Appendix A: Transnational hacker debriefs

Insights into their target selection and tactics, techniques and procedures
– Michael D’Ambrosio, Deputy Assistant Director, United States Secret Service

Over the past fifteen years, the United States Secret Service has successfully identified, located, and arrested numerous high-value cybercriminals. These individuals were responsible for some of the most significant and widely publicized data breaches of public and private industry networks. Over this period, the Secret Service’s Cyber Division has cultivated mutually beneficial partnerships with law enforcement agencies around the globe, which has extended the reach of the Secret Service’s investigative efforts far beyond its traditional limits. This network of collaborative partners has enabled the Secret Service to successfully extradite criminal suspects located overseas and have them face prosecution in the United States. The Secret Service continues to forge new international partnerships in furtherance of its mission to pursue and apprehend cybercriminals regardless of their geography.

As part of its mandate to combat financially motivated cybercrime, the Secret Service combines its investigative efforts with educational outreach programs. These are aimed at strengthening the ability of private and public sector entities to protect themselves against a range of cybercrimes. The Secret Service conducts in-depth analyses of the activities, tools, and methodologies used by the cybercriminals during the commission of their crimes to better assess the evolving threats that cybercriminals pose to financial institutions and other potential targets. The Secret Service then shares the results of these reviews with its network of public and private partners through its outreach programs.

The Secret Service’s Cyber Division has learned that the most prescient information about cybercrime trends often comes from the cybercriminals themselves. The Secret Service conducts extensive debriefings of arrested cybercriminals and uses their first-hand knowledge to understand more fully the spectrum of variables they used to identify and select a particular target for intrusion and exploitation. The Secret Service has recently completed such debriefings with a handful of highly skilled cybercriminals who were responsible for some of the most significant network intrusions in history, and has found that the ways in which these individuals select their targets and perpetrate their crimes share certain common features.
Cybercriminals prey upon human error, IT security complacency, and technical deficiencies present in computer networks all over the world. Individually, each of these tactics, techniques and procedures (TTPs) discussed below are not always initially successful and may seem easily mitigated; it is when multiple TTPs are utilized in concert that cybercriminals are able to gain and maintain access to a computer network, no matter their motives. Once they are inside a network their process is almost always the same: establish continued access, escalate or obtain administrator privileges, move slowly and quietly to map the entire network, look for open ports, locate the “crown jewels,” and exfiltrate the data undetected for as long as possible.

The selection of a target is a continual process. Cybercriminals do their research. Almost always during these interviews, the hackers referred to gathering valuable intelligence from the same cybersecurity blogs, online IT security publications, and vulnerability reports that network administrators should be monitoring. They know that once a vulnerability is revealed, they still have a limited amount of time to try to exploit that vulnerability at a potential victim organization. Every time a vulnerability is disclosed or a system update or patch is released, a hacker sees an opportunity. They research the disclosure or update notes to learn if they can exploit the vulnerability and where, searching for their best opportunity to monetize the vulnerability. Hackers also communicate vulnerability information and exploit techniques on hacking forums. Once a target is selected, the hacker conducts thorough research into the victim organization and their network(s), often using free and commercially available Internet scanning tools that reveal extremely useful information about the victim company’s network.

Webserver and/or webpage hacking has been a highly successful primary attack vector, as there are various potential avenues for exploitation. These include the main website of an institution or a less protected linked website, which in turn can provide access to the main network. The added use of Structured Query Language (SQL) database injections of malicious code has been a very effective attack vector because these types of intrusion techniques can be deployed at any access point of a website. There are additional webserver attack vectors such as overlooked or forgotten IP addresses, possibly from development or beta-testing and external webservers or data servers that share the same or common domain. Unmanaged servers that still utilize Unicode can be exploited via encoding the URL with certain characters to bypass application filters.

Other traditional and effective attack vectors should not be overlooked. These include spear phishing for login credentials or malware delivery and “Man in the Middle” attacks through poorly secured routers or web gateways. Botnets are a relatively inexpensive tool that have been used to degrade or brute force attack networks in connection with parallel tactics. A very skilled hacker admitted to the Secret Service that he ended up paying a collusive employee (insider threat) when all of his other hacking attempts to access a foreign bank’s network were unsuccessful.
Once inside a network, cybercriminals continue to do their research and reconnaissance. Hackers often examine a webserver's default error pages because those pages expose a lot of the target network's system information. Cybercriminals take all of network information they can collect and utilize virtual machines (VMs) to build a mock system to emulate the network of the victim company. This is done both for testing their methods of exploitation and for better understanding the types of network defenses present within the system.

The exploits used by cybercriminals inside a target network depend on the installed network defenses. Undoubtedly, the hacker will try to install a web shell to ensure access into the system. Another sustainment method is the use of cross-site scripting (XSS) for session hijacking (cookie stealing) of a valid user through malicious code injections into a user's JavaScript, ActiveX, Flash, or other code bank. The use of malware delivered to the valid user via spear phishing is a key component of this process.

