Using EPSS to Predict Threats and Secure Your Network

fortinet.com/blog/threat-research/predict-threats-and-secure-networks-with-epss

April 29, 2022

In the world of cybersecurity and threat intelligence, understanding what threats you might face next is critical to effectively securing your network. The Common Vulnerability Scoring System (CVSS) has been a valuable tool in this fight because it highlights how exploitable different vulnerabilities are. And now, EPSS (Exploit Prediction Scoring System) can supplement CVSS by providing dynamic insight into the likelihood that a vulnerability will be exploited. It produces a probability score of a rational number between 0.0 and 1.0, and the higher the score, the greater the probability that a vulnerability will be exploited.

In this blog, I will review the CVSS scoring system used worldwide to prioritize software vulnerabilities. I will also highlight best practices for using CVSS at its full potential, which, unfortunately, few companies leverage.

I will then introduce EPSS, a new score that can predict the likelihood of exploitation of software vulnerabilities, and describe how it should be used in conjunction with CVSS to prioritize vulnerability patching. I will also illustrate why EPSS is more dynamic than CVSS using a popular vulnerability published last year that affected the library Log4j.

Finally, I will describe how Fortinet supports the calculation of EPSS scores by providing the daily telemetry data needed to establish the “Ground Truth” for the probabilistic training of the model.

Intro to CVSS

The Common Vulnerability Scoring System (CVSS) is an open framework for communicating the characteristics and severity of software vulnerabilities. CVSS is owned and managed by FIRST.Org, Inc. (FIRST), a US-based non-profit organization. Its mission is to assist computer security incident response teams worldwide. The official CVSS documentation can be found at https://www.first.org/cvss/.

Two common uses of CVSS are calculating a CVSS rating, or score, that attempts to provide a standardized measurement of the severity of vulnerabilities discovered on one's systems and as a factor in the prioritization of vulnerability remediation activities. While this is an invaluable tool for threat assessment, the only public entity that produces consistent and reliable CVSS scores is the NVD Database, which is widely adopted as a reference source for many private and public companies worldwide.

Let's take a deeper look at how the CVSS operates and how those scores are computed. Most security practitioners forget that CVSS consists of three metric groups: Base, Temporal, and Environmental. By examining these different metrics, you will better understand how CVSS works and its value to researchers and developers.

Figure 1: CVSS Metric Groups

The Base group represents the intrinsic qualities of a vulnerability that are constant over time and across user environments. These are broken down into two main groups: Exploitability metrics, such as how complex an attack must be to exploit a vulnerability and whether the user has to participate in the attack, and Impact metrics that address issues like confidentiality, integrity, and availability. The Temporal group reflects the characteristics of a vulnerability that change over time, such as the maturity of available exploitation code and the effort required for remediation. And the Environmental group looks at the characteristics of a vulnerability that are unique to a user's environment.

The Base metrics produce a severity score ranging from 0 to 10, 10 being the most severe. This score can then be modified by combining the Temporal and Environmental metrics. For example, a vulnerability that might have a severe impact on a device may be downgraded

because the exploit required to take advantage of this vulnerability is just too complex to be widely distributed. However, while this method was the authors' original intent for the CVSS scores, most companies only use the Base Score.

Figure 2: CVSS Metrics and Equations

Generally, the Base (always available from NVD) and Temporal metrics are specified by vulnerability bulletin analysts, security product vendors, or application vendors because they typically possess the most accurate information about the characteristics of a vulnerability. End-user organizations specify environmental metrics because they can best assess the potential impact of a vulnerability within their own computing environment.

The scores are computed in sequence. The Base Score is used to calculate the Temporal Score, and the Temporal Score is used to calculate the Environmental Score (see Figure 2 and the calculator here: https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator).

Temporal metrics are very hard to track because it is a manual-intensive process and requires analysts to be constantly aware of exploit availability and maturity, remediations available, and the reliability of available information.

Scoring CVSS metrics also produces a vector string, a textual representation of the metric values used to score the vulnerability. This vector string is a specifically formatted text string that contains each value assigned to each metric and should always be displayed with the vulnerability score.

Because of its complexity, especially the need for analysts to be constantly aware of a wide variety of threat data, CVSS cannot be used at its full capacity by the majority of the community. Instead, it needs to be combined with data-driven threat information, like EPSS, to better prioritize vulnerability remediation efforts.

Introduction to EPSS

The Exploit Prediction Scoring System (EPSS https://www.first.org/epss) is an open, data-driven effort for estimating the likelihood (probability) that a software vulnerability will be exploited in the wild. Its goal is to assist network defenders in better prioritizing vulnerability remediation efforts in conjunction with an existing CVSS score.

As explained above, while other industry standards such as CVSS have been helpful in capturing the innate characteristics of a vulnerability and providing measures of severity, they are often limited in their ability to predict the likelihood of a threat occurring. EPSS fills that gap because it uses current threat information from the CVE database (https://www.cve.org/ResourcesSupport/FAQs) combined with real-world exploit data for its predictions. EPSS then produces a probability score of between 0 and 1 (0 and 100%). The higher the score, the greater the probability that a vulnerability will be exploited in the next 30 days.

