



# 2017

## STATE OF VULNERABILITY RISK MANAGEMENT REPORT



IN PARTNERSHIP WITH



# TABLE OF CONTENTS

|  |    |
|--|----|
| I. Introduction  | 1  |
| A. The DNC Email Dump  | 1  |
| B. The Dyn DDoS Attack of October 2016   | 1  |
| C. The Global WannaCry Ransomware Attack   | 2  |
| II. Tracking Attacks   | 3  |
| III. Vulnerability Landscape   | 4  |
| A. Data and Methodology  | 4  |
| B. Security Vulnerabilities by Industry  | 5  |
| C. Security Vulnerabilities by Vendor  | 6  |
| IV. Risk Analysis: Social Media and Dark Web   | 8  |
| A. Social Media Activity   | 9  |
| B. The Dark Web Element  |    |
| 1. Dark Web Activity   | 10 |
| 2. Look Through the Other End of the Telescope   | 12 |
| V. Malware Data Analysis and Other Input by the AlienVault Labs Security Research Team | 13 |
| VI. Conclusion   | 14 |
| VIII. About NopSec Labs  | 15 |

## **I. Introduction**

The past year was particularly turbulent for the cybersecurity landscape. In just twelve short months, we encountered several online attacks that were notable not just by their potency, but also by the level of disruption they caused. There was also a troubling uptick in terms of how frequent online attacks -- ranging from data breaches, to disruptive DDoS attacks, to virulent ransomware outbreaks -- were encountered.

Security issues, which used to be near-exclusively the focus of infosec professionals and the specialist technology press, are now so damaging, they capture the attention of the media and general public, and dominate the news cycle. Over the past twelve months, there have been three glaring examples of this, which everyone reading this report will likely be familiar with.

### **The DNC Email Dump**

The US election of 2016 was an undeniably fraught affair. Not only was cybersecurity a major policy talking-point, but it also played a role in the campaign with the Democratic National Committee (DNC) email leaks.

The controversial transparency site WikiLeaks published a cache of nearly 20,000 emails and 8,000 attachments obtained from the DNC. Wikileaks purportedly obtained these from a hacker by the name of Guccifer 2.0; itself a homage to the original Guccifer, Romanian hacker Marcel Lazăr Lehel, who have reportedly hacked several US political figures, including Colin Powell and George W. Bush.

### **The Dyn DDoS Attack of October 2016**

It's not uncommon for distributed denial of service (DDoS) attacks to target individual websites and services. It's far less common to hear of an attacker attempting to collapse the infrastructure of the Internet under a flood of packets and bits with the use of Mirai-infected Internet-connected devices (or Internet of Things/IoT devices).

October 2016 was, therefore, a watershed moment. An attacker (or attackers) targeted Dyn -- one of the most widely used DNS platforms on the Internet -- with a gargantuan amount of traffic, with the aim of rendering it inaccessible to legitimate users. This resulted in several popular Internet services slowing to a crawl, or being altogether unavailable to users.

## **The Global WannaCry Ransomware Attack**

WannaCry was an astonishingly destructive specimen of ransomware. In just one day, it spread to over 300,000 computers, many of these in critical enterprise and government environments. Intelligence sources also believe that the WannaCry ransomware was developed by the North Korean government, and propagated by “Lazarus Group,” which is the prime suspect in the Sony Pictures hack of 2014.

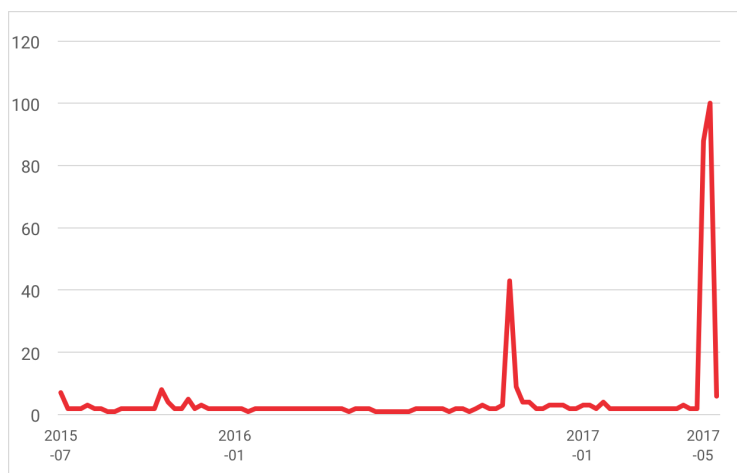
In total, WannaCry was said to have spread to 99 countries. As of July 18, 2017, the three BitCoin wallets associated with it have earned a total of just 51.65 Bitcoins. At current market rates, this translates to just \$219,565.

