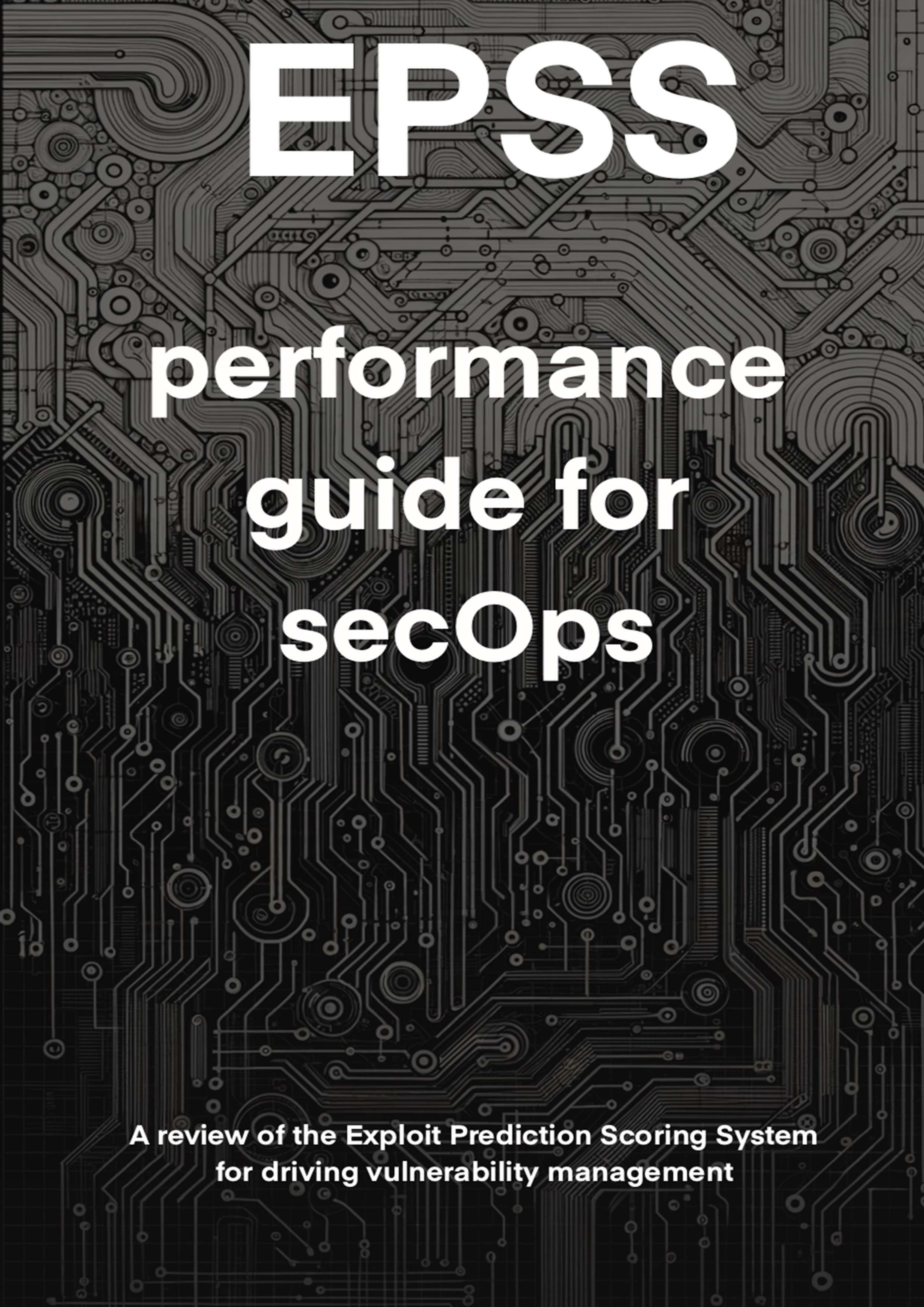


EPSS performance guide for secOps

A review of the Exploit Prediction Scoring System for driving vulnerability management

Introduction

For each tracked CVE, The Exploit Prediction Scoring System (EPSS) v3 calculates a daily probability of exploit within the next 30 days.

Based on these premises, EPSS authors[1] suggest an optimal process for prioritizing vulnerabilities management: by remediating only 14000 vulnerabilities (about 7.3% of the whole), corresponding to an EPSS score of 8.8% or higher, they claim to get a coverage of about 82%.

It means that whenever an exploit is published, its likelihood to belong to the 14000 vulnerabilities is 82%.

Our Work

In October 2023, we performed an independent survey of this claim, the results of which are presented in this document. The insights are structured as follows:

- highlights for secOps;
- methodology of the survey;
- comprehensive CVE and EPSS data analysis.

