A VISUAL EXPLORATION OF EXPLOITATION IN THE WILD

THE INAUGURAL STUDY OF EPSS DATA AND PERFORMANCE

JULY 2024

INTRODUCTION

Have you ever heard claims about a security product or model that never made any attempt to validate whether actual performance lived up to those claims? Nah, neither have we. That's obviously a joke—not backing our mouths with measurement is the norm in our field. But we don't want to follow that script with the Exploit Prediction Scoring System (EPSS).

For years now, we've been collecting evidence of exploitation activity from data contributors. This data was used to train the EPSS model that produces the daily scores that are freely available to the security community. With the passage of time, we now have a rich history of predictions that we can test with the benefit of hindsight.

This inaugural study seeks to evaluate EPSS performance over the last few years. In addition, we tackle a host of questions related to understanding the ins and outs of vulnerability exploitation in the wild. We hope it offers measured KPIs for EPSS as well as valuable insights for using it to manage and prioritize vulnerabilities in your environment.

About EPSS

EPSS is a data-driven effort for estimating the likelihood (probability) that a published vulnerability will be exploited in the wild. Its goal is to assist defenders to better prioritize vulnerability remediation efforts. While other industry standards have been useful for capturing innate characteristics of a vulnerability and provide measures of severity, they are limited in their ability to assess threat. EPSS fills that gap by using current threat information targeting CVEs along with real-world exploit data. The EPSS model produces a daily updated prediction of the probability that a given vulnerability will be exploited in the next 30 days.

A growing set of organizations contribute data to EPSS (and you can join them!). The Cyentia Institute developed the EPSS model and crunches the data to generate the daily scores. First.org hosts the EPSS SIG and makes the data available to the community. Find out more at https://www.first.org/epss/

The Cyentia Institute and FIRST.org have made all of the charts in this report avalible for download. These resources provide valuable insights into vulnerability exploitation patterns and EPSS performance.

Access and download the full set of charts here.

TABLE OF CONTENTS

INTRODUCTION FROM EPSS CREATOR & FIRST.ORG SIG CHAIR 4
ACKNOWLEDGMENTS 5
SPONSOR COMMENTS 6
EXPLOITATION ACTIVITY 8
How many vulnerabilities have been published? Exploited? 10
What proportion of vulnerabilities have been exploited? 12
Does exploitation activity fluctuate over time? 14
What's the typical pattern of exploitation activity? 16
What's the ratio of new vs. old exploitation? 18
How long since exploitation was last observed? 20
How long until exploitation was first observed? 22
How "old" is current exploitation activity? 24
How widespread is exploitation among organizations? 26
EVALUATING EPSS PERFORMANCE 28
How do we evaluate exploit predictions? 30
How does CVSS perform? 32
How does the KEV perform? 34
Can metadata help predict exploitation? 36
Can exploit tools help predict exploitation? 38
How does EPSS perform? 40
How do EPSS and CVSS compare? 42
What EPSS score warrants priority remediation? 44
CONCLUSION: WHAT'S NEXT FOR EPSS? 46
APPENDIX: EPSS OVERVIEW & HISTORY 48

OPENING THOUGHTS

From EPSS Creator Jay Jacobs

I've been fortunate in my career to have worked with some very interesting data sets. Data often surprises me and challenges many commonly held beliefs across the security industry. But more importantly, they can generate opportunities to learn if we are ready to do so. This is one of those opportunities.

The opportunity to learn generally comes in only one form: feedback. If we want to learn how to play better golf we hit a golf ball and get feedback by watching what happens. While "practice makes perfect", it's actually the feedback we receive while practicing that creates improvement. How quickly would someone improve if they couldn't hear the sound coming out of their instrument? How fast could someone improve their free throws if they couldn't see what happened after the basketball left their hands? The same is true in vulnerability management. When is the last time anyone went back to what was prioritized in the last cycle to collect feedback on their decisions? It generally doesn't happen, but that's exactly what we are doing here.

Now, I don't want to spoil the surprise, but EPSS is not perfect. It will rate some vulnerabilities very low that end up with exploitation activity, and some very high that don't. However, perfection isn't an option for anyone in reality, so EPSS (and every other prioritization strategy) needs to be compared to real and practical alternatives. We explore some of those comparisons in this research with CISA's Known Exploited Vulnerability (KEV) list and the Common Vulnerability Scoring System (CVSS).

We have two major goals in this research. First, we want to investigate all of the exploitation activity we were able to collect and ask some seemingly simple questions. We want to understand everything we can about the timing, volume and prevalence of exploitation activity. As you'll see in the first half of this research, "exploited in the wild" is a relatively meaningless label. Exploitation today does not always mean exploitation tomorrow and me seeing exploitation activity doesn't also mean you'll see exploitation. Exploitation activity is incredibly varied across time, targets and volume and we need better language to talk about it.

Second, we want to collect and analyze feedback on how the Exploit Prediction Scoring System (EPSS) is performing. EPSS generates a score every day for every published vulnerability (with a CVE ID) on how likely it is that we will observe exploitation activity in the following 30 days. Well, EPSS has been publishing scores for over three years now, that's a lot of predictions over many 30 day windows. With the power of hindsight, we can look back at each and every daily prediction and compare against the actual exploitation activity we (our data partners) observed in the 30 day windows following each prediction.

Speaking of data partners, I want to personally thank each and every one of them for their contribution, so in no particular order, thank you to GreyNoise, Shadow Server Foundation, Fortinet, AlienVault, Cisco, F5, Efflux and Cyentia. EPSS would be nothing without their contributions, so please join me in thanking them!

ACKNOWLEDGMENTS

PLATINUM SPONSOR: tenable

GOLD SPONSOR: JupiterOne

GOLD SPONSOR: Nucleus

SILVER SPONSOR: SOFTRAMS

SILVER SPONSOR: VULCAN.

SILVER SPONSOR: KYND

SILVER SPONSOR: APPSALUTE

SILVER SPONSOR: FOSSA

SILVER SPONSOR: ACDS ADVANCED CYBER DEFENCE SYSTEMS

SILVER SPONSOR: edgescan

The Cyentia Institute and FIRST.org offer our thanks to the sponsors of this study. Your commitment to supporting a resource that's increasingly used by organizations around the world is commendable.

As we explore the intricacies of exploits in the wild and assess the efficacy of the Exploit Prediction Scoring System (EPSS), we recognize the invaluable role of community contributions. Your participation in sharing exploitation activity data is crucial for refining our predictive models and enhancing the security landscape. We invite you to join our efforts in advancing the EPSS initiative by becoming a data contributor. Together, we can build a more robust and accurate system that benefits the entire security community. Visit the Cyentia website to learn how you can get involved and contribute to our ongoing projects.

COMMENTS FROM PLATINUM SPONSOR TENABLE

EPSS is an effective input for risk-based vulnerability management.

The Exploit Prediction Scoring System (EPSS), plays a crucial role in the risk formula by providing a predictive measure of the likelihood that a specific vulnerability with a Common Vulnerabilities and Exposures (CVE) identifier will be exploited. EPSS helps organizations prioritize and triage known vulnerabilities based on the likelihood of exploitation. By assigning a probability score to each CVE, EPSS enables security teams to efficiently allocate resources to address the most pressing threats. This targeted approach enhances the overall risk management strategy and ensures the most critical vulnerabilities are addressed promptly.

EPSS is just one input. Understanding context is key.

Despite its strong performance in both coverage and efficiency as noted in this report, EPSS should not be used in isolation for the effective prioritization of efforts in a vulnerability management practice. Environmental and organizational factors outside the scope of EPSS (e.g. asset criticality, network exposure and business impact) are crucial for assessing overall risk. By design, EPSS does not account for the criticality of affected assets, their role in business perations, or their interconnectedness within the network. This focus can lead to misaligned prioritization, where vulnerabilities deemed likely to be exploited are addressed at the expense of those that, while less likely, could have severe consequences if exploited. Integrating EPSS with inputs like threat intelligence, patch availability and compliance requirements offers a more comprehensive risk management approach. EPSS must be used in conjunction with this contextual information to provide a more complete picture and ensure effectiveness in guiding holistic vulnerability management strategies.

Tenable brings it all together with VPR.

As this report highlights, there will always be more risk than you can address in your environment. Focusing on the exposures that matter - we call them the critical few - through an effective vulnerability prioritization strategy is key. EPSS brings hope to the future of vulnerability management by demonstrating that this sea of risk can be drained down to a manageable pond.

Using EPSS as a supplemental input alongside Tenable's proprietary scoring system, the Vulnerability Priority Rating (VPR), sharpens that focus even further. VPR helps organizations improve their remediation efficiency and effectiveness by rating vulnerabilities based on severity level determined by two components: technical impact and threat. Technical impact measures the impact on confidentiality, integrity and availability following exploitation of a vulnerability. The threat component reflects both recent and potential future threat activity against a vulnerability. Examples of such threat sources include intelligence feeds, observations of Indicators of Compromise (IoC), reports of exploitation on social media or code repositories, and more. VPR provides context that is otherwise missing from EPSS. In other words, not only does VPR tell you how bad a vulnerability is, but it tells you why it's bad. Using these scores in parallel provides a much more holistic risk prioritization approach.

INAUGURAL EPSS PERFORMANCE REPORT BY FIRST & CYENTIA INSTITUTE BRINGS HOPE TO THE FUTURE OF VULNERABILITY MANAGEMENT

tenable®

EXPLOITATION ACTIVITY

Before measuring the predictive performance of EPSS, we first analyze our data sources for the exploitation of vulnerabilities. We start with some historical trends and then examine activity patterns, timelines, and prevalence of exploit activity in the wild.

IN THIS SECTION

How many vulnerabilities have been published?

What proportion of vulns have been exploited?

Does exploitation activity fluctuate over time?

What's the typical pattern of exploitation activity?

What's the ratio of new vs. old exploitation?

How long since exploitation was last observed?

How long until exploitation was first observed?