NopSec has pioneered the research, measurement, and analytics of vulnerability threats since 2013. Its annual State of Vulnerability Risk Management reports are widely used and cited in the cybersecurity industry for its insights and actionable information. As presented in this report, vulnerability threats are ever more expanding and evolving, and NopSec is once again leading the research for new ways to expose these threats and protect valuable assets from getting compromised.

## II. Tracking Attacks

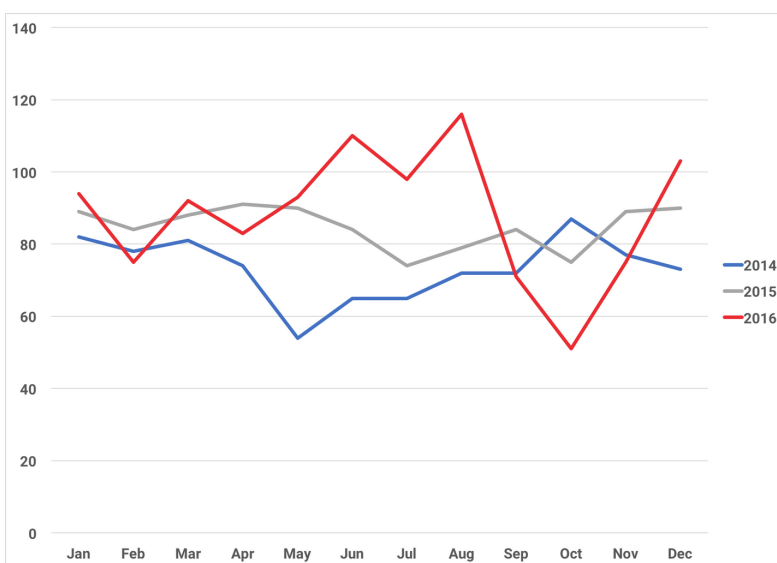
For obvious reasons, it is extremely challenging to create an accurate and comprehensive survey of every single cyber attack. In order to identify trends with respect to size and proliferation, we must look at a diverse array of sources.

This chart from Google Trends (see *Figure 1*) tracks instances of the phrase “cyber attack” in news articles published online. As you can see, there are several instances where it spikes dramatically. These spikes coincide with several key events, including the attack on Dyn, and the WannaCry infection earlier this year.



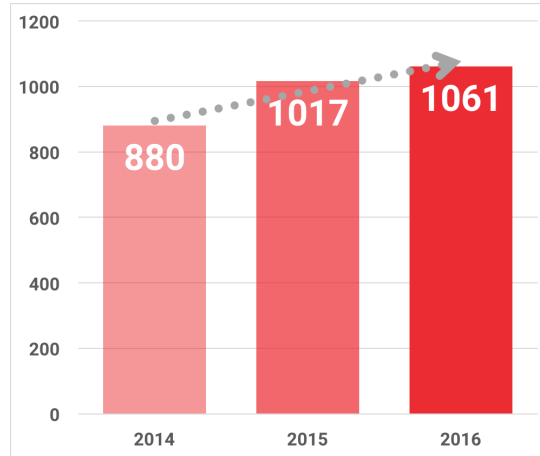
*Figure 1.*  
Google News Trends  
"Cyber Attack" Results by Time

Similar insights can be gleaned from this chart by Hackmageddon (see *Figure 2*), which tracks cyber attacks per month. These date from 2014, 2015, and 2016. As you can see, the number of attacks vary per month. The only clear trend is that the summer of 2016 saw a flurry of activity unparalleled by previous years.



*Figure 2.*  
Number of Cyber Attacks  
by Time

When we summarize the results in an annual basis, we notice that the spate of attacks has soared between 2014 to 2016 (see *Figure 3*).



*Figure 3.*  
Annual Number of Cyber Attacks  
from 2014 to 2016

### III. Vulnerability Landscape

Thanks to NopSec's position in the market, we are able to gain a "bird's-eye" view of the overall "vulnerability landscape." The following section of the report explores two threads of thought. The first looks at what industries are being targeted by bad actors. The second examines the proliferation of vulnerabilities on a vendor-by-vendor basis, in order to determine the most common avenues of ingress for an attacker.

#### Data and Methodology

The analysis in this report is based on aggregated anonymized NopSec Unified VRM client data, which consists of over 1 million unique vulnerabilities found on our clients' systems.

