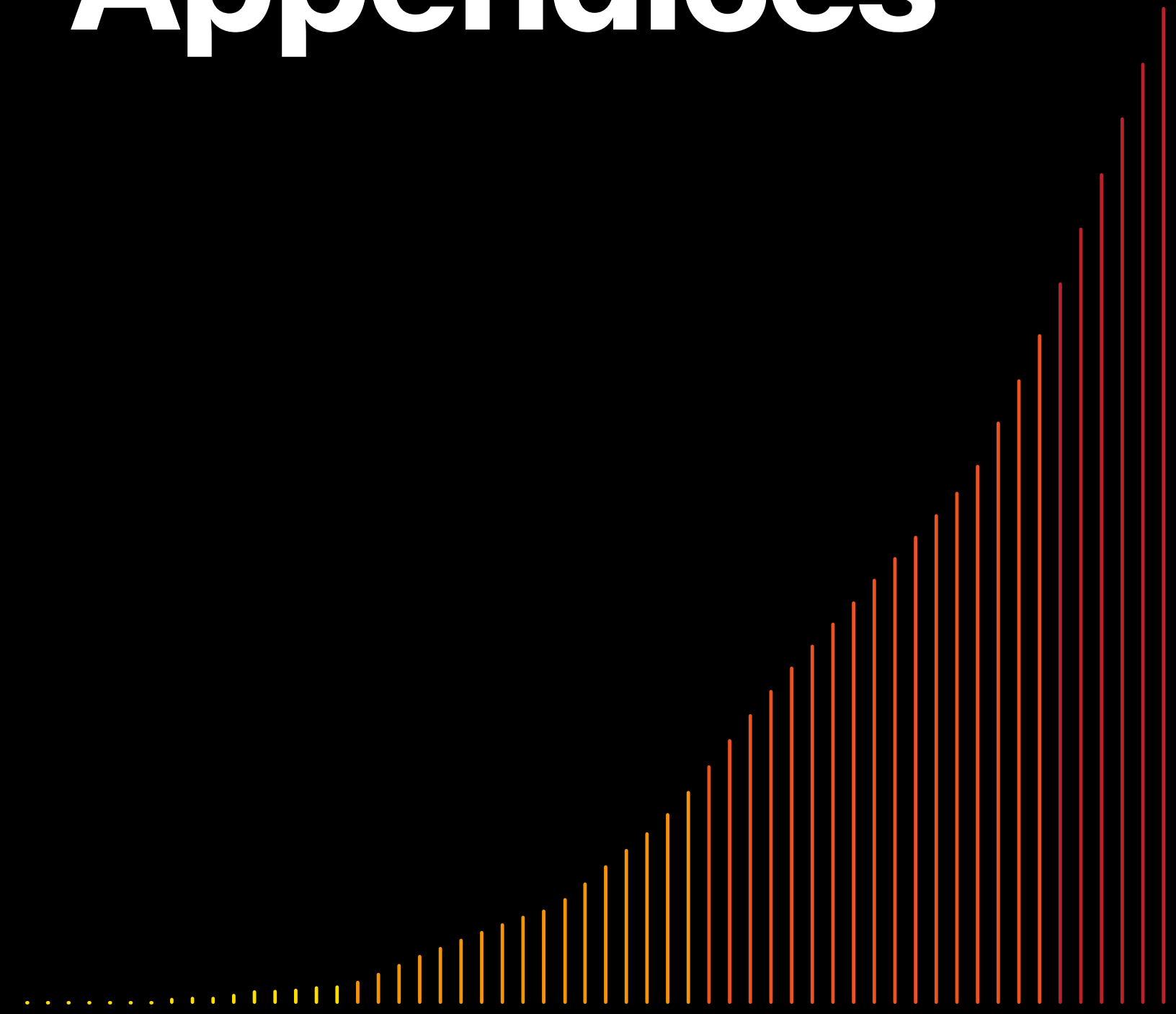
7 Appendices



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/2023/corrections/.

Second, we check our work. The same way the data behind the DBIR figures can be found in our GitHub repository,[62] 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.

Third, science comes in two flavors: creative exploration and causal hypothesis testing. The DBIR is squarely in the former. While not perfect, we believe we provide the best obtainable version of the truth (to a given level of confidence and under the influence of biases acknowledged below). However, proving causality is best left to randomized control trials. 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 must 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 we believe 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 conviction in this grows as we gather more data and compare it to that of others), bias exists.