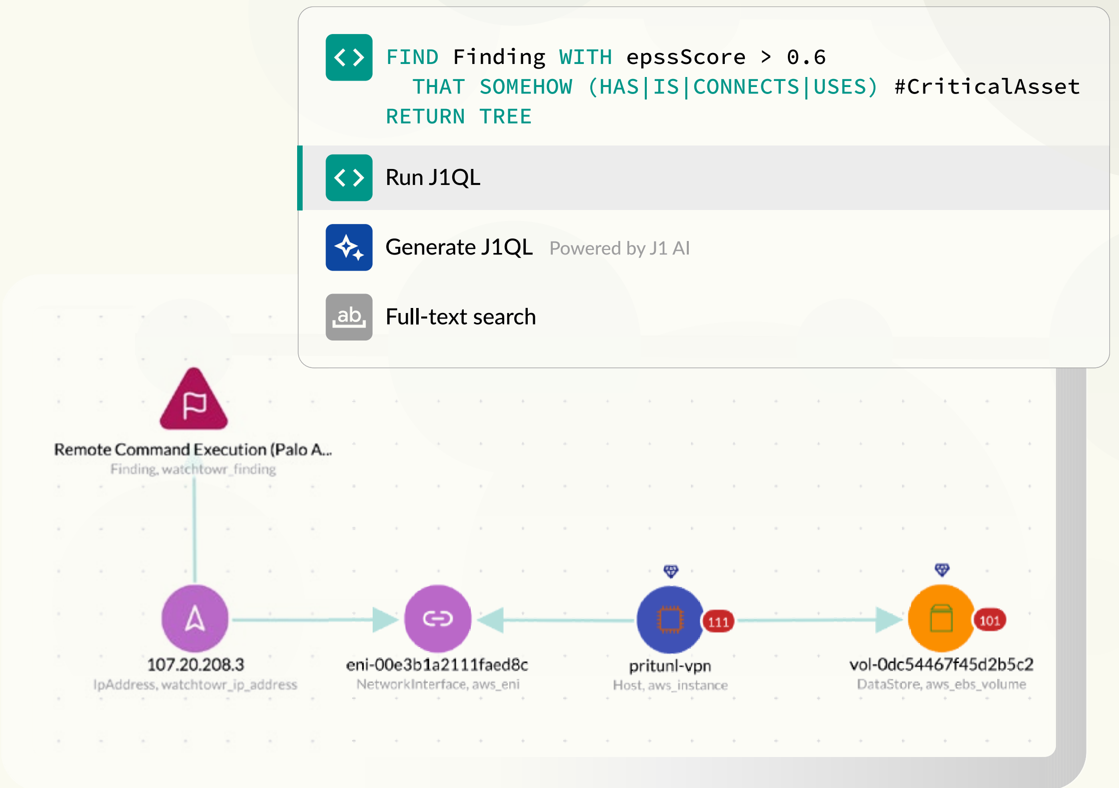
How "old" is current exploitation activity?

How widespread is exploitation among organizations?

> "EPSS is a positive step forward for the industry as organizations now have an independent risk-focused scoring metric to augment the long-standing CVSS severity metrics that have been the underpinnings of many VM programs. Coupled with the contextualization of vulnerability intelligence data as well as the impacted assets, organizations will have the ability to better make true risk-based prioritization decisions that are oriented towards their environments.
>
> **- Luke Tamagna-Darr | Senior Director, Engineering, Tenable**

REMARKS FROM JUPITERONE

Prioritize Effectively: The Power of Time-to-Exploitation Metrics



```
FIND Finding WITH epssScore > 0.6
  THAT SOMEHOW (HAS|IS|CONNECTS|USES) #CriticalAsset
RETURN TREE
```

Run J1QL
Generate J1QL
Powered by J1 AI
Full-text search
Remote Command Execution (Palo A...
Finding, watchtowr_finding
107.20.208.3
IpAddress, watchtowr_ip_address
eni-00e3b1a2111faed8c
NetworkInterface, aws_eni
pritunl-vpn
Host, aws_instance
111
vol-0dc54467f45d2b5c2
DataStore, aws_ebs_volume
101

Traditional vulnerability management approaches often overwhelm security teams with numerous alerts, many of which may not pose immediate threats. For security analysts this can lead to alert fatigue and inefficient use of time and resources. Time-to-exploitation (TTE) metrics address this problem by helping teams focus on vulnerabilities that are most likely to be exploited soon.

As organizations strive to reduce their risk and enhance vulnerability management, incorporating EPSS and TTE metrics alongside traditional vulnerability scores offers a holistic view that integrates severity with exploitation probability.

With JupiterOne and EPSS, eliminate guesswork and focus on what really matters. Many vulnerabilities aren't exploited immediately—don't waste resources on non-urgent patches.

Prioritize effectively, stay secure, and maintain control.

JupiterOne is the asset, attack surface and exposure management platform for security and IT, that empowers organizations to prioritize and remediate what matters most. Continuously monitor exposure with complete visibility across assets and relationships. See out key takeaways at jupiterone.com/epss.

JupiterOne

HOW MANY VULNERABILITIES HAVE BEEN PUBLISHED? EXPLOITED?

Let's begin with the big picture. There's been no shortage of charts created that depict the number of published vulnerabilities over time. But it's an appropriate starting point for this study, so here's one more. We're nearing a quarter million published CVEs, and that's been growing faster in recent years. There are many contributing factors behind this trend, which we can't dig into in this report. Suffice it to say that more vulnerabilities don't necessarily mean the world is less secure; much of this growth is a reflection of changes in the CVE disclosure process.

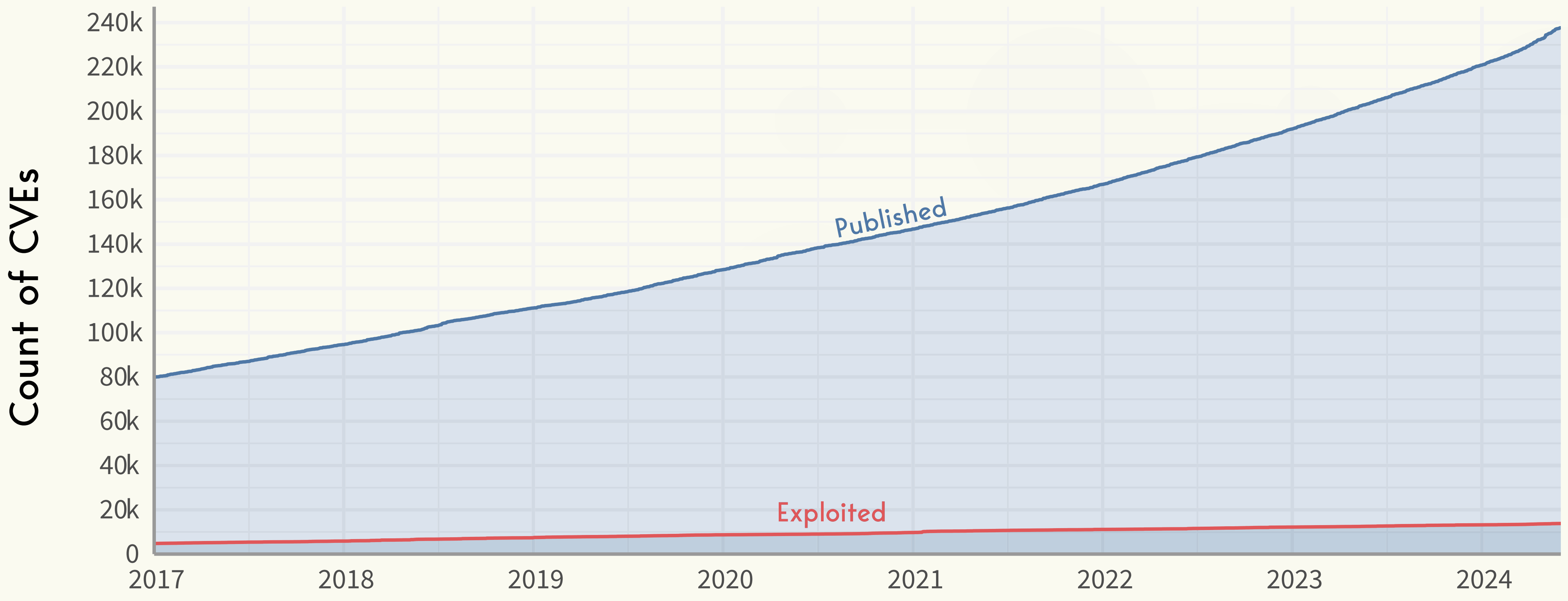
This rising tide of vulnerabilities inundates VM teams with the challenge of assessing and remediating them all. Given the volume of vulnerabilities out there, tracking which ones have been exploited or attacked becomes imperative to managing risk. Per the chart, the number of CVEs known to be exploited keeps rising… though not as quickly as the rate of publication. We'll zoom into that red "Exploited" line next.

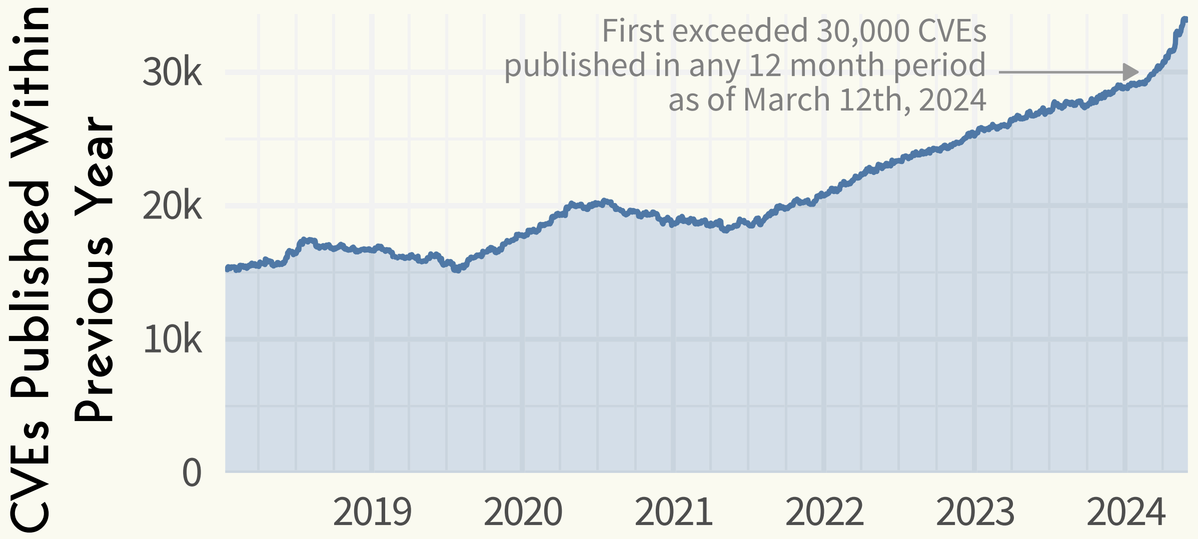
TAKEAWAY: The rising tide of vulnerabilities will overwhelm VM teams if remediation can't be prioritized.

HISTORY OF PUBLISHED AND EXPLOITED CVEs

There were 237,687 published CVEs as of May 31, 2024, with 13,807 being observed with exploitation activity as shown in the top plot. The bottom plots show that we just passed 30,000 CVEs published in the last 12 months with the annual rate varying around the average of 16%.