For our purposes, we define a unique vulnerability as a unique combination of client, vulnerability ID, asset, and port affected. We use this definition because a vulnerability's intrinsic attributes are only one part of risk – the context in which a vulnerability is present is often just as important.

Our clients span a wide range of industries, but for the purposes of this report, we have classified them into one of four broad industry categories: Financial, Technology, Healthcare, and Other. We have integrated our client data with information on public exploits (from sources such as the Exploit DB and Metasploit), malware correlation data coming from several threat intelligence feeds (30+ unique feeds), social media information from Twitter, and other sources (such as CVE, CWE and CPE information) to add additional context, classification, and to give a comprehensive overview on the State of Vulnerability Risk Management for our clients.

We are going to look at NopSec's client vulnerability data from various angles: an industry view, to see that "one size does not fit all" for vulnerability management on various industry verticals; a product view, to highlight the fact that various vendors have different impact on the total vulnerability impact to the networks; and an analysis of social media and dark web impact in terms of vulnerability risk regardless of their original CVSS score and public exploit availability.

It is important to note that our analysis comes from a convenience sample of our clients – as such, we do not claim that this is a definitive analysis of all possible vulnerabilities and threats an average organization could face. The possibility of sample bias exists, and this should be kept in mind throughout the report. However, we believe that our research offers important insight into how companies in various industries address vulnerabilities, universal weaknesses companies across industries share, and factors that should be incorporated into a comprehensive threat detection and remediation program.

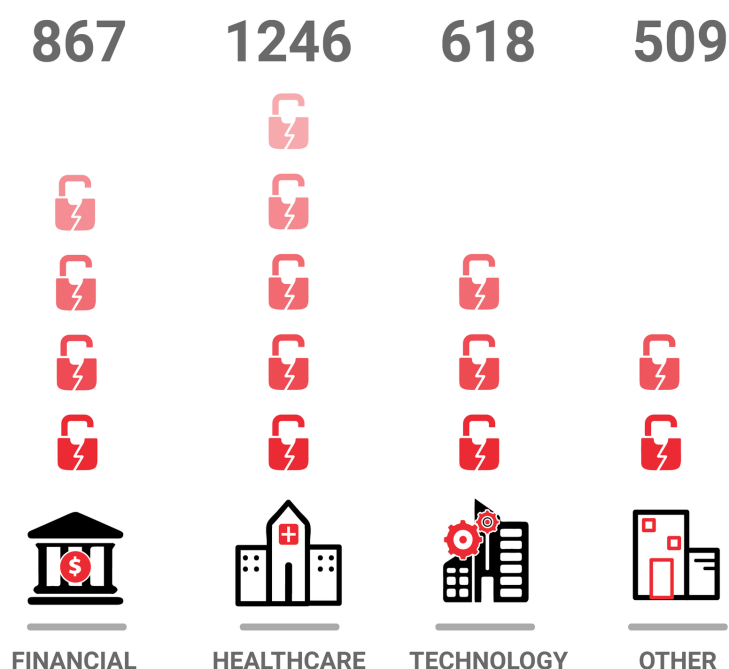
## **Security Vulnerabilities by Industry**

We begin our analysis with an overview of vulnerability counts for clients in each of our industry categories. By examining overall and per scan vulnerability counts, we can gain insight into the magnitude of threats clients in different industries face, and overall remediation trends.

A statistical note: because the distribution of vulnerability counts for clients is right-tailed (there are a few companies with very large numbers of vulnerabilities), we found that a few clients had an overly large impact when we averaged our data. Therefore, we chose to take the median, as it is significantly more robust to outliers. We will continue to use the median of our data throughout this report, as we believe it gives a better representation of the threat landscape for a "typical" client.

Following is a graph (see *Figure 4*) depicting the median number of vulnerabilities discovered per client \*per scan. These numbers illustrate the total number of vulnerabilities discovered on typical NopSec client systems since they began using Unified VRM.

Figure 4.  
Vulnerability Discovered Per Client  
Per Scan by Industry

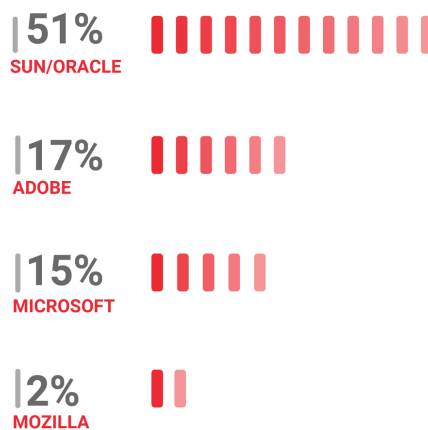


The conclusions to be drawn from this is that each of the industry verticals has its own, specific security challenges. In many cases those arise from the fact that each particular technology stack that is common across all organizations in a given industry has its own exposure profile. The tools that an organization is using to manage vulnerabilities and track remediation efforts, therefore, needs to be flexible enough and configurable enough to address that exposure profile.