Takeaways for secOps

Takeway #1: Age At Exploit

In our investigation, we found that the coverage is strongly dependent on the vulnerabilities' Age At Exploit (A@E):

1. Long vulnerabilities (A@E greater than one year) have a perfect coverage of 100%
2. Medium vulnerabilities (A@E between 1 year and 1 month) have a coverage of 56%
3. Short vulnerabilities (A@E less than a month) enjoy the poorest coverage, at 40.7%

Takeaway #2: Overfitting

Nearly all long vulnerabilities in our sample are overfitting, meaning not a probability. Since their coverage is so good, however, their rating is extremely valuable and must be considered as a filter.

Medium and short vulnerabilities are nearly never overfitting, this reinforces the importance of handling long vulnerabilities differently from medium and short ones.

Takeaway #3: Lead time

The lead time (LT) is the number of days between a CVE being tracked by EPSS with a non-overfitting score greater than 8.8% and the exploit publication.

In our sampling, we found that:

1. It often takes a long time before EPSS starts tracking a CVE once it is published, therefore reducing the LT;
2. The LT of long vulnerabilities usually doesn't make sense, due to overfitting
3. The LT of medium vulnerabilities is about 34 days
4. The LT of short vulnerabilities is just below 6 days

Takeaway #4: recommendations for prioritizing short vulnerabilities

The low coverage (40.7%) and limited lead time (6 days) of short vulnerabilities suggest it might not be worth it for many companies to use EPSS for prioritizing short vulnerabilities.

A lighter option could be to just track publications from exploit sources.

Takeaway #5: recommendations for prioritizing medium vulnerabilities

The decent coverage (55%) of medium vulnerabilities, even if well below the EPSS authors' estimate (82%), is a useful metrics for prioritization, because a lead time of 34 days provides enough leeway for teams to remediate efficiently.

Takeaway #6: recommendations for prioritizing long vulnerabilities

For long vulnerabilities, we agree with EPSS authors that a score of 8.8% or greater is reasonable when used as a filter to populate the remediations set.

Still, the secOps are left with a whopping 14000 unsorted vulnerabilities to remediate. This set is unordered because long vulnerabilities, which are quite numerous, are prone to overfitting (their rating doesn't make statistical sense). It would be extremely beneficial for the secOps community if the overfitting issue was overcome in a next version of EPSS, so that the remediation set could be at least partially sortable.

Findings [scope and methodology]

We conducted our survey by sampling all exploited vulnerabilities from Exploit-DB over a period of 2 months, . We counted how many were true positives (i.e. how many belonged to the predicted remediation set).

A CVE is a true positive if and only if the 3 conditions hold true:

1. The CVE was published in Exploit-DB during the sampling period
2. EPSS started tracking the CVE before exploit (seems obvious but many CVEs from our sample do not match this criteria)
3. The EPSS score at age of exploit is greater or equal to 8.8%

A CVE is a false negative if and only if the 2 conditions hold true:

1. The CVE was published in Exploit-DB during the sampling period
2. EPSS didn't start tracking the CVE before exploit OR the score at age of exploit was lower than 8.8%

We calculated the coverage sticking to EPSS authors definition: the ratio of true positive over the sum of true positives and false negatives

The reason we chose Exploit-DB as a source of exploit was that, according to EPSS authors, it makes for the majority of exploits from which EPSS was trained.

Note that, from a secOps perspective, a 2 months period is quite long: when we sample the behavior of a firewall or an IDS to determine their « efficiency in the field », we use similar time ranges.

Findings [global coverage]

70 exploited vulnerabilities were published in Exploit-DB during the reviewed period:

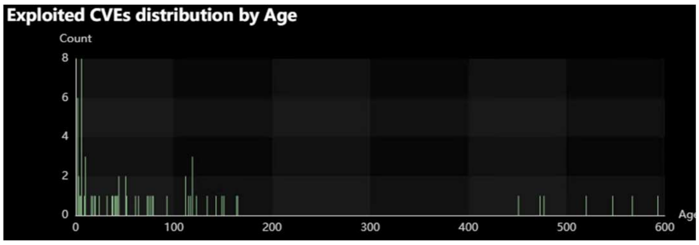
- 39 exploits were true positives
- 19 were not tracked before exploit
- 12 had an EPSS score below 8.8% at age of exploit

The global coverage is therefore: 39/(39+19+12)=55.7%

Findings [split by age at exploit]

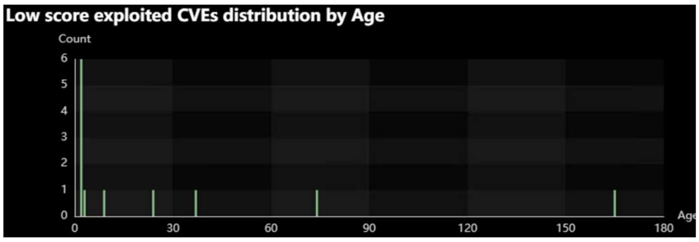
Digging more into the data set and looking at the distribution of exploited CVEs by age, we noticed a large empty gap between long vulnerabilities and shorter vulnerabilities:

- A set of vulnerabilities aged 451 days or more
- Another set of vulnerabilities aged 165 days or less.