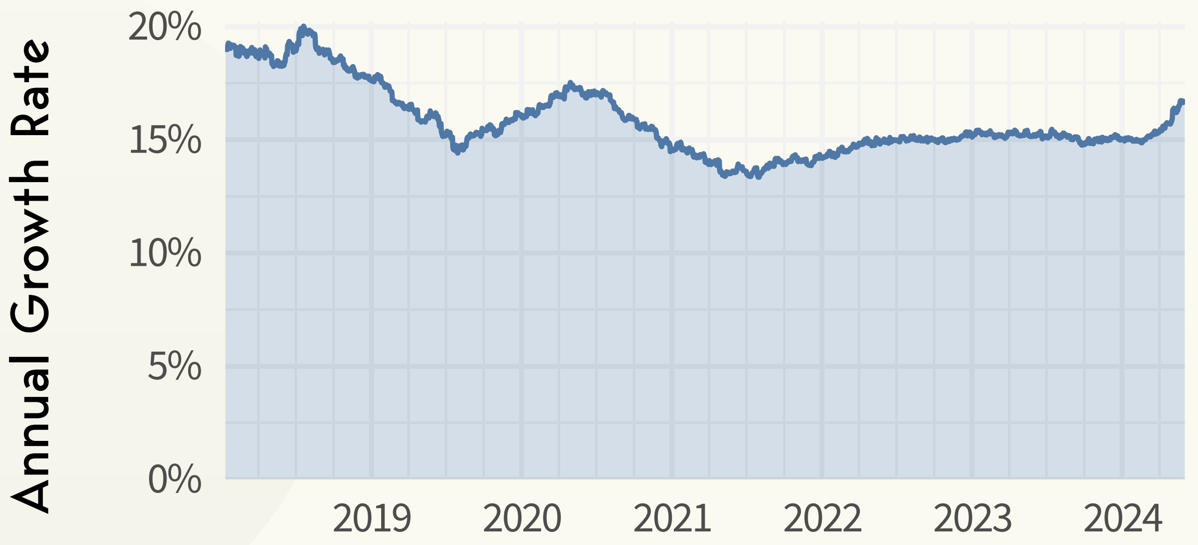
We're nearing a **quarter million** published CVEs.





First exceeded 30,000 CVEs published in any 12 month period as of March 12th, 2024

We'll easily add **30k+** CVEs to the public record during **2024.**



That number has grown at a rate of **16%** annually over the last **7 years.**

WHAT PROPORTION OF VULNERABILITIES HAVE BEEN EXPLOITED?

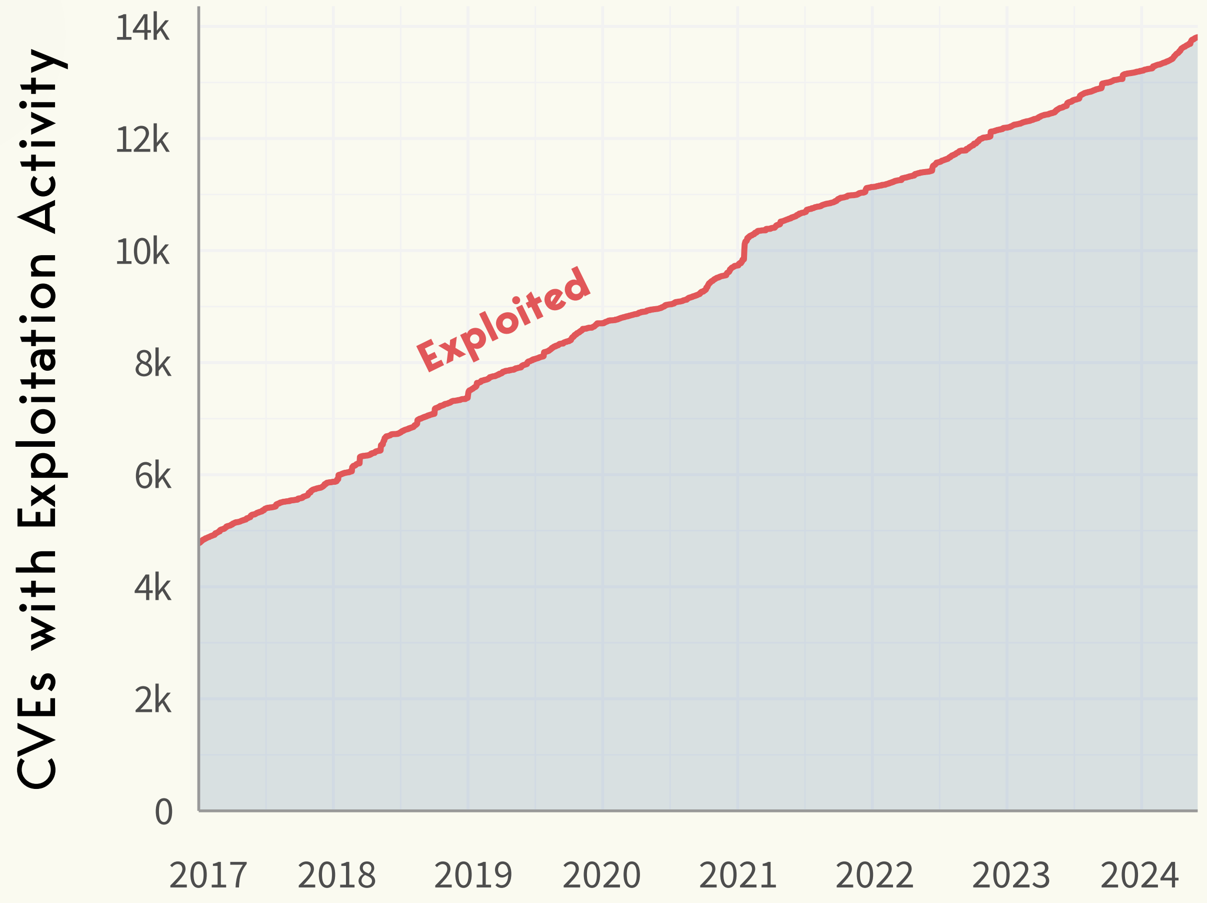
Let's take a closer look at the red exploitation trendline from the previous chart. The top left chart in the figure below shows steady growth in the number of vulnerabilities with known exploitation in the wild. Keep in mind that this doesn't mean that ~14,000 vulnerabilities are actively being exploited right now. It shows that we know of ~14,000 CVEs that have, at some point in their history, been reported as exploited by primary sources. We'll examine the age, duration, and prevalence of exploitation over the next several pages of this report.

While the total number keeps rising, the bottom right chart shows that the proportion of published CVEs known to be exploited remains fairly steady, fluctuating around the 6% mark. The apparent decline over the last few years isn't so much a decline as it may be a delay. As we'll soon see, the majority of vulnerabilities aren't immediately exploited when initially published. It can take time for attackers to discover them and develop exploits and for defenders to detect exploitation activity. Monitoring these precursors of exploitation via its many data contributors is what drives updates to EPSS scores on a daily basis.

TAKEAWAY: Tracking (and predicting) the subset of exploited CVEs is critical for efficient remediation.

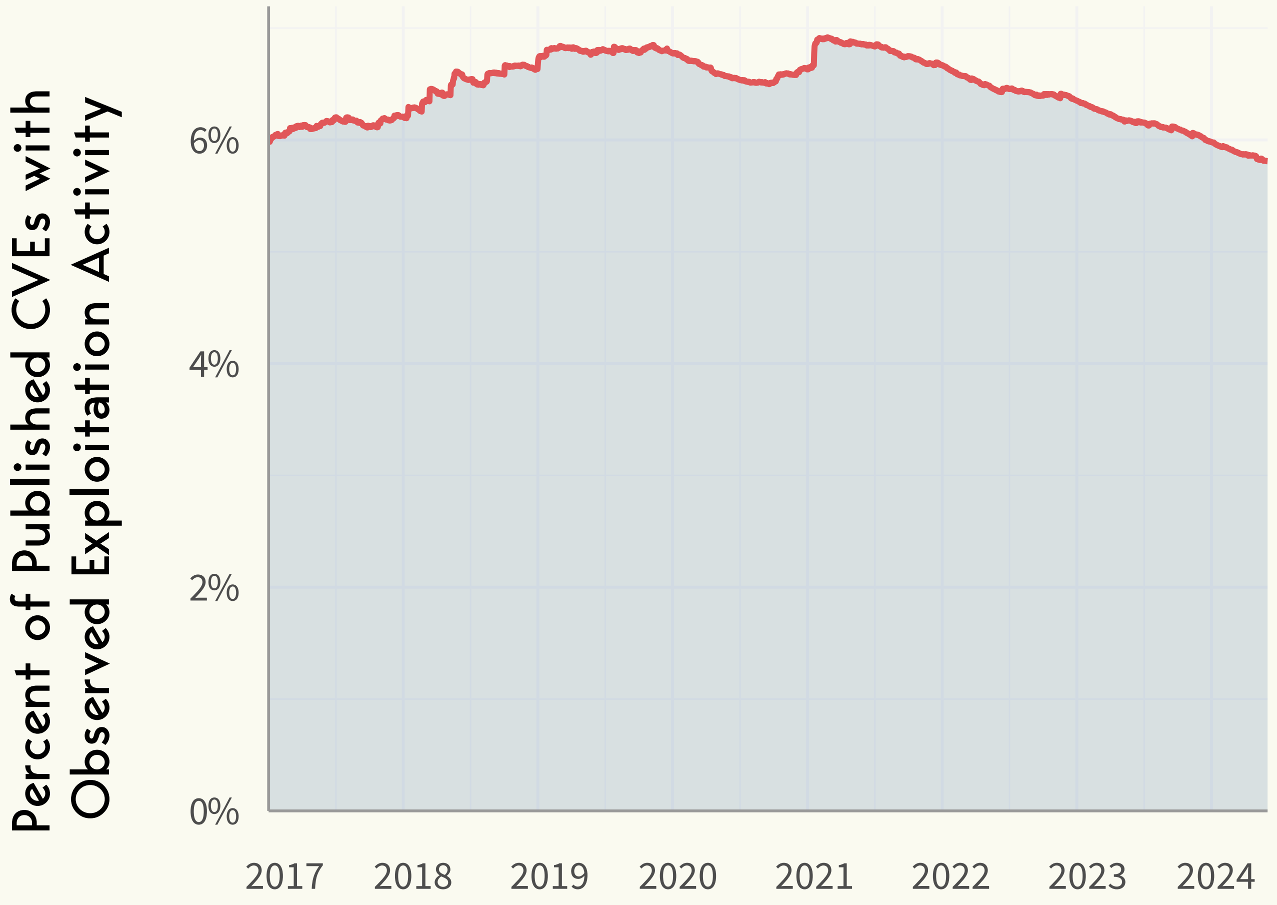
VULNERABILITIES WITH EXPLOITATION ACTIVITY

Newly observed exploitation actvity has been rather steady over the last few years. The top left plot shows the cumulation of 13,807 CVEs with exploitation activity over time, while the bottom right plot shows the count as a percentage of published CVEs over time.



The number of known-exploited vulns is steadily approaching 15k.

About 6% of all published CVEs have been exploited; that rate is holding relatively steady.



DOES EXPLOITATION ACTIVITY FLUCTUATE OVER TIME?