## Security Vulnerabilities by Vendor

Next, we examine top vendors by industry in order to determine which are the most vulnerable. Again, in order to get the best picture of what a "typical" client faces, we measure the median number of vulnerabilities each client has. The chart below (see *Figure 5*) shows the vendors with the most vulnerabilities per client associated with them.

Figure 5.  
Top 4 Most Vulnerable Vendors  
Overall





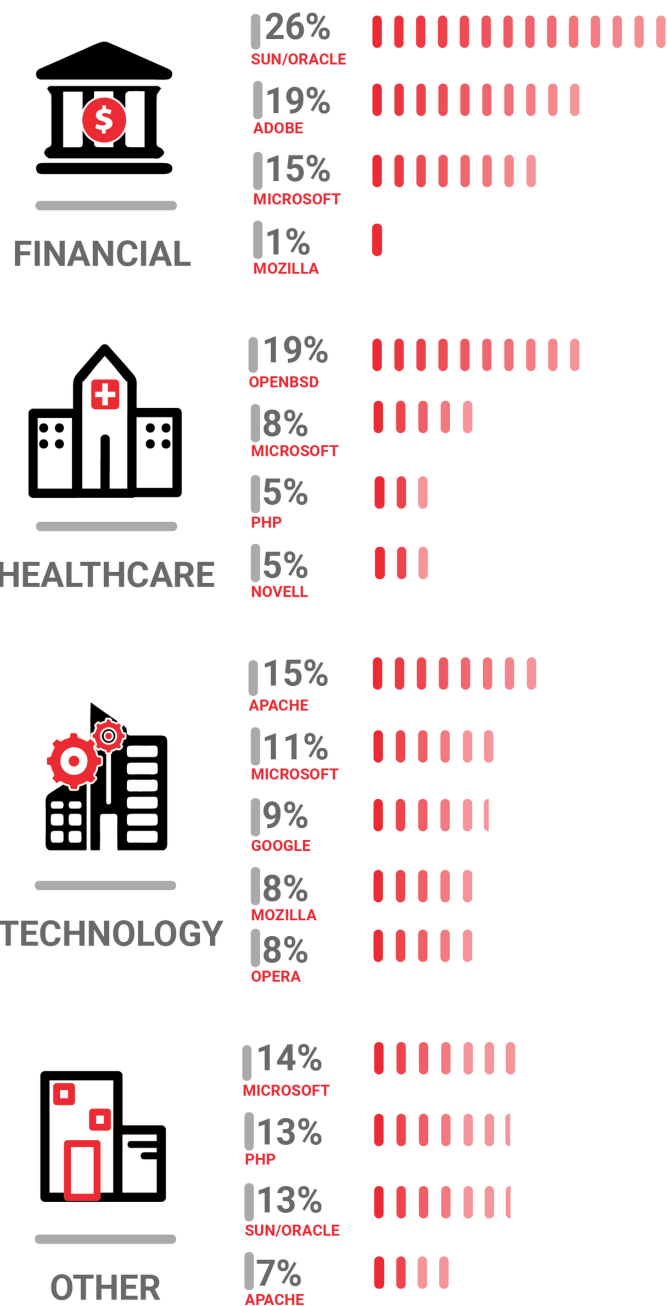


Figure 6.  
Top Vulnerable Vendors  
by Industry

Microsoft and Adobe vulnerabilities are always top of the chart (see Figure 6). However, Oracle products and Java vulnerabilities are important to consider.

## IV. Risk Analysis

Research indicates that social media, particularly Twitter, is becoming the go-to resource for security researchers and attackers looking to disseminate proof-of-concept exploits. While some vulnerabilities have thousands of Twitter interactions -- counted by tweets mentioning their CVE identifier -- most vulnerabilities are never tweeted about or or tweeted only once. The top-most tweeted CVEs reflect interactions focused on well-publicized, dangerous vulnerabilities. This phenomenon reflects a so-called *right tail distribution*.

Because of this correlation between social media activity and the degree of risk that a vulnerability poses to an organization, NopSec collects and incorporates Twitter data as one component of its risk evaluation process.