In addition, hackers utilize directory transversal attacks (directory climbing, backtracking, etc.) on web servers to attempt to reach otherwise restricted directories, such as Secure Socket Layer (SSL) private keys and password files. Hackers can even execute commands on the server by accessing such directories. After administrator privileges are obtained, it is common for the prized data to be exfiltrated by tunneling via a remote access protocol. Cybercriminals will also scan for open ports and attempt to install software of their choosing on non-standard ports for a variety of malicious uses. If the targeted network has the potential to provide valuable data continuously, diligent hackers will continuously clean up their “tracks” within the exploited network to obfuscate their presence indefinitely. Another prominent hacker described having persistent access into a company's networks for 10 years using multiple “backdoors” (web shells) and continually cleaning up his “work” to go undetected. In reality, many of the hackers we debriefed often stated that they could see traces of other hackers in the targeted network which sometimes made it harder to hide their hacking exploits.

These are just some of the tactics, techniques and procedures the Secret Service has observed used by criminal groups to exploit victim networks. The threat is real and the adversary is constantly evolving, driven by diverse and varying motivations. Their success is more often dependent on how well network administrators can adapt their defenses to potential vulnerabilities as they are revealed.

The Secret Service will continue to pursue, arrest, and prosecute cybercriminals no matter where they are and we will continue to provide valuable attack methodology analysis from our investigations to better improve the cybersecurity efforts of our partners in law enforcement, academia, and the public and private sectors alike.
Appendix B: 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.

Our overall methodology 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 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.

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 three months as data is collected and analyzed.

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 detail necessary to pass the quality filter, the incident must be within the timeframe of analysis, (November 1, 2017 to October 31, 2018 for this report). The 2018 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 personal laptop was hit with CryptoLocker 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 sample bias.

Acknowledgement of sample bias

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. Unfortunately, we cannot measure exactly how much bias exists (i.e., in order to give a precise margin of error). We have no way of knowing what proportion of all data breaches are represented because we have no way of knowing the total number of data breaches across all organizations in 2018. 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).

While we believe many of the findings presented in this report to be appropriate, generalization, bias, and methodological flaws undoubtedly exist. However, with 73 contributing organizations this year, we’re aggregating across the different collection methods, priorities, and goals of our partners. We hope this aggregation will help minimize the influence of any individual shortcomings in each of the samples, and the whole of this research will be greater than the sum of its parts.

Statistical analysis

We strive for statistical correctness in the DBIR. In this year’s data sample, the confidence interval is at least +/- 2% for breaches and +/- 0.5% for incidents. Smaller samples of the data (such as breaches within the Espionage pattern) will be even wider as the size is smaller. We have tried to treat every statement within the DBIR as a hypothesis based on exploratory analysis and ensure that each statement is accurate at a given confidence level (normally 95%). We’ve tried to express this confidence in the conditional probability bar charts explained in the “tidbits” that precede the Table of Contents.

18Bayes method, 95% confidence level.
19If you wonder why we treat them as hypotheses rather than findings, to confirm or deny our hypothesis would requires a second, unique data set we had not inspected ahead of time.
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 percent. For example, if there are 5 botnet breaches, the sample size is 5. However, since each botnet used phishing, installed keyloggers, and used stolen credentials, there would be 5 social actions, 5 hacking actions, and 5 malware actions, adding up to 300 percent. 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 too little 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. Finally, “Not Applicable” (normally “NA”) may be counted or not counted depending on the hypothesis.

**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 analyze botnet-related incidents.

Both subsets were separately analyzed last year 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 the aforementioned two subsets.

**Non-incident data**

Since 2015, 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 reporting on the median organization rather than the average of all data). 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.
Last year in the “Feeling vulnerable?” appendix, we discussed the services or weaknesses attackers look for in spray and pray internet scans, and how those aren’t necessarily the same things they look for in targeted attacks. In this section, we again examine what services are open to the internet and the adversary activity against them. At the risk of stating the obvious, what the attacker looks for tells you a great deal about what is of value to them.

**Any port in a storm**

Ports that offer at least some value to, and at the same time require the least amount of investment from the attacker garner a lot of attention. An economist might call the amount invested by the actor per attack the marginal cost. The very best attacks from the criminal’s point of view would cost almost nothing per target. We will refer to these as zero-marginal-cost attacks.