On the prior page, we showed that nearly 14,000 vulnerabilities have evidence of exploitation and caveated that not all of them are actively being exploited right now. That's actually a really important point because many people have the misconception that exploitation is a static or persistent trait. So, we've devoted the next several charts in this report to exploring the ebbs and flows of exploitation activity.

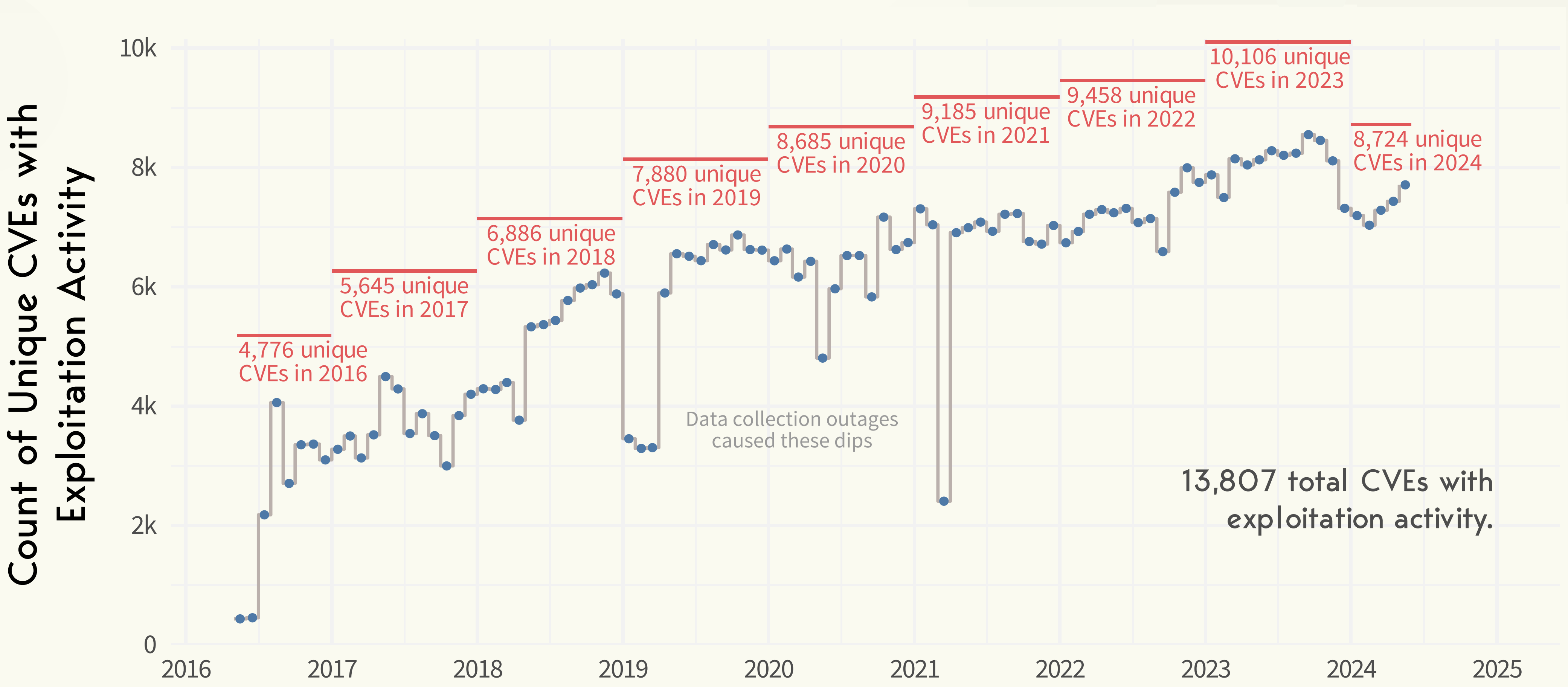
As we've already established, the number of vulnerabilities with exploit activity detected within each year rises over time. But the monthly tally fluctuates quite a bit (sometimes because of data issues). Of the ~14,000 CVEs known to be exploited, about 10,000 had observed exploitation activity in 2023. Thus, there's definitely a temporal element to track and consider when prioritizing vulnerability remediation based on exploitation activity.

What should VM teams do in light of this pattern of sporadic exploitation? The answer has a lot to do with risk tolerance. Risk-averse organizations may wish to take a "once exploited, always exploited" approach to eradicate any vulns with a history of exploitation, however brief. Risk-tolerant or resource-challenged organizations may be best served by prioritizing those exploited recently and/or those most likely to be targeted in the near future. **EPSS provides data to support whatever strategy you choose.**

TAKEAWAY: The number of actively exploited vulns grows as some drop off and others get attacked.

UNIQUE CVES WITH EXPLOITATION ACTIVITY

Counting the unique CVEs with exploitation activity within each month (blue) and within each calendar year (red), there is evidence of sporadic exploitation activity and an indication that once a vulnerability is exploited it may not always be exploited.



Vulns exploited in the past aren't all being attacked right now.

Don't get too excited about the dip in 2024; it's not over yet.

WHAT'S THE TYPICAL PATTERN OF EXPLOITATION ACTIVITY?

What does this fluctuating pattern of exploitation activity look like? Well, that depends on the vulnerability in question. Some vulns are continuously exploited for long periods of time. Some are just a flash in the pan. Exploits of others come in fits and starts. Some real-world examples of these patterns are demonstrated in the chart below, which depicts observed exploitation activity for five CVEs over the course of 2023.

Exploitation of this CVE was short-lived and very sparse.

This one saw fairly regular, albeit sporadic, weekday activity.

Daily to weekly exploit attempts with a spike in mid-Dec.

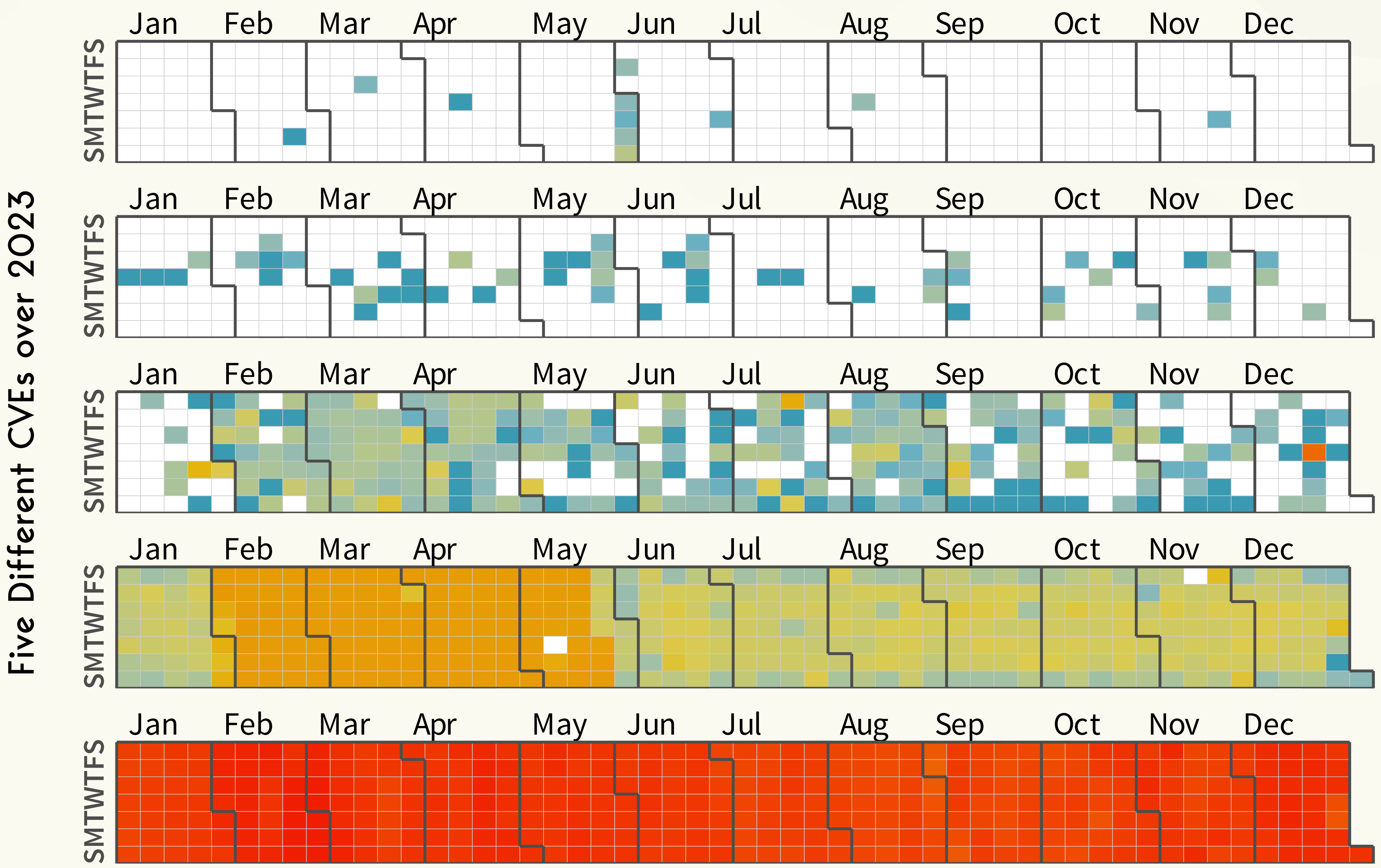
Sustained daily exploitation at its highest in Q1–Q2.

Extremely high rate of unrelenting exploitation activity.

TAKEAWAY: Don't treat "Exploited" as a binary variable; intensity and duration matter for prioritization.

DISPARITY IN OBSERVED EXPLOITATION ACTIVITY

Five out of the 10,106 CVEs with observed exploitation activity are shown here to highlight the volume and variety. Each data source measures "volume" on different scales, so they are normalized here with red representing the highest volume and blue is just a trickle of activity. Not shown is that most of the exploitation activity looks a lot more like the top CVEs than the bottom shown here.



WHAT'S THE RATIO OF NEW VS. OLD EXPLOITATION?