One of the questions that we recently asked is whether a similar signal can be extracted from Dark Web activity around particular vulnerabilities. The results are mixed. In order to understand the approach, it is important to understand the methodology employed with the research using Twitter data.

An indicator of risk associated to vulnerability is the availability public exploits for that vulnerability. There are several databases that aggregate this data, and NopSec includes these as part of its threat intelligence gathering. Using these data sources, it is possible to build a quantifiable indicator of risk for a CVE based on the number of Twitter interactions it receives. In the language of Machine Learning, this called a Bayesian model. The particular model of interest is expressed by the probabilistic relationship:

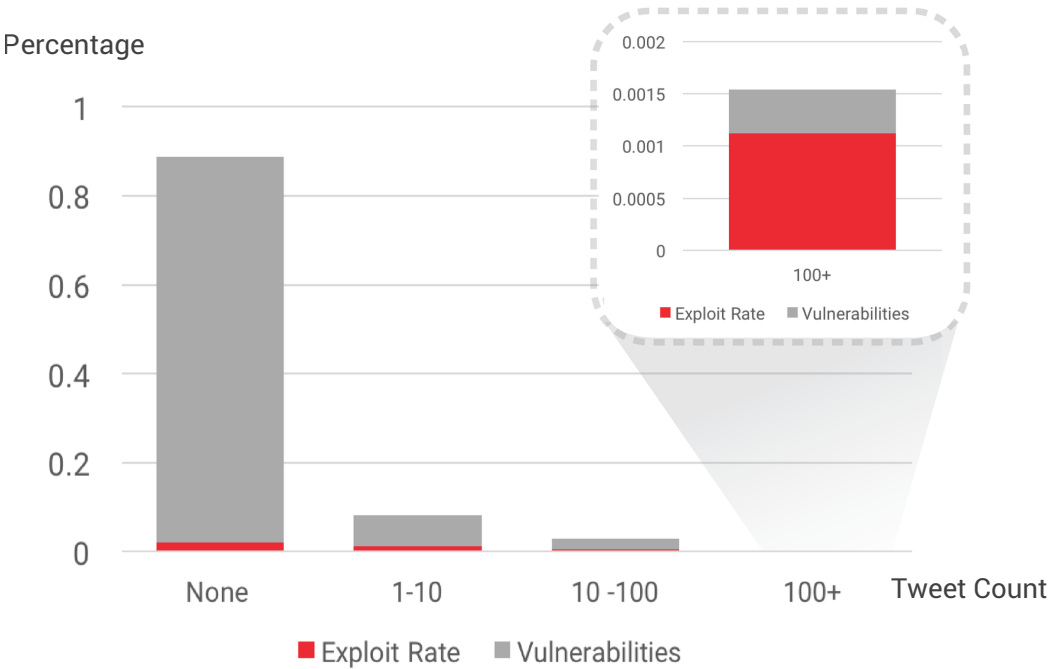
$$p(\text{has exploit} \mid \text{tweets}) \propto P(\text{has exploit}) p(\text{tweets} \mid \text{has exploit})$$

In words: the likelihood of a vulnerability having an exploit, given its number of twitter interactions, is proportional to the the rate of Twitter interactions for those vulnerabilities that do have an exploit.

It is important to keep in mind that this equation does not attempt to paint a causal relationship, but instead aims to extract a quantifiable signal from a correlation in the data. In the language of Machine Learning, again, it is what is known as a smoothing technique.