The DBIR process

Our overall process remains intact and largely unchanged from previous years.[63] All incidents included in this report were 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 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 reanalyzed. This process runs nightly for roughly two months as data is collected and analyzed.

62 https://github.com/vz-risk/dbir/tree/gh-pages
63 As does this sentence

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 (or not one of the other bars if found in a bar chart). Finally, "Not Applicable," (normally "NA"), may be counted or not counted depending on the claim being analyzed.

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 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:

- 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.
- 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, 2021, to October 31, 2022, for this report). The 2022 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.

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. Unfortunately, this process introduces bias.

Breaches

Figure 64. Individual contributors per action

Breaches

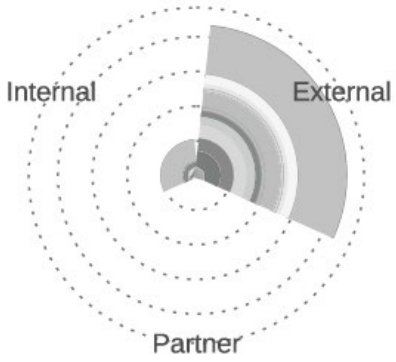


Figure 65. Individual contributors per actor

Breaches

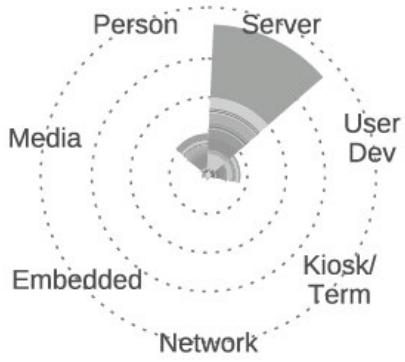


Figure 66. Individual contributors per asset

Breaches

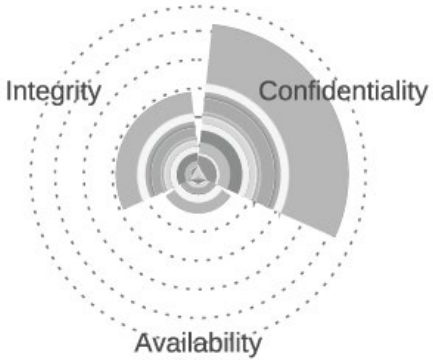


Figure 67. Individual contributors per attribute

The first type of bias is random bias introduced by sampling. This year, our maximum confidence is +/- 0.7% for incidents and +/- 1.4% 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 complementary cumulative density (slanted) bar charts, hypothetical outcome plot (spaghetti) line charts and quantile dot plots.

The second source of bias is sampling bias. We strive for "the best obtainable version of the truth" 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.

The four figures on the left are an attempt to visualize potential sampling bias. Each radial axis is a VERIS enumeration, and we have stacked bar 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 of data and a lot of 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 those axes.

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, 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, we cannot test specific hypotheses. Until we develop a collection method for data breaches beyond a sample of convenience, this is probably the best that can be done.

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 horses, not zebras.[64] 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, though may not be written about if no relevant findings were, well, found. 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 separated the last six 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 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 and DDoS. 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.

64 A unique finding is more likely to be something mundane, such as a data collection issue, than an unexpected result.

Appendix B: VERIS mappings to MITRE ATT&CK®

When it comes to sailing the stormy seas of the cybersecurity world, a map comes in handy to help you chart your direction. We consider the DBIR to be one of those maps, helping organizations navigate the complicated and ever-changing conditions of the cybersecurity landscape. To make sure this map is the most accurate possible, we have created the VERIS Framework,[65] which captures most of the important components of data breaches in order to facilitate risk-oriented decision making for our weary cyber mariners.

Over the years, new guiding frameworks have been created that provide different levels of detail, MITRE ATT&CK® being by far the most popular. We have worked with MITRE Engenuity and the Center for Threat Informed Defense[66] to capture the relationships between VERIS to ATT&CK so that organizations can leverage the benefits of both in their navigation.