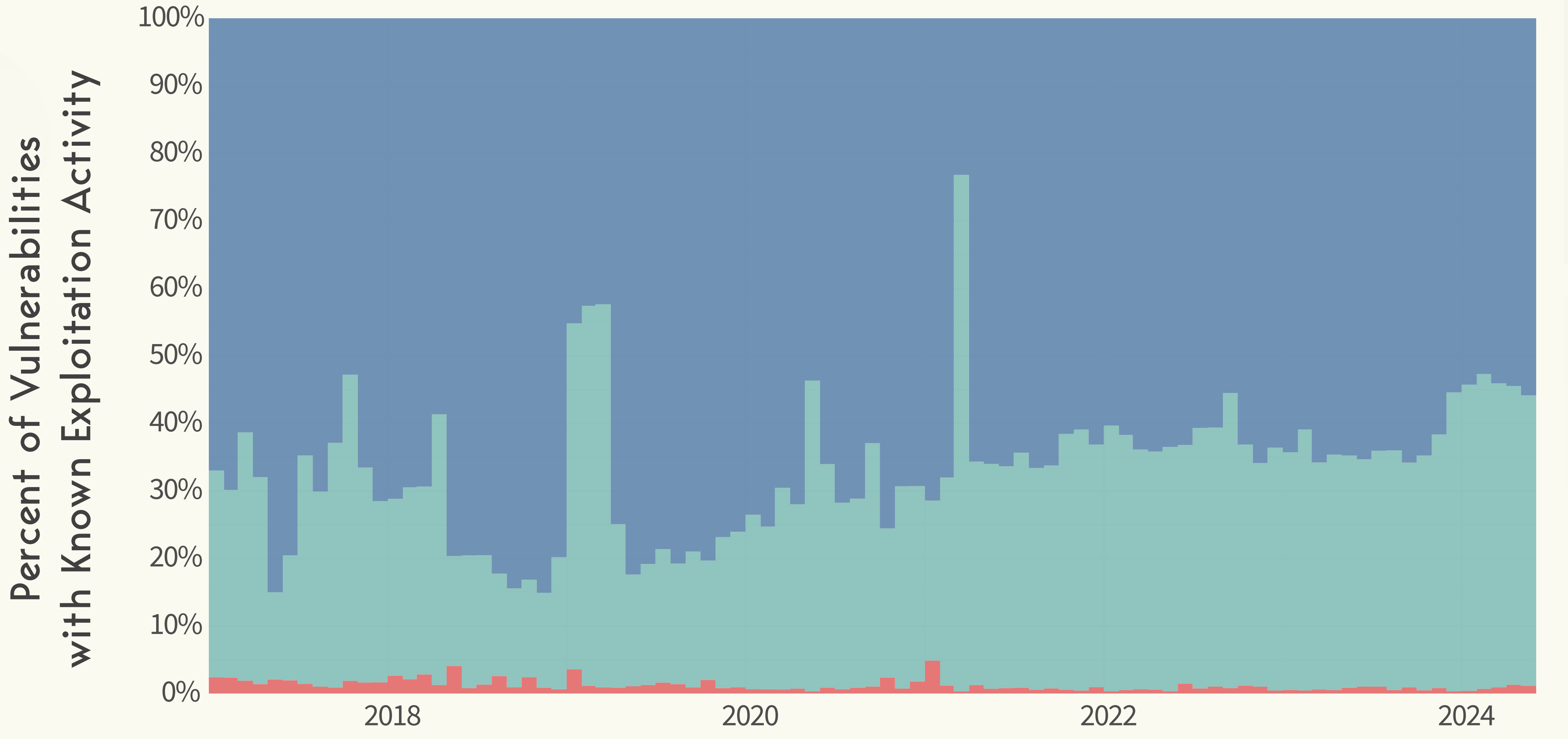
We've seen that exploitation activity targeting vulnerabilities ebbs and flows over time, but what proportion constitutes an ebb vs. a flow? The chart below plots that distinction over the last several years. In it we see that the majority of observed exploitations in a given month flows over from the previous month (represented by the blue area). Also apparent is the third or so of exploit activity that ebbs away—temporarily, at least (the teal area).

You'll also notice a splash of red flitting across the bottom of the chart. That represents net new exploitations that have never before been detected. It's just a fraction of the overall activity, but those are the attacks that keep many VM teams up at night (and sometimes working over the weekend).

TAKEAWAY: Newly exploited vulns get the most attention, but the older ones get the most action.

VULNERABILITIES WITH KNOWN EXPLOITATION ACTIVITY

We break up the monthly exploitation activity into three categories: exploitation activity has been observed this month and observed previously (blue), exploitation activity has been observed before but not in this month (sea green), no previous exploitation activity has been observed (red).



"We've seen these before, and we are seeing them again"

The majority of exploitation activity observed in any given month has been previously reported

"We've seen these before, but not this month"

30-40% of previous exploitation activity is not observed in the current month

"We are seeing these for the first time this month"

A small percent of exploitation activity has never been observed previously

The vast majority of monthly exploitation activity has been seen before.

About a third of previously observed exploitations will periodically go dormant.

HOW LONG SINCE EXPLOITATION WAS *LAST* OBSERVED

Imagine we could take the vulnerabilities represented in the blue and teal areas of the prior chart and more precisely measure how long it's been since they were last exploited.

Good news — no imagination needed. We've visualized it for you!

For half of the nearly 14,000 known exploited vulnerabilities, the most recent detected activity was within the last week. Another quarter of CVEs have been attacked in the last twelve months and the remaining quarter have been dormant for longer than a year. You can triangulate any point on the horizontal and vertical axis to pick out whatever stats you like.

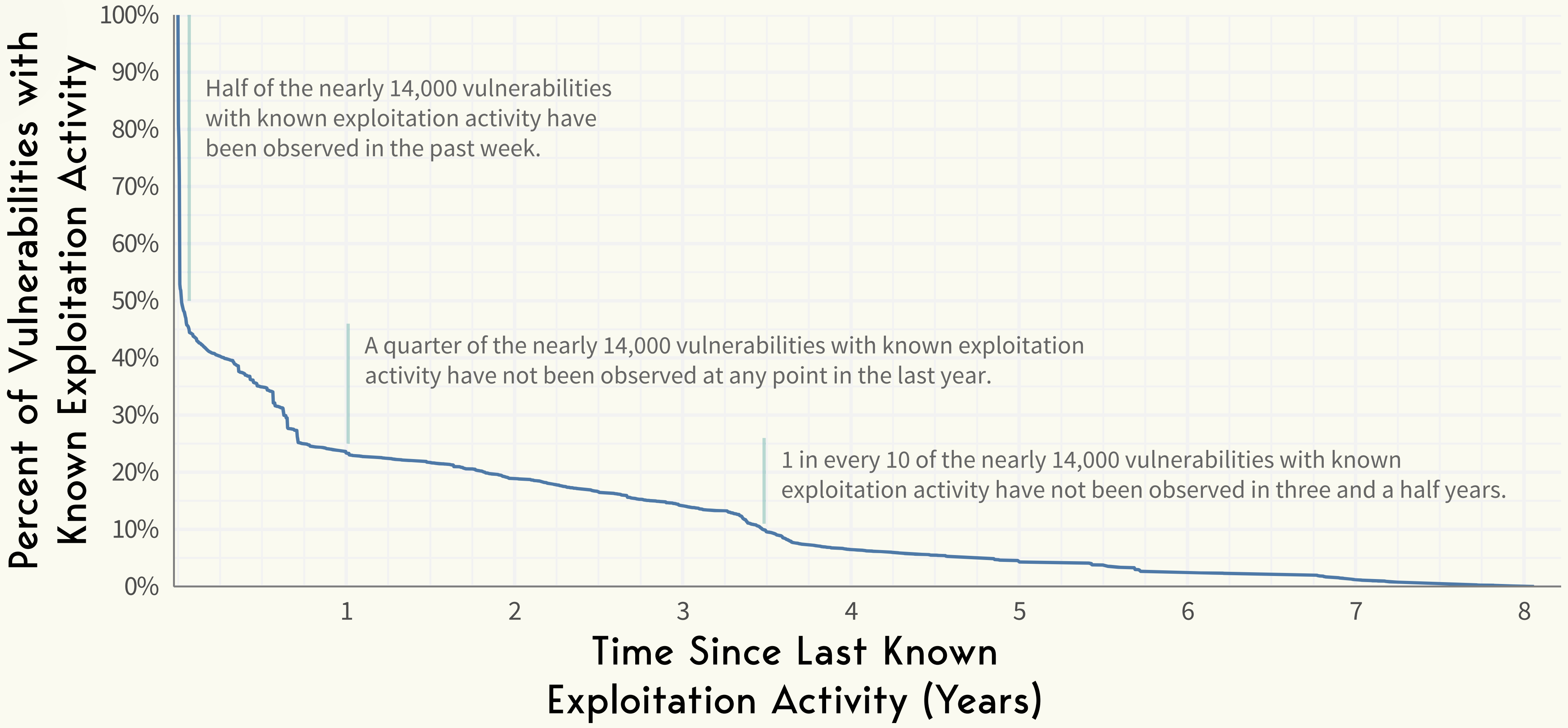
Here's a memorable one: 5% of CVEs had exploitation activity over five years ago and haven't been seen or heard from since.

Some may have a "So what?" reaction here, but look at it this way: how long do you need to keep previously exploited vulnerabilities on your prioritization radar in case they wake up again? This chart can help answer that question and should help you rethink your remediation efforts.

TAKEAWAY: Just because a vulnerability is known to have exploitation activity, doesn't mean it always will.

THE RECENCY OF EXPLOITATION ACTIVITY

Just because something has been reported as exploited in the wild does not mean it will always be exploited in the wild. This chart looks at all of the exploitation activity and how recently vulnerabilities have had observed exploitation activity.



Half of the nearly 14,000 vulnerabilities with known exploitation activity have been observed in the past week.

A quarter of the nearly 14,000 vulnerabilities with known exploitation activity have not been observed at any point in the last year.

1 in every 10 of the nearly 14,000 vulnerabilities with known exploitation activity have not been observed in three and a half years.

Most exploitation activity is a continuation of recent attacks.

It's rare for exploits that haven't been seen in years to flare up again.

HOW LONG UNTIL EXPLOITATION WAS *FIRST* OBSERVED?

What about the flip side of exploitation duration? Once a vulnerability is published with a newly minted CVE, how much time typically passes until detected exploitation in the wild begins? Some people assume that happens immediately, often urging VM teams to drop everything else to remediate “critical” vulns ASAP. Others take the opposite approach and presume there’s plenty of time before they’ll see actual attacks. As with most things, the truth is somewhere in the middle.

Among the ~14,000 known exploited vulnerabilities in our dataset, 8% were targeted BEFORE the CVE was published. Rather than zero days, most of these are “reserved but public” CVEs that, while not officially published, contain information sufficient for them to be incorporated into vulnerability scanners, detection tools, and exploit kits. Within a month of publication, 40% of CVEs observed exploitation in the wild.