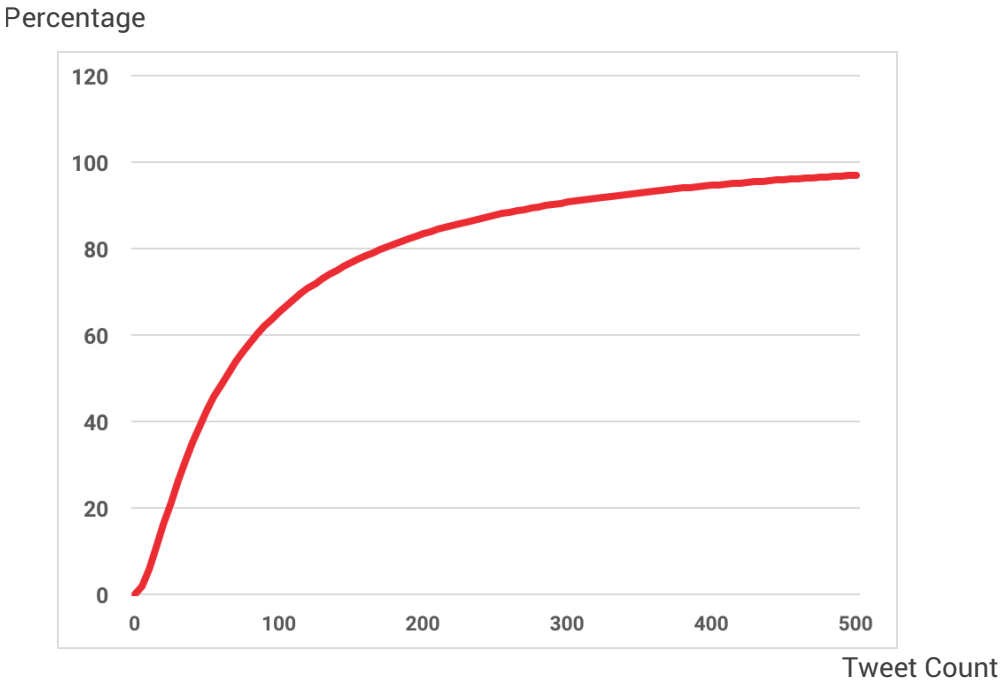
The right hand side of the equation is straightforward to model directly from the raw data. The expression " $p(\text{tweets} \mid \text{has exploit})$ " tells most of the story, and is summarized in the next figure (see *Figure 7*).

Figure 7.  
Relationship Between  
Exploit Likelihood  
and Tweet Count



This breaks down the population of CVEs into bins based on the number of Twitter interactions, and splits each bin into subpopulations of those vulnerabilities with an associated exploit and those without. Although less than 0.2% of all vulnerabilities fall into the bin of 100 Twitter interactions or more, almost three quarters of those have an associated exploit. The next figure (see Figure 8) captures the "smoothed" signal of exploit likelihood, rescaled as a percentage, as a function of the number of Twitter interactions.

Figure 8.  
Relationship Between  
Fitted Exploit Likelihood  
and Tweet Count



This summarizes the left-hand side of the Bayesian model, expressing the correlation between Twitter interactions and exploit likelihood, but does so as a predictive tool.

## Dark Web Activity

The utility of a Machine Learning model, in addition to extracting a signal from noisy data, is that it yields a way to experimentally evaluate additional sources of data.

As part of an early stage research project, NopSec acquired a dump of Dark Web data covering conversations about software vulnerabilities from a crawl of Dark Web sources. In total, about two thousand distinct "pages" from different Dark Web sources that included the string "CVE-" as part of the search results were returned. These search results were subsequently normalized and analyzed. These sources are heterogeneous -- forums and marketplaces -- but excluded any pages that are behind a login or required authentication to access. In other words, the Dark Web data captured was less public than that accessible via Twitter, but includes only what is visible from the "public" onion network without any registration or login.

On the question of whether a heterogeneous data set of Dark Web sources is comparable to data sourced from a single social network, there are two points of comparison which suggest that Dark Web data is amenable to the same mining technique as applied to the Twitter data:

1. The distribution of CVE reference counts within the corpus of Dark Web data follows a similar long-tailed distribution as the tweet count from the corpus of Twitter data.
2. The coverage of distinct CVE references from each data source is roughly the same size as seen in the following table.

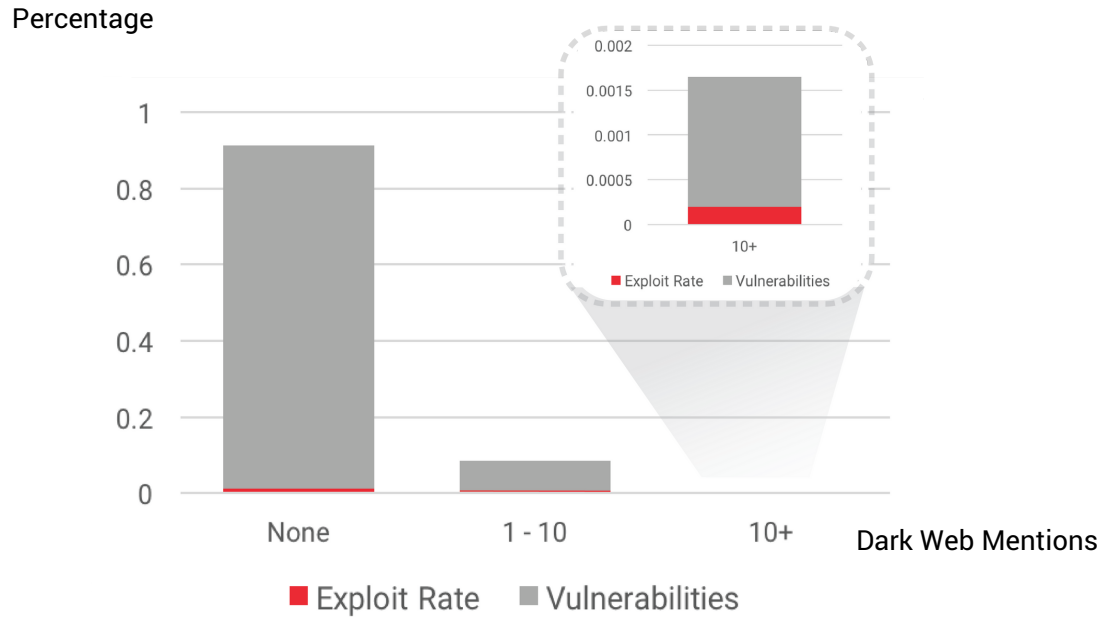
| Source        | Distinct CVEs |
|---------------|---------------|
| Twitter Data  | 11,800        |
| Dark Web Data | 9,300         |