This prompted us to make 2 separate calculations: one for each set.

When inspecting the set of younger vulnerabilities, we noticed that many vulnerabilities younger than a month were false negatives because they enjoyed an EPSS score lower than 8.8%.



So, we broke down this second set into 2 subsets.

We ended up with 3 sets.

Long-term set: vulnerabilities older than a year

This set features 9 exploited CVEs, all of which are true positives.

So the coverage for the long-term set is 100%.

Medium-term set: vulnerabilities aged 1 month to 1 year

This set features 34 exploited CVEs,

- 19 are true positives
- 12 were not tracked before exploit
- 3 were not scored above 8.8%

The coverage for the medium-term set is: 19/(12+12+3)=56%

Short-term set: vulnerabilities aged at most 1 month

The last set contains 27 vulnerabilities:

- 11 are true positives
- 7 were not tracked before exploit
- 9 were not scored above 8.8%

The coverage for the short-term set is: 11/(11+7+9)=40.7%

Findings [Overfitting]

Overfitting happens when a prediction model is having a hard time to generalize from its training set.

EPSS v3 was trained using all vulnerabilities from November 1st, 2021 to October 31st, 2022.The activity included in the training cover July 1, 2016 to December 31st, 2022 [1].

Since then, vulnerabilities have "grown old".

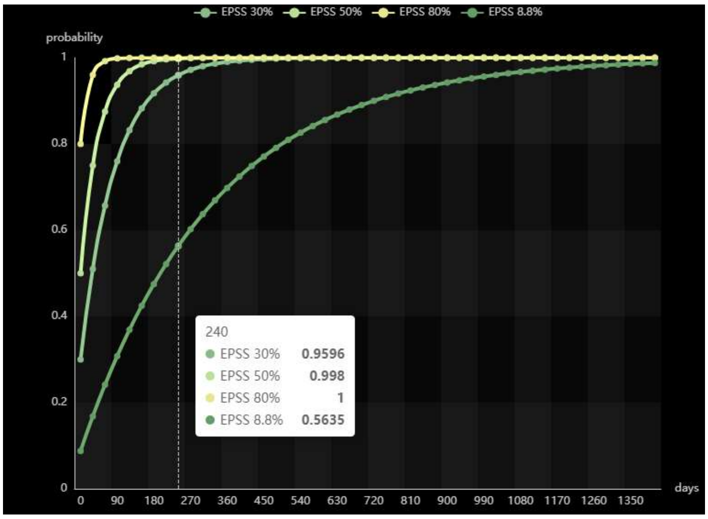
To measure the impact of time in the quality of the prediction, let's consider 4 vulnerabilities with EPSS probabilities of 8.8%, 30%, 50% and 80% respectively. Suppose these probabilities either remain constant or increase over time.

Now let's be conservative and assume the probabilities remain constant. It means that every 30 days, we get the same EPSS probability.

Under these assumptions, if p is the initial EPSS probability of a CVE being exploited within 30 days, the probability of a CVE being exploited within 60 days will be: 1-(1-p)^2

We can generalize and find the probability of a CVE being exploited within nx30 days: 1-(1-p)^n

Leveraging this, let’s plot the long-time evolution of the 4 CVE exploit probabilities:



In the white square from the graph above, we see thatn after 240 days (or n=8), 3 in 4 CVEs have a probability of exploit that is greater than 95%. Only the CVE with lowest initial EPSS score is still only 56% of being exploited.

Based on this graph, we have 3 examples of overfitting: unexploited CVEs older than 240 days with an initial EPSS probability of 30% or more are overfitting with 95% confidence. Their initial EPSS probability is not to be taken for granted.

During the reviewed period, we found that:

- No short vulnerabilities were overfitting (not a surprise)
- 1 medium vulnerability was overfitting with 95% confidence (CVE-2023-1389)
- 9 of 10 long vulnerabilities were overfitting.

Findings [Lead time]

Overfitting is an important metrics not for calculating coverage, but for calculating lead time and prediction power.

Lead time and prediction power don’t make sense for a CVE featuring an overfitting EPSS score.

Data set details

The reviewed data set is available from GitHub:

https://github.com/labyrinthinesecurity/EPSS/blob/main.EPSSv3_2023.csv

Reference

[1] “Enhancing Vulnerability Prioritization: Data Driven Exploit Predictions with Community Driven Insights”, by J. Jacobs at al, February 2023: https//arxiv.org/abs/2302.14172

Study Author

This studied was conducted by Christophe Parisel (linkedin.com/in/parisel)