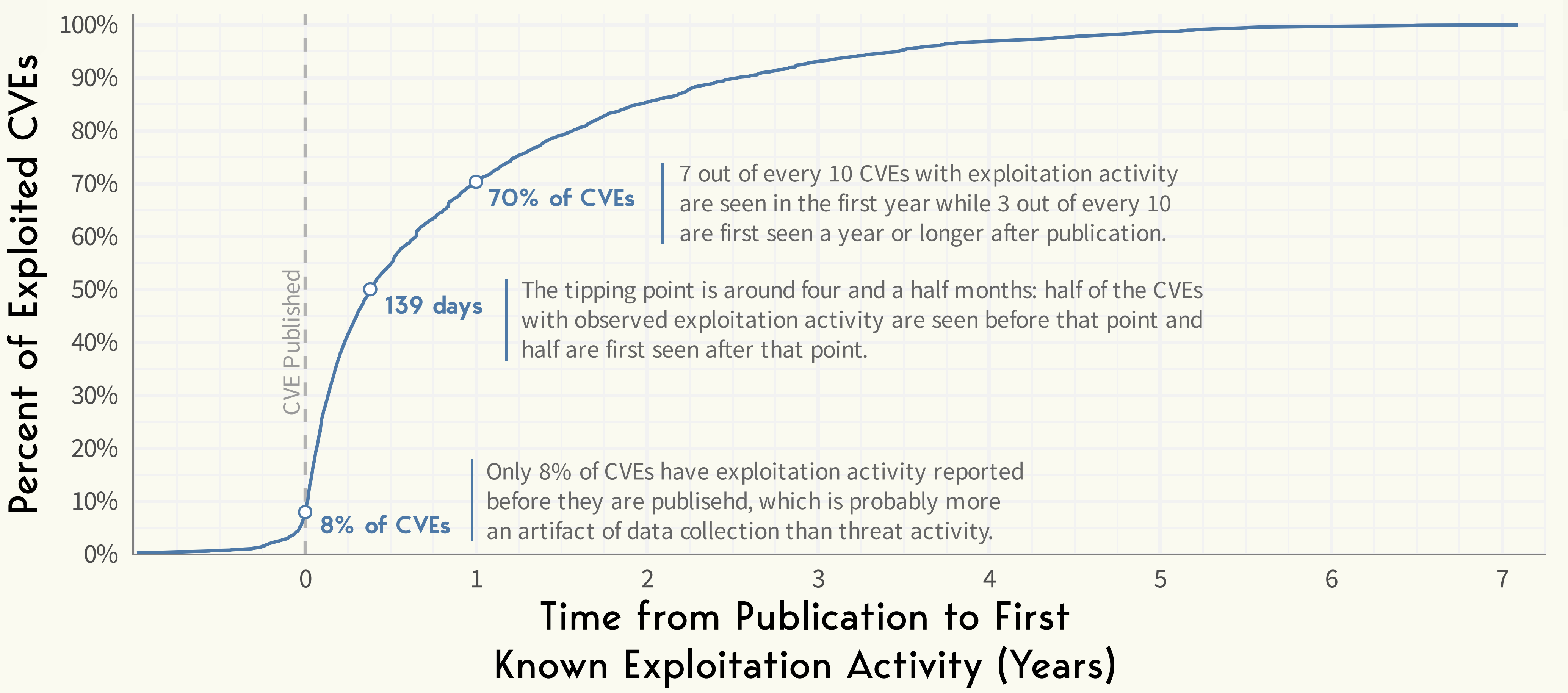
A strong majority (70%) of vulns see initial attack activity in a year or less. It levels out quickly from there. Just 7% of published CVEs go three years before being exploited.

These statistics are all fine and dandy. The rub is in determining how long before THIS PARTICULAR vulnerability is likely to be exploited. That’s where EPSS comes in by helping to remove the guesswork. It gives a daily assessment of the probability that any given CVE will be exploited within the next month.

TAKEAWAY: What if remediation SLAs were based on time-to-exploitation instead of vulnerability severity?

THE URGENCY OF EXPLOITATION ACTIVITY

A vulnerability being published is usually accompanied by a range of other possible events (patches, disclosures, scanner and detection signatures, etc.), but how soon are we observing exploitation activity? Roughly about 1 in every 9 CVEs with observable exploitation activity are observed before the end of the first week after publication.



7 out of every 10 CVEs with exploitation activity are seen in the first year while 3 out of every 10 are first seen a year or longer after publication.

The tipping point is around four and a half months: half of the CVEs with observed exploitation activity are seen before that point and half are first seen after that point.

Only 8% of CVEs have exploitation activity reported before they are publisehd, which is probably more an artifact of data collection than threat activity.

The countdown to initial exploitation is often pretty quick.

That said, there are hundreds of CVEs that went several years before being attacked.

HOW "OLD" IS CURRENT EXPLOITATION ACTIVITY?

Beyond time to/since exploitation, this question gets at another important temporal aspect of exploitation. We analyzed all observed exploitation and recorded the age of all CVEs when they were targeted with exploitation activity. The point is to understand whether attackers are targeting older or newer vulnerabilities on the whole. The chart below will aid that understanding.

What do you do with this information? Well, for starters, we suggest that VM teams maintain their long-term memory. The data clearly shows that the hackings will continue until security improves. Don't be fooled into thinking that attackers only look for the cool new exploit. They are still probing for decade-old vulnerabilities and are happy to exploit them if found.

Here are a few statistical highlights to help you interpret what the chart conveys:

About 1% of observed exploitations target unpublished CVEs (first bar).

About 5% of exploitation activity targets CVEs less than a year old (second bar).

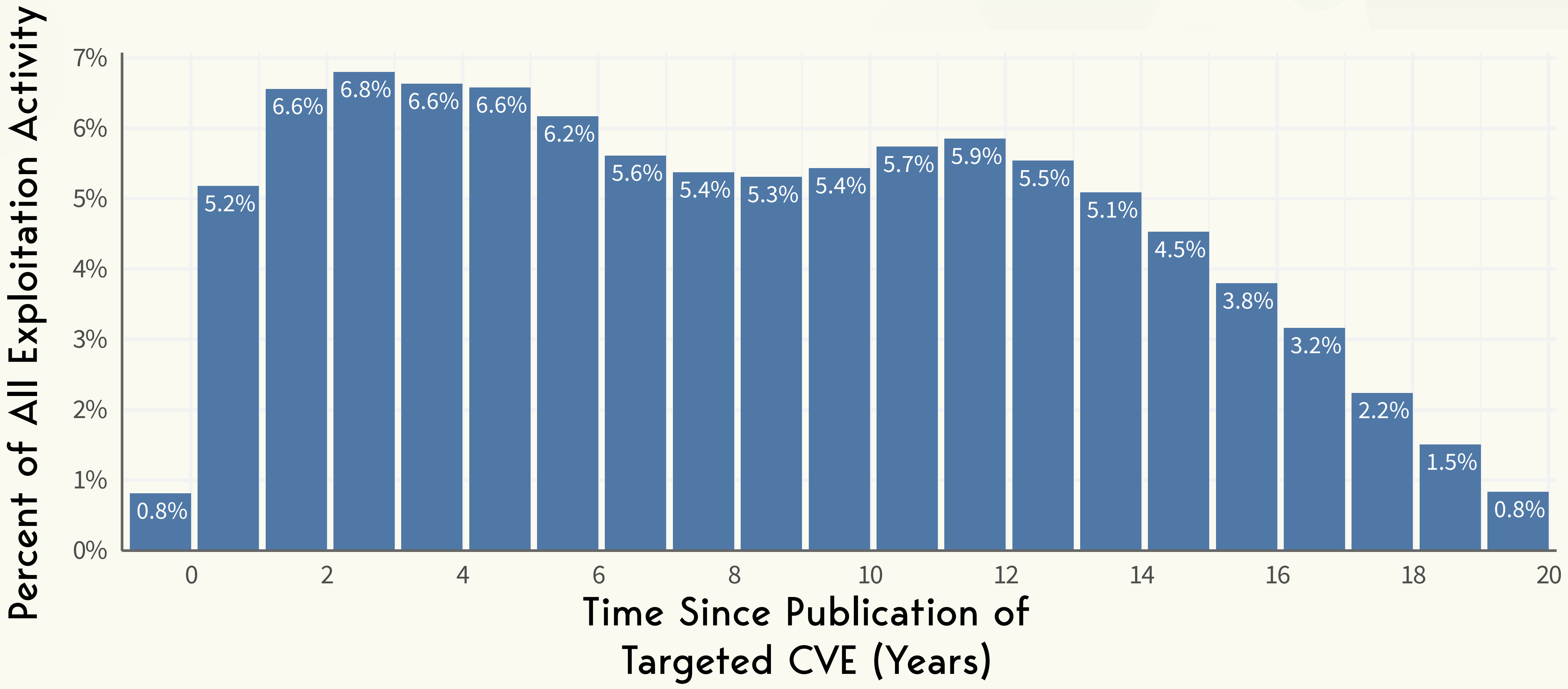
About 6% of current exploits target CVEs released 12 years ago.

About 39% of exploit observations target CVEs that are five or fewer years old.

TAKEAWAY: Attackers are content to keep exploiting the "oldies but goodies" as long as we let them.

THE TIMELINESS OF EXPLOITATION ACTIVITY

The typical CVE with exploitation activity is observed a median of 284 days. This chart breaks down over 8.6 million unique observations of daily exploitation activity and the difference between the publication of the target vulnerability and the date exploitation activity was observed.



< 1 year

Only **6%** of the **8.6 million** daily observed exploitation attempts targeted vulnerabilities before they were a year old

> 10 years

38% of the exploitation attempts targeted vulnerabilities more than 10 years after they were published

3X more exploits target CVEs 10+ years old than those published in last 2 years.

The rate of exploitation for unpublished CVEs equals those published 20 years ago.

HOW WIDESPREAD IS EXPLOITATION AMONG ORGANIZATIONS?

This one was an eye-opener for us. Rather than exploited CVEs or timelines, let's examine the prevalence of exploitation observed across a large population of 100,000+ organizations around the world. Before looking at the figure below, ask yourself this question: what percent of organizations typically see exploitation targeting a particular vulnerability? Perhaps 1% of them? Or 10%? Half?

It turns out that widespread exploitation in the wild is a pretty rare feat. The chart (you can look now) records this reality. Half of all known exploited CVEs are never observed by more than 0.02% of organizations!

Less than 5% of exploited vulns hit more than 1 in 10 organizations. The scope of exploitation becomes important when trying to discern whether your organization is in the crosshairs.