EPSS currently collects information from multiple data sources, usually daily, to score threats. These sources include, but are not limited to, the following:

- MITRE's CVE List - Only CVEs in the "published" state are scored
- Text-based "Tags" derived from the CVE description and other sources talking about the vulnerability
- Count of how many days a CVE has been published
- Count of how many references are listed in the CVE
- Published Exploit code in any of: Metasploit, ExploitDB, and/or Github
- Security Scanners: Jaeles, Intrigue, Nuclei, sn1per
- CVSS v3 vectors in the base score (not the score or any subscores), as published in the National Vulnerability Database (NVD)
- CPE (vendor) information, as published in NVD
- Ground Truth: Daily observations of exploitation-in-the-wild activity from AlienVault and **Fortinet**.

The process that generates this score is based on machine learning. A simplified representation is found in the image below:

Figure 3: Representation of score generation through machine learning

As can be seen, any EPSS deployment must first go through a training phase, where it is given historical vulnerability data and daily exploitation activity to develop and refine its predictive model. Once established, the predictive model can then be used in combination with vulnerability metadata on a daily basis to predict future exploitation activity.

The resulting EPSS score can be used for such things as prioritizing which software to patch based on a threshold, with the advantage of requiring organizations to patch fewer vulnerabilities compared to using a patching strategy based solely on a CVSS ranking. There is too much math involved to describe the process here, but you can go to this site to learn more (https://www.first.org/epss/model). But the takeaway is that organizations that properly deploy and train EPSS will need to patch fewer vulnerabilities than they would have using the classic CVSS strategy while maintaining the same level of protection.

You are warned, though. EPSS should never be treated as a risk score. Other factors, such as how accessible vulnerable assets are to attackers, the type of weakness a vulnerability presents, the asset's purpose and value, etc., are all factors to consider when prioritizing which vulnerabilities should be addressed. Fortunately, those factors are typically contained in the string vector of the CVSS score.

Practical Example: Log4j

I will now demonstrate how EPSS tracked the probability of exploitation for the infamous Log4j vulnerability that took the Internet by surprise this last December. The following annotated chart shows how the EPSS score evolved over time based on specific events that affected the dependent variables of the model.

Figure 4: Log4Shell through the eyes of EPSS

The official CVE, CVE-2021-44228 was published on December 10th, 2021. The NVD entry is available here (https://nvd.nist.gov/vuln/detail/CVE-2021-44228).

Figure 5: NVD Analysis

Security analysts provided the initial CVSS scores for v2 and v3 on the 13th of December, with a Base score of 10.0 (critical). That has not changed as of the writing of this blog.

Figure 6: NVD Analysis

EPSS published its daily scores on the morning of December 11th. Even before the CVSS score was assigned on the 13th, it predicted the likelihood of exploitation at 0.355. In terms of all other vulnerabilities scored on that day, Log4j scored higher than 96.8% of all CVEs.

You can see how the EPSS score changed over time based on variations of the 1,164 variables of the model. The last change was when Metasploit added the module “exploit/multi/http/log4shell_header_injection”. That pushed its score to 0.944, the highest it received.

Once again, it is important to note that the CVSS base score has not changed, so organizations that relied solely on that score would not have detected this spike in risk. And while, technically, the temporal CVSS score should have changed to reflect this difference, there are no public tracking services for the temporal component.

The chart below shows our Fortinet telemetry following the release of the CVE-2021-44228 on 9 December 2021. We began receiving telemetry on 10 December 2021 when our IPS engines were set to passively monitor the CVE.

The spike in detecting this signature on 13 December caused our IPS engine to be upgraded to blocking action, ensuring that our customers were automatically protected. You can see the very fast growth spike, reaching a peak of 31 million events on 15 December 2021, followed by a second wave and a sudden decrease from 29 December 2021.

Figure 7: Log4j events

You can also observe that the vulnerability spread fast across the globe, reaching a peak of 207 of the 248 countries listed in our database (this list contains official countries and small independent areas, etc., as well).

There are a few important observations about this chart: The first is that it represents a volumetric count of all the FortiGate firewalls in our network reporting this threat telemetry. And second, we can't confirm whether these FortiGates were fronting an actual application powered by Log4j. This means that we can't be sure that of those counts, how many could have been successful attempts of exploitation. However, we can clearly see the massive jump on 14 December, followed by a small trail of activity.

We can compare the EPSS scores and the volume of detections together by normalizing the magnitude, showing that the EPSS score started at 0.355 on 10 December, indicating that there was a 35% chance in the next 30 days of exploitation of the vulnerability, and looking at that time frame we can see that the subsequent activity ranged from 40,000 events to 31 million events.

Figure 8: EPSS scores and detection volume

Less intuitive is that when EPSS finally reached its maximum score of 94%, the exploitation volume had already subsided. However, this is perfectly fine because the EPSS score does not attempt to predict the volume of events but rather the binary decision of whether or not it will be exploited.