The results of that work are remarkable: ATT&CK provides excellent tactical and technical details into the specific techniques the threat actors leverage, while VERIS provides a strategic view of the landscape, covering a wider range of possible mishaps. Errors, for instance, are present in 9% percent of breaches this year but are out of scope in ATT&CK. When VERIS and ATT&CK are combined, they provide you with a clearer view of what type of assets were impacted and what type of victims those assets belonged to while still preserving the specifics of the attack techniques that were leveraged.

This combination of forces is timely due to the increased regulatory pressure of reporting data breaches to governments, although there is no commonly accepted format in how this reporting should be done. We, of course, cannot opine on the need for such regulations,[67] but we would like to do our part to make sure that organizations have the right tooling to reduce their burden as new laws come to fruition.

The second version of this mapping has just been released as of April 6, 2023, and we are very excited about it. In addition to VERIS Actions, a lot of thought was put into mapping Attributes. To make it better, Actors were mapped to ATT&CK Groups.[68] There are also new mappings to ATT&CK for Mobile and ATT&CK for ICS.

If this interests you at all, please hop over to https://center-for-threat-informed-defense.github.io/attack_to_veris/ for all the details of the work. Even if it doesn't,[69] you are already reaping the benefits of the work thanks to the ATT&CK Technique mappings we have added to some select Incident Patterns to help you in your epic journey to "full control coverage."

Our team puts a lot of thought and energy into trying to make the VERIS Framework more accessible and helpful for all. If you are curious about the framework or have tried it in the past and want to check what's new, get in touch with the DBIR team at dbir@verizon.com.

65 https://verisframework.org/
66 https://mitre-engenuity.org/cybersecurity/center-for-threat-informed-defense/
67 Who are we kidding? We would love to have more data to analyze!
68 https://attack.mitre.org/groups/
69 How dare you?

Appendix C: VTRAC 20-year retrospective

By Chris Novak,
Managing Director
Cyber Security Consulting
Verizon

It's hard to believe that the Verizon Threat Research Advisory Center (VTRAC) is 20 years old! I've had the unique pleasure of being part of the team since the very beginning—or should I say the "zero-day"?

Over those 20 years, we've had a few different names but always the same passionate team behind the scenes. Back then, I was part of a small gaggle of geeks in New York City, always having a suitcase packed and ready to hop a flight to anywhere to take on the next big data breach investigation. Our forensic lab at the time was a collection of systems that didn't even fill a single full-height server rack.

It bears reminding that, 20 years ago, "cybersecurity" was not a commonly used or understood word. If you asked the average person what "cyber" was, you would probably get back responses that sounded like something from a science fiction movie. There was no such thing as a cybersecurity college degree—the closest thing that existed at the time was a computer science or engineering degree. Today, there are hundreds of universities around the world that not only offer cybersecurity bachelor's degrees, but also master's degrees and Ph.D.s.

I can still remember some of the first data breaches I ever investigated. Old timers will appreciate the days when we showed up onsite with our "medical bag"—typically a bag that had a binder of bootable floppy disks, a collection of assorted cables, and a variety of hard drives and enclosures. As mentioned above, hardly anyone knew what cybersecurity was back then, and the average person had no idea of the purpose of the equipment in that medical bag. In a world just following 9/11, going through airport security with that bag of odd-looking electronics and cables guaranteed that I was frequently the lucky winner of "random" extra screening. If only that luck carried over into a few of the trips to Vegas ...

Today, we rarely need to travel. We have enterprise tools that can facilitate remote forensic evidence collection from anywhere in the world. Taking advantage of our telecommunications backbone and advances in cellular connectivity, we're even able to provide immediate emergency and out-of-band communications via 5G, allowing us to collect forensic data at speeds in excess of 1 Gbps, even if the victim organization has its own network, systems or infrastructure outages.

The then and now comparisons over the last 20 years are staggering to consider. Today, we have exponentially more people on our team, with incredible diversity of backgrounds and geographic locations. The VTRAC supports organizations across more than 100 countries. We not only have several physical lab locations around the world but also cloud-based and client on-premises lab locations to care for nearly every conceivable data privacy and sovereignty concern.

Of course, I cannot forget to mention the incredible work of the DBIR team that makes this very publication possible. Many have heard me say that the DBIR is my third child. It was born 16 years ago as part of an early incarnation of VTRAC (back then we were called the RISK Team) with a vision of sharing our data breach insights with the world. Metaphorically, I heard it say its first words and watched it take its first steps alongside the other co-creators. Thankfully, I don't have to save for the DBIR's college tuition.[70]

I couldn't be prouder of what the past and present members of the VTRAC have built and accomplished over the past 20 years. It is the passion and dedication of each and every team member that contributes to our long client tenure, never having missed a contractual service level agreement, world-class thought leadership and consistent rating as a leader by industry analysts.