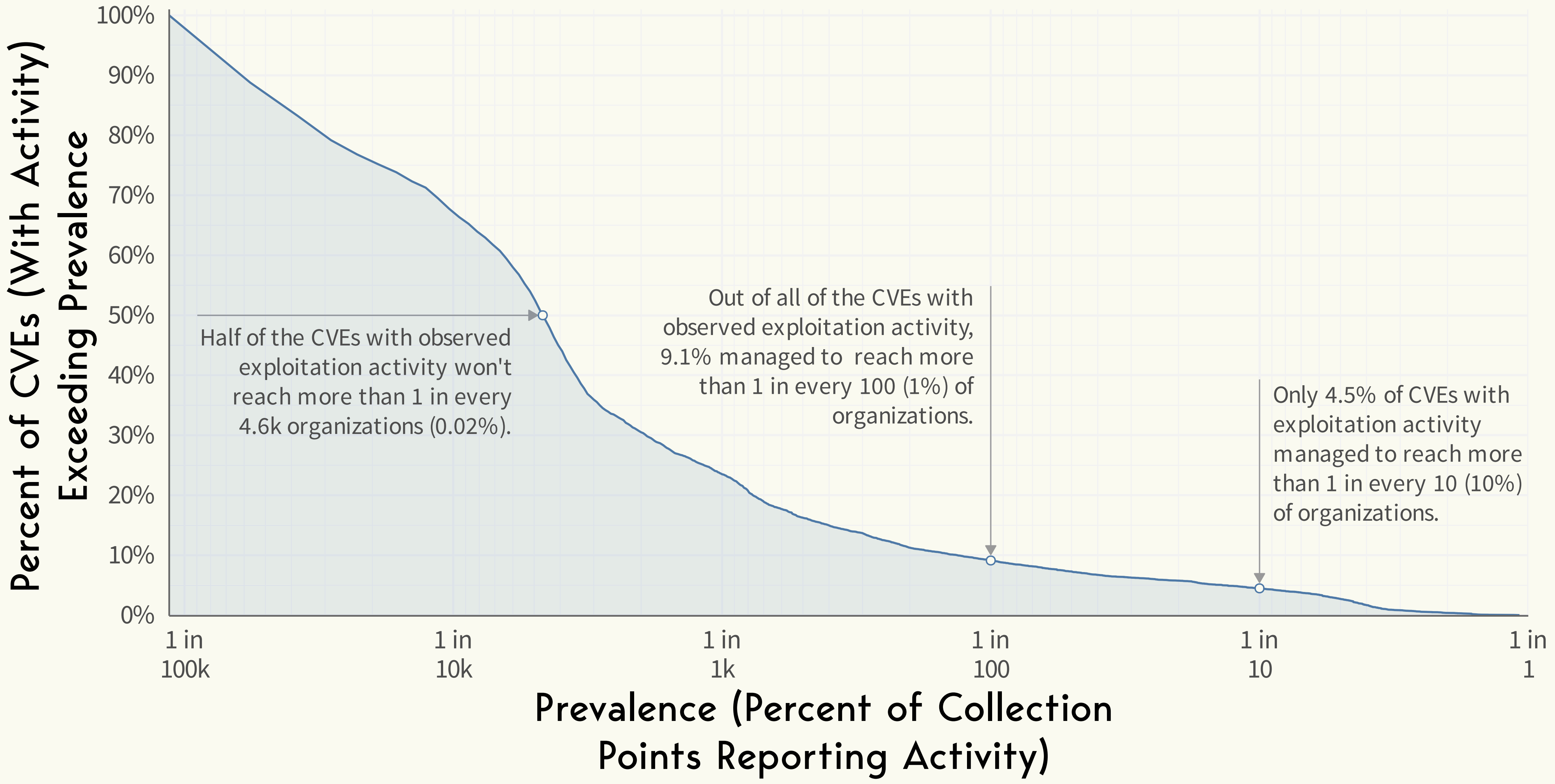
There is another challenge here to conventional thinking. When vulnerabilities are reported as being exploited in the wild, they are generally portrayed as being exploited everywhere. This is clearly not the case. When someone cries, "This is being exploited!", we should request more information about the nature and scope of that exploitation rather than treating all such reports equally.

TAKEAWAY: A small minority of CVEs "go big," achieving widespread exploitation across organizations.

Q: How many organizations see exploit activity from a typical vulnerability?

THE PREVALENCE OF EXPLOITATION ACTIVITY

By identifything which data collection point reports the exploitaiton activity we can get a sense of how far the activity spreads around the world. In the case of published vulnerabilities, it's relatively rare to have widespread exploitation: only 5% of exploited CVEs reach more than 10% of collection points.



Half of the CVEs with observed exploitation activity won't reach more than 1 in every 4.6k organizations (0.02%).

Out of all of the CVEs with observed exploitation activity, 9.1% managed to reach more than 1 in every 100 (1%) of organizations.

Only 4.5% of CVEs with exploitation activity managed to reach more than 1 in every 10 (10%) of organizations.

A: Not many. Exploits hitting more than 1 in 10 organizations are pretty rare.

EVALUATING EPSS PERFORMANCE

This section leverages the clarity of hindsight to measure the reliability of EPSS' predictions of exploitation. We start by describing the methodology used to measure performance and then put several vulnerability rating and prioritization approaches to the test before turning to EPSS.

IN THIS SECTION

How do we evaluate exploit predictions?

How does CVSS perform?

How does the KEV perform?

Can meta-data help predict exploitation?

Can exploit tools help predict exploitation?

How does EPSS perform?

How do EPSS and CVSS Compare?

What EPSS score warrants priority remediation?

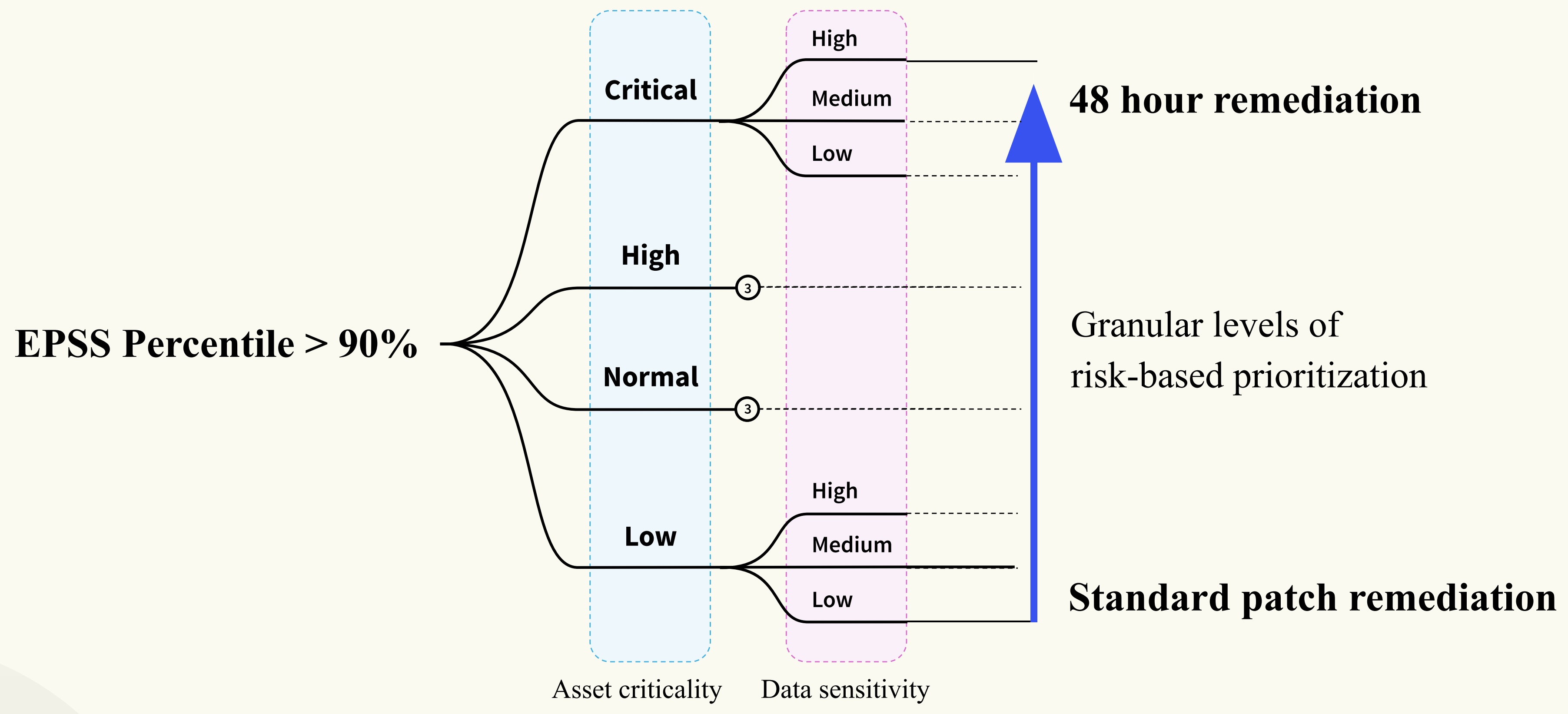
> "**Threat centric** scoring systems like EPSS are the foundation of data-driven vulnerability management programs. Organizations must leverage the insights and context these scores provide, but they can't stop there. They must also determine their organization's unique risk tolerance and contextualise assets based on the business impact caused by a critical vulnerability being exploited on those systems, to prioritise remediation and mobilise response."
>
> - Gavin Millard | VP, Product Management, Tenable

REMARKS FROM NUCLEUS SECURITY

EPSS Thresholds Operationalized with Business Context

As risk–based vulnerability management programs mature, they shift their focus from 'What is being exploited now?' to 'What is likely to be exploited next?'. EPSS uniquely addresses the latter question. It provides an estimate of the likelihood that a software vulnerability will be exploited in the wild based on probability and machine learning.

Setting an EPSS threshold based on the organization's risk tolerance is the first step to operationalizing EPSS. However, this only provides a global prediction. Without organizational context, the effectiveness of using EPSS as a measure of prediction is limited. To manage risk-based prioritization at enterprise scale, Nucleus combines your EPSS threshold with extensive asset and business context including internet accessibility, data sensitivity, asset criticality, and compliance scopes. This unified approach enables teams to effectively operationalize EPSS scores and shift from reactive to proactive prioritization.



As the leader in unified vulnerability management, Nucleus enables enterprises to prioritize and mitigate vulnerabilities faster, at scale. Powered by the Nucleus Data Core, the platform automatically unifies, organizes, and operationalizes finding, threat, and business, data from all your tools.

Nucleus

HOW DO WE EVALUATE EXPLOIT PREDICTIONS?

At the outset of evaluating the performance of EPSS, it makes sense to discuss what that entails and how we measure it. A perfect prediction model will correctly identify all vulnerabilities that are exploited (true positives) with zero omissions (false negatives) or false positives. No prioritization method is perfect, of course, which can be seen in the diagram below that depicts the accuracy of using CVSS scores above 9 to predict exploitation. This sets up the classic performance metrics of precision and recall that are widely used to evaluate classification and prediction models. In the context of VM, we term these efficiency (precision) and coverage (recall) to make the concepts more memorable and practical.

Coverage (recall)

Measures the completeness of prioritizing the exploitation activity. What percentage of all known exploited vulnerabilities were correctly prioritized? If 100 vulnerabilities get exploited but only 40 of those were prioritized, the coverage is 40%. Technically, coverage is the true positives divided by the sum of the true positives and false negatives.