The reader may also notice the sudden drop in volume on the last day of December. This is quite common. We often see this when attackers see diminishing returns on a campaign and switch to exploit other vulnerabilities. In fact, we can see this using another comparative plot, shown below. As the orange RCE (remote code execution) was decreasing in popularity, another one in green—which was a DOS vulnerability (denial of service)—initiated considerable activity, which then faded out around the 6th of January 2022.

Figure 9: All Log4j Vulnerabilties

In conclusion, organizations should consider EPSS as "pre-threat intel." Of course, if they have intel that something is being exploited (via their own telemetry sensors or OSINT), then they should use that as an indication of activity in the wild. But for those without any evidence of exploitation or that lack threat intel, then EPSS is a great fit.

How to use both CVSS and EPSS for prioritization?

Any given company should have a database of vulnerabilities reported daily by their security vendors. For each vulnerability discovered, they can extract the CVE score (the base score, as mentioned before, is available for free from NVD, and paid vendors provide the temporal/environmental scores) and the EPSS score (via a simple API). They can then be plotted with a scatter plot layout, as shown below.

Figure 10: EPSS score compared to CVSS Base Score (NVD)

The X-axis represents the CVSS version 3.1 score (1 to 10), and the Y-axis represents the EPSS score (0 to 1.0). The color of the dots indicates how many vulnerabilities are concentrated in the same spot, so darker colors indicate more, and lighter colors indicate less.

Vulnerabilities in the bottom left represent those that have both a lower probability of being exploited and would incur a lower severity impact to the information system. They can therefore be deprioritized.

Vulnerabilities in the upper right quadrant, on the other hand, are the most critical kinds of vulnerabilities as they are both more likely to be exploited and could fully compromise the information system. They should therefore be patched first.

The best strategy to remediate those vulnerabilities is to rank them from the top right to the bottom left, following the blue line. The defender can now spend fewer resources while patching more vulnerabilities that are much more likely to be exploited.

Summary

This quick table shows the key feature points of CVSS vs. EPSS.

	CVSS	EPSS
Score Range	0 to 10	0 to 100%
Score Meaning	Severity of a vulnerability	Probability of a vulnerability to be exploited in the next 30 days
Maintainer	FIRST ORG	FIRST ORG
Dependencies	Security analysts to compute base, temporal, and environmental score	A variety of sources, including the CVSS base score from NVD
Update	Base scores are provided by NVD when a new vulnerability is published (plus a few adjustments for errata)	Scores are updated daily via an ML-based prediction pipeline

Human Annotation Required	Yes, by CNA/NVD	Indirectly, because the machine learning extracts factors from a variety of database sources maintained by the community

How does Fortinet contribute to EPSS?

The ground truth of EPSS is derived from daily observations of exploitation-in-the-wild activity provided by AlienVault and Fortinet.

For each vulnerability, FortiGuard Labs provides the daily count of detections (anonymized) from our network of FortiGates distributed around the world. This is critical information for building an accurate machine learning predictor for the exploitability score.

FortiGuard Labs also contributes to the codebase of the EPSS API in the official FIRST repository.

FortiGuard Labs is Driving the Future of CybersecurityFortiGuard Labs has been active in the threat intelligence industry since its founding. We contributed to developing the STIX/TAXII protocols with MITRE and OASIS, starting back in 2012. In 2014, we co-founded the Cyber Threat Alliance, which focuses on threat intelligence and information sharing. In 2018, we began to work with the World Economic Forum's Centre for Cybersecurity on a series of threat intelligence projects, including the Cybercrime ATLAS. And for the last couple of years, FortiGuard Labs has been working with MITRE on a variety of projects to increase the awareness and capabilities of defenders, particularly as a leading contributor to MITRE Engenuity Center for Threat-Informed Defense.

FortiGuard Labs has also been working with the Center for Threat-Informed Defense and its members on a project called ATT&CK Sightings Ecosystem, which includes the "2021 ATT&CK Sightings Report." As quoted in the report, this effort "created a picture of common adversary behavior, including which techniques adversaries use, how their use changes over time, and how adversaries sequence techniques. Defenders can use this information to create a threat-informed defense against what they are most likely to see, not just the latest cyber threat headlines."

Fortinet also provided data sets to help train the model for the Exploit Prediction Scoring System (EPSS) project and FortiGuard Labs has been an active participant. EPSS is an open model for predicting the likelihood that a vulnerability will be exploited in the wild. Coupled with the TTP (tactics, techniques, and procedures) data we are providing through the Threat intelligence Insider, we're building the future of cybersecurity protections and decision making.

References:

https://www.first.org/epss/model

https://www.first.org/epss/articles/log4shell

https://www.first.org/epss/articles/prob_percentile_bins

https://www.first.org/epss/user-guide

Learn more about Fortinet's FortiGuard Labs threat research and intelligence organization and the FortiGuard Security Subscriptions and Services portfolio.