As in the analysis of Twitter data, the proxy for indication-of-risk is the existence of an associated public exploit. Instead of "tweet count" for a CVE as the variable of interest, the reference count for a CVE within the corpus of Dark Web data is taken as the indicator variable.

This aligns with the observation that many of the "pages" captured in the Dark Web crawl constitute a discussion thread discussion in a forum -- much like a conversation on Twitter would be represented in a connected sequence of distinct tweets from different handles.

Our main conclusion is that that the correlation between our indication-of-risk signal with Dark Web activity is much weaker than the correlation with Twitter activity. The main story is evident in the following chart (see *Figure 9*). The population of CVEs is partitioned into bins according to the number of CVE references. The relative frequency of indication-of-risk within each subpopulation, however, remains relatively flat.

*Figure 9.*  
Relationship Between  
Exploit Likelihood  
and Dark Web Mentions



This is in contrast to the Twitter data, which reflects roughly an order-of-magnitude increase in relative risk with each successive subpopulation of CVEs in the "long tail." Attempting to fit a smoothed model for exploit likelihood from Dark Web reference counts did not even converge to stable statistical indicator.

## Looking Through the Other End of the Telescope

Perhaps there is good reason for the failure of this particular model of risk associated to Dark Web activity. Scrutiny may be applied to the signal that this model employs for indication-of-risk, namely, the existence of an associated public exploit. The Dark Web constitutes a channel for conversations that are intentionally invisible to the larger public web, and it would stand as significant if the conversations about exploitable vulnerabilities were a mirror of the topic set of vulnerabilities with a public exploit. Another challenge to the assumptions of that line of research is that it is restricted only to vulnerabilities with a CVE identifier; it may be that there is significant Dark Web activity around vulnerabilities that fall outside of this set, and that therefore would not have been easily captured in the original data crawl.

A separate line of research was undertaken to address these challenges. The focus of this approach was aimed at capturing a set of vulnerabilities' exploits that are explicitly "for sale" among Dark Web sources, and to analyze this subset for potential patterns relevant to the landscape of vulnerability management.

On an exchange like Inj3ct0r [onion: <http://mvfjfugdwc5uwho.onion>], which discusses hundreds of exploits, including remote exploits, web exploits, proof-of-concept and denial-of-service attacks, the majority are offered for free, or reference an associated public exploit, especially those included in the Exploit BD website. About three dozen are marketed as "premium " exploits and have a price ranging between 0.008 and 1.448 Bitcoin (or between \$34 and \$6,174). For comparison, the same exchange includes about 390 exploits under the category of "Remote Exploit" which are offered for free or which have an existing public exploit. The 34 exploits for sale are summarized in the following table (see *Figure 10*).




|   | No<br>CVE | Has<br>CVE | Public<br>Exploit | Has<br>Patch |
|---|-----------|------------|-------------------|--------------|
| <br>Remote Exploit | 0         | 6          | 1                 | 6            |
| <br>Web Exploit    | 14        | 1          | 0                 | N/A          |
| <br>None of Above  | 10        | 13         | 3                 | 0            |

Figure 10.  
Comparisons Across Different  
Exploits

The two most significant observations from this data set are highlighted in the top row: 6 exploits for sale already have a CVE identifier assigned, but no associated public exploit, and therefore would constitute "false negatives" using this as an indication-of-risk factor.