![Figure 65. Comparison of ports in DDoS and honeypot attacks](image-url)
Figure 65 illustrates the ports that are in the top 50 for both honeypot activity and DDoS attacks (with “1” in the upper right being the most common and the rest decreasing from that point). We can consider how often attackers look for a given port as an indicator for how valuable they are to the attacker. Ports below the red line, such as cLDAP (389), DNS (53), and NTP (123) are more valuable due to their DDoS amplification potential. The ports above the red line are more valuable for their non-DDoS malevolence including SSH (22), telnet (23), HTTP (8080), NetBIOS (445), and others.

**Portémon Go**

Probably the most effective way to judge perceived value for the attacker for a given port in zero-marginal-cost attacks is to examine their ranking in honeypot scans vs their general population ranking on the Internet. There are a myriad of organizations that scan the internet regularly, and there are a few of those who are gracious enough to contribute to the DBIR. As a result, we can share this data in Figure 66.
Figure 66 lists the top 10 ports by ratio of honeypot activity to internet prevalence. Some of these, for example, Telnet, NetBIOS, and SQL Server – legacy services with known weaknesses that are old enough to vote – may not be as common as dirt, but they still exist and when an attacker finds them you can almost hear the intro to Pink Floyd’s “Money” floating in the ether. If your organization has any of these services exposed to the internet, it’s probably a good idea to go and take care of that now. We’ll wait here. Take your time. This report changes once a year, but those ports are being hammered daily.

**Dime a dozen**

The above section begs the question, “If those ports are what attackers frequently search for but rarely find, which open ports are plentiful but rarely sought?” We are glad you asked. For the most part they are unassigned or ephemeral ports. Of more interest are the ports that appear in vulnerability scans, but do not show up in honeypots. Figure 67 gives us some insight into that area. The main takeaway is that there are a lot of ports far down on the list from a honeypot perspective (the big cluster in the lower left of the figure) that get reported often in vulnerability scans. Those are the vulnerabilities that may be useful for attackers but either only for niche attacks or internal pivoting, or are of absolutely no interest whatsoever to the attacker.

---

"For example, if a port was the top ranked port in honeypot scans and the 15th most common on the internet, its ratio would be 15x."
Take action

There may only be seven seas, but there are 65,535 ports. While not all are found in the figures above, a great many are. So now what? We suggest you take a look to ascertain if you are vulnerable to any zero-marginal-cost attacks (easily identified by their honeypot to internet scan ratio). If so, you are operating below a critical security threshold and you need to take action to get above it. Are you running a honeypot yourself? If not, why is that port open? Finally, take a cue from the Unbroken Chains section and be smart about what else you mitigate. Understand the paths attackers are most likely to take in order to exploit those services.
Appendix D: Contributing organizations
| A | Akamai Technologies  
    | Apura Cyber Intelligence  
    | AttackIQ  
    | Avant Research Group, LLC |
|---|---|
| B | BeyondTrust  
    | BinaryEdge  
    | BitSight  
    | Bit-x-bit |
| C | Center for Internet Security  
    | CERT Insider Threat Center  
    | CERT European Union  
    | Checkpoint Software Technologies Ltd  
    | Chubb  
    | Cisco Security Services  
    | 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)  
    | Cyance |
| D | Dell  
    | DFDR Forensics  
    | Digital Edge  
    | Digital Shadows  
    | Dragos, Inc |
| E | Edgescan  
    | Emergence Insurance |
| F | Federal Bureau of Investigations Internet Crime Complaint Center (FBI IC3)  
    | Fortinet |
| G | Gillware Digital Forensics  
    | Government of Telangana, ITE&C Dept., Secretariat  
    | GRA Quantum  
    | GreyNoise Intelligence |
| I | Interset  
    | Irish Reporting and Information Security Services (IRISS-CERT) |
| J | JPCERT/CC |
| K | Kaspersky Lab  
    | KnowBe4 |
| L | Lares Consulting  
    | LIFARS |
| M | Malicious Streams  
    | McAfee  
    | Mishcon de Reya  
    | Moss Adams (formerly ASTECH consulting)  
    | MWR InfoSecurity |
| N | National Cyber-Forensics and Training Alliance (NCFTA)  
    | NetDiligence  
    | NETSCOUT |
| P | Paladion  
    | Palo Alto Networks  
    | Proofpoint |
| Q | Qualys |
| R | Rapid7  
    | Recorded Future |
| S | S21sec  
    | Shodan  
    | Social-Engineer, Inc.  
    | SwissCom |
| T | Tripwire |
| U | US Secret Service  
    | US Computer Emergency Readiness Team (US-CERT) |
| V | VERIS Community Database  
    | Verizon Cyber Risk Programs  
    | Verizon Digital Media Services  
    | Verizon DOS Defense  
    | Verizon Managed Security Services  
    | Verizon Network Operations and Engineering  
    | Verizon Professional Services  
    | Verizon Threat Research Advisory Center  
    | Vestige Ltd |
| W | Wandera  
    | West Monroe Partners  
    | Winston & Strawn LLP |
| Z | Zscaler |