collective struggle. Let’s work together.
Diego Curt
Chief Compliance Officer
State of Idaho, Office of the Governor—Information Technology Services

State of Idaho enhances incident response program with VERIS.

We hear it all the time. We need to share incident and breach information for improved decision-making. The State of Idaho was facing the same issue, trying to get different agencies to share incident and breach information for improved decision-making and better cyber-defense investment. In order to address this, the State of Idaho designed a program that gained approval from various stakeholders, including the legal department. The program consists of two fundamental components and three core components.

The two fundamental components are:

1. Cyber Kill Chain® developed by Lockheed Martin, Inc.—used to promote actionable intelligence-process thinking and serves as a blueprint for building an effective cybersecurity program

2. National Institute of Standards and Technology (NIST) Cybersecurity Framework—a risk reporting framework used to assess the readiness and maturity of cybersecurity controls throughout the enterprise

The three core components of the program are:

1. NIST SP 800-53 Incident Response Control Family—used to govern and ensure all control processes are addressed and matured on a continuous basis

2. Vocabulary for Event Recording and Incident Sharing (VERIS)—an easy-to-use, systematically structured language/taxonomy used to gather intelligence from incidents and breaches for better decision-making and information sharing

3. A commercial web-based application that brings together first responders, emergency management, National Guard, cyber-incident response handlers, etc., into one platform that houses the VERIS language/taxonomy

67 https://www.nist.gov/cyberframework
68 https://nvd.nist.gov/800-53
At the heart of the program is the VERIS taxonomy. VERIS is a language/taxonomy designed to help an organization hurdle over the issues many organizations are concerned about—sharing confidential data with outsiders. Without the capability to incorporate a common language (VERIS) designed to share incident information, the State of Idaho would never have been able to gain approval from various stakeholders (including the legal department) to share incident and breach information both internally (other agencies) and externally (DHS, FEMA, etc.).

Some of the areas in which VERIS has helped improve the State of Idaho’s ability to share information are:

- It has created awareness and interest that there is a better way to gather and use intelligence information from adverse events that we respond to from time to time
- It is an open source framework that works well with other incident response frameworks
- It is an easy-to-use full-schema taxonomy/language designed to be incorporated and implemented within a short period of time
- It provides a way for business executives to get involved with their organization’s cybersecurity efforts and simplifies intelligence gathering by repetitively asking four basic questions: Whose actions affected the asset? What actions affected the asset? Which asset was affected? How was the asset affected?

VERIS provides a solid language foundation that can be used to build a strong intelligence-driven incident response program. Couple that with other open source frameworks and you have one heck of an incident response program.
## Appendix E: Contributing organizations

**A**
- Akamai Technologies
- Avara Cyber Intelligence
- AttackIQ
- Australian Federal Police

**B**
- BeyondTrust
- Bit Discovery
- Bit-x-bit
- BitSight

**C**
- Center for Internet Security
- CERT European Union
- CERT Insider Threat Center
- CERT Polska
- Check Point Software Technologies Ltd.
- Chubb
- Cisco Talos Incident Response
- Coalition (formerly BinaryEdge)
- Computer Incident Response Center Luxembourg (CIRCL)
- CrowdStrike
- Cybercrime Central Unit of the Guardia Civil (Spain)
- CyberSecurity Malaysia, an agency under the Ministry of Science, Technology and Innovation (MOSTI)

**D**
- Defense Counterintelligence and Security Agency (DCSA)
- Dell (formerly EMC-CIRC)
- DFDR Forensics
- Digital Shadows
- Dragos, Inc.

**E**
- Edgescan
- Elevate Security
- Emergence Insurance

**F**
- F-Secure (formerly MWR InfoSecurity)
- Federal Bureau of Investigation—Internet Crime Complaint Center (FBiIC3)
- Financial Services Information Sharing and Analysis Center (FS-ISAC)

**G**
- Government of Telangana, ITE&C Dept., Secretariat
- Government of Victoria, Australia—Department of Premier and Cabinet (VIC)
- GreyNoise

**H**
- Hasso-Plattner Institut
- Hyderabad Security Cluster

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

**J**
- JPCERT/CC

**K**
- Kaspersky
- KnowBe4

**L**
- Lares Consulting
- LMG Security

**M**
- Malicious Streams
- Micro Focus (formerly Interset)
- Mishcon de Reya
- mnemonic
- Moss Adams (previously AsTech Consulting)

**N**
- National Cybersecurity and Communications Integration Center (NCCIC)
- NetDiligence
- NETSCOUT

**P**
- Paladion Networks Pvt Ltd.
- Palo Alto Networks
- ParaFlare Pty Ltd
- Proofpoint (formerly Wombat Security)

**Q**
- Qualys

**R**
- Rapid7
- Recorded Future

**S**
- S21sec
- SecurityTrails
- Shadowserver Foundation
- Shodan
- SISAP—Sistemas Aplicativos
- SwissCom

**T**
- Tetra Defense (formerly Gillware Digital Forensics)
- Tripwire

**U**
- United States Computer Emergency Readiness Team (US-CERT)
- 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
- Verizon Threat Research Advisory Center (VTRAC)
- Vestige, Ltd.
- VMRay

**W**
- Wandera
- WatchGuard Technologies

**Z**
- Zscaler