The 5 exploits which fall into this class include:

- 2 Oracle Java Uninitialized Memory-type vulnerabilities
- 2 Adobe Flash Player Out-Of-Bounds Access-type vulnerabilities
- 1 Adobe Acrobat Reader Memory Corruption-type vulnerability

For reference, a sampling of those with no CVE from the year 2017 include:

- remote code execution vulnerability for Joomla 3.6.5
- proof-of-concept for sending unauthorized emails from Wix sites

## V. Malware Data Analysis and Other Input by the AlienVault Labs Security Research Team

In a collaboration with our industry partner, AlienVault®, NopSec performed an analysis with Open Threat Exchange® (OTX™) data along the lines of the "indication-of-risk" analysis carried out for social media and Dark Web activity.

OTX enables anyone in the security community to actively discuss, research, validate, and share the latest threat data, trends, and techniques. It has more than 65,000 participants in 140 countries, who contribute over 14 million threat indicators daily. Members can share threat data via "pulses" in OTX which provide a summary of the threat, a view into the software targeted, and related indicators of compromise (IOC) that can be used to detect the threat. Many of the OTX pulses include an explicit CVE identifier reference, giving a new signal for indication-of-risk from the standpoint of vulnerability assessment.

The basic question we attempted to answer was how well the OTX signal correlates with other indicators of risk, specifically social media activity. The brief analysis is highly encouraging (see Figure 11):

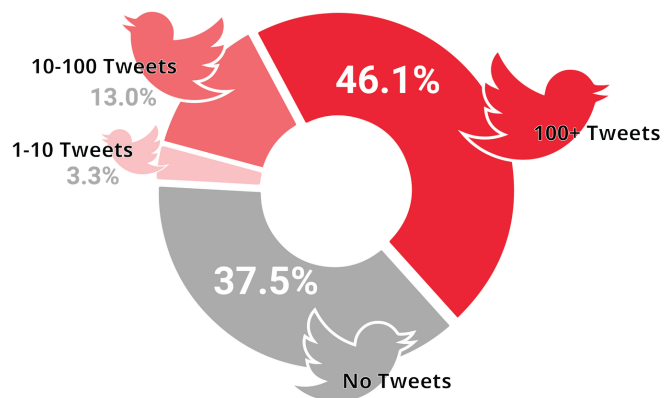


Figure 11.  
AlienVault OTX Social Media  
Pulses

For the pulses that are captured that reference a specific vulnerability, 46% have social media activity in the "critical" range (100+ tweets) of interest. The analysis validated the quality of AlienVault's threat intelligence data, and served only to increase NopSec's confidence in its ability provide exceptional insight and boost the prioritization results by NopSec Unified VRM.



AlienVault OTX captures a wide range of data from public and private sources, delivering indicators of compromise to our users to complement their threat intelligence processes and technologies.

The focus of NopSec's State of Vulnerability Risk Management Report is extracting insight from indicators of risk that point to specific signals from social media. This correlates well with the vulnerabilities reported by users of OTX. The data in OTX also displays a long-tail distribution, with a small set of vulnerabilities accounting for the bulk of our users reports about attacks they have seen in the wild. This analysis has helped validate the quality of OTX data.

## VI. Conclusion

Results drawn from this research suggest a cautionary tale for approaches that combine data mining and Machine Learning toward the challenge of evaluating the risk associated to individual vulnerabilities.

In general, models that are trained to rank vulnerabilities based on an indication-of-risk signal do provide value. In the absence of all other information, given a pair of vulnerabilities, such a model can be queried to render a decision as to which vulnerability to prioritize for remediation efforts. This research, however, illustrates that there are many different signals available in this space, and moreover, those signals move at different speeds or address different facets of coverage. Twitter activity moves at a speed that is faster than publication activity of organizations such as NVD, whose aim is to deliver risk summarized in a standardized, digestible format. Activity in Dark Web market forums serves as a signal exposing the gaps in coverage of other public data sources.

Re-established correlations and fresh perspectives emerge from the analysis of our 2017 report. Organizations continue to struggle to efficiently and accurately prioritize vulnerability risks due to information overload, inefficient manual processes, and the acute workforce shortage.



## **IX. About NopSec Labs**

The research contained in this report was conducted by NopSec Labs. All NopSec data presented in this report is compared with the population data from National Vulnerability Database, where relevant.

NopSec has a leading research and data science practice focused on analyzing malware, exploit, vulnerability, and other cyber threat risk patterns. Our team of data scientists applies that knowledge to help organizations forecast the probability of a data breach and improve prioritization, remediation, and reporting of critical vulnerabilities. Customers of NopSec's Unified VRM platform are provided with a variety of reports specific to their organization and similar to the data contained in this report.