I look forward to the adventures, innovation and excitement to come in our next 20 years!

Happy 20th birthday, VTRAC!

—Chris Novak

70 Editor's note: We hope the DBIR is actually helping you pay for tuition for your human children.

Appendix D: Contributing organizations

A

Akamai Technologies

Ankura

Apura Cybersecurity Intelligence

B

Bit-x-bit

BitSight

BlackBerry

C

Censys, Inc.

Center for Internet Security

Cequence Security

CERT Division of Carnegie Mellon University's Software Engineering Institute

CERT – European Union

CERT Polska

Check Point Software Technologies Ltd.

Chubb

Coalition

Computer Incident Response Center Luxembourg (CIRCL)

Coveware

CrowdStrike

Cybersecurity and Infrastructure Security Agency (CISA)

CyberSecurity Malaysia, an agency under the Ministry of Communications and Multimedia (KKMM)

Cybersixgill

CYBIR

D

Dell

Department of Government Services, Victorian State Government, Australia

DomainTools

E

Energy Analytic Security Exchange (EASE)

Edgescan

Elevate Security

Emergence Insurance

EUROCONTROL

Eviden

F

Federal Bureau of Investigation – Internet Crime Complaint Center (FBI IC3)

Fortinet

G

Global Resilience Federation

GreyNoise

H

HackEDU

I

Irish Reporting and Information Security Service (IRISS-CERT)

Ivanti

J

JPCERT/CC

K

K-12 Security Information Exchange (K-12 SIX)

Kaspersky

KordaMentha

L

Legal Services – ISAO

M

Malicious Streams

Maritime Transportation System ISAC (MTS-ISAC)

mnemonic

N

NetDiligence®

NETSCOUT

O

Okta

OpenText Cybersecurity

P

Palo Alto Networks

Proofpoint

S

S21sec

SecurityTrails, a Recorded Future Company

Shadowserver Foundation

SISAP – Sistemas Aplicativos

Shodan

Swisscom

U

U.S. Secret Service

V

VERIS Community Database

Verizon Cyber Risk Programs

Verizon Cyber Security Consulting

Verizon DDoS Defense

Verizon Network Operations and Engineering

Verizon Threat Research Advisory Center (VTRAC)

Vestige Digital Investigations

W

WatchGuard Technologies, Inc.


Akamai
ankura
apura
bit-x-bit
BITSIGHT
BlackBerry
censys
CIS Center for Internet Security
CEQUENCE
Carnegie Mellon University Software Engineering Institute
CERT-EU
NASK CERT.PL
CHECK POINT
CHUBB
Coalition
CIRCL Computer Incident Response Center Luxembourg
COVEWARE
CROWDSTRIKE
CISA
CyberSecurity MALAYSIA
cybersixgill
CYBIR
DELL Technologies
VICTORIA State Government Australia
DomainTools
EASE Energy Analytic Security Exchange
edgescan
Elevate Security
emergence
EUROCONTROL
EVIDEN
FORTINET
GLOBAL RESILIENCE FEDERATION
GREYNOISE

Hack_EDU	IRISS Irish Reporting and Information Security Service	ivanti	JPCERT CC®	K12 SIX
kaspersky	KordaMentha	LEGAL SERVICES ISAO	Malicious Streams	Maritime Transportation System Information Sharing and Analysis Center
mnemonic Securing your business.	NetDiligence®	NETSCOUT.	okta	opentext™
paloalto NETWORKS	proofpoint. Security Awareness Training	S21 SEC Cyber Solutions by Thales	SecurityTrails A Recorded Future® Company	SHADOWSERVER Lighting the way to a more secure Internet
SISAP Sistemas Aplicativos	SHODAN	swisscom		
Verizon Cyber Risk Programs	**Verizon Cyber Security Consulting**	**Verizon DDoS Defense**	**Verizon Network Operations and Engineering**	**Verizon Threat Research Advisory Center (VTRAC)**
VESTIGE Digital Investigations	WatchGuard