Efficiency (precision)

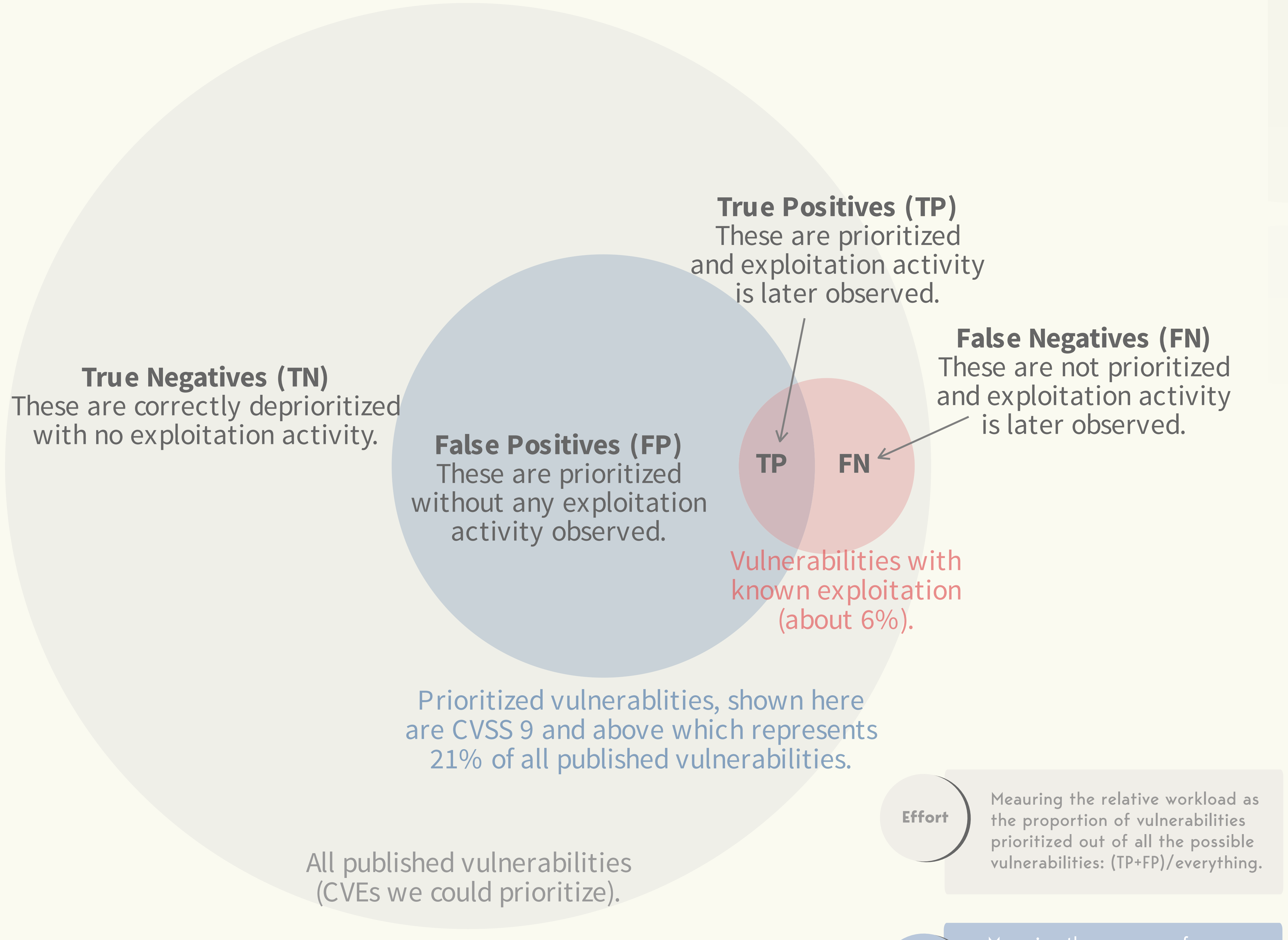
Measures the accuracy of prioritizations. What percentage of vulnerabilities prioritized (for remediation) were actually exploited? If 100 vulnerabilities were predicted to be exploited but only 60 had observed exploitation activity, the efficiency is 60%. Technically, efficiency is the true positives divided by the sum of the true positives and false positives.

Effort

Measures the overall workload created by the prioritization strategy and is simply the percentage of prioritized vulnerabilities out of all vulnerabilities. Typically, we can improve our coverage by increasing our effort, but this comes at the expense of our efficiency. We can only increase all three metrics at the same time by having a better prioritization strategy.

MEASURING PERFORMANCE OF PRIORITIZATION

No matter what strategy is used, there is a tradeoff between true and false positives and true and false negatives. We highlight what each of those mean for vulnerabilities by measuring the performance of a strategy to prioritize CVSS "critical" (9 and above) vulnerabilities.



True Positives (TP)
These are prioritized and exploitation activity is later observed.

True Negatives (TN)
These are correctly deprioritized with no exploitation activity.

False Negatives (FN)
These are not prioritized and exploitation activity is later observed.

False Positives (FP)
These are prioritized without any exploitation activity observed.

Vulnerabilities with known exploitation (about 6%).

Prioritized vulnerablities, shown here are CVSS 9 and above which represents 21% of all published vulnerabilities.

All published vulnerabilities (CVEs we could prioritize).

Effort Meauring the relative workload as the proportion of vulnerabilities prioritized out of all the possible vulnerabilities: (TP+FP)/everything.

Efficiency Meauring the accuracy of our strategy as the vulnerabilities with exploitation activity out of the all prioritized vulnerabilities: TP/(TP+FP).

Coverage Meauring the completeness of our strategy as the vulnerabilities we prioritized out of the all vulnerabilities with exploitation activity: TP/(TP+FN).

The perfect model would identify all exploited vulns (TPs) with zero omissions (FNs) or false positives (FPs).

No perfect approach exists, but we can objectively compare their performance with well-established metrics.

HOW DOES CVSS PERFORM?

Since CVSS was used as the example of how to measure performance in the prior topic, we might as well see that through to actually measure its performance. CVSS has long been a de facto input for many organizations in determining which vulnerabilities should be prioritized for remediation. So, it makes sense to establish a predictive performance baseline with CVSS.

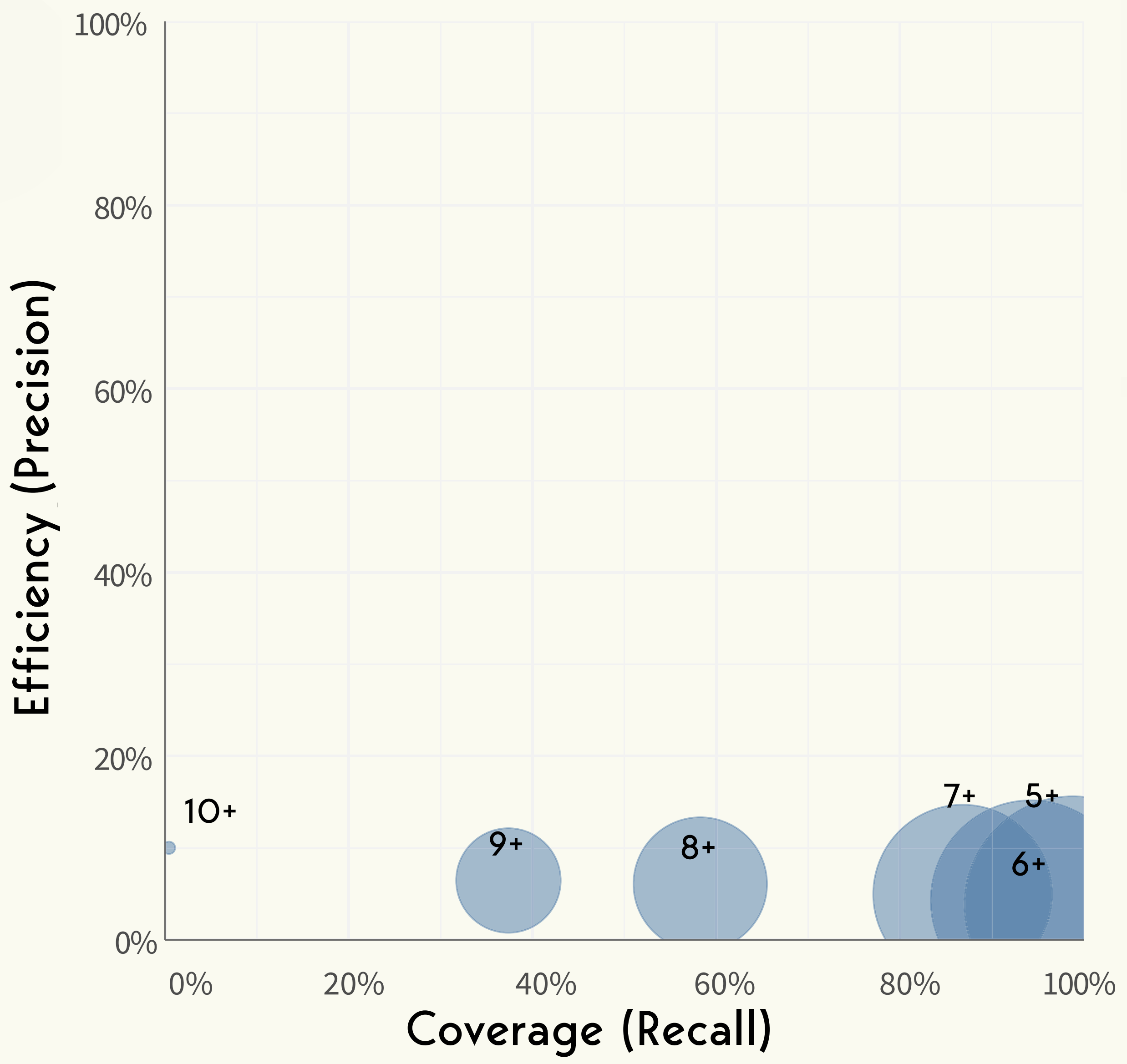
The left chart below plots the coverage (x-axis), efficiency (y-axis), and effort (dot size) achieved by using various CVSS score thresholds to predict exploitation. It's not a great look. A strategy of remediating vulnerabilities with a score of 7 or above—a common recommendation in security and compliance standards—would address the majority (63%) of known exploited CVEs. However, the efficiency is quite low at 10%, indicating quite a bit of misplaced effort spent prioritizing vulnerabilities that did not have any observed exploitation activity.

It's only fair to mention here that CVSS wasn't made to predict exploitation. That said, people often use it that way, and there's a general belief that vulnerabilities with higher scores are more likely to be attacked and should therefore be remediated ASAP. Thus, measuring its performance for this purpose is fair game. Overall, CVSS achieves coverage by increasing effort with a rather low and consistent efficiency.

TAKEAWAY: CVSS is a very inefficient predictor of exploitation; it just wasn't designed for that purpose.

THE PERFORMANCE OF CVSS

Even though CVSS was not designed specifically for exploitation prediction, most people will think of CVSS as having some predictive power for exploitation activity. However, there is very little correlation between a higher CVSS score and observed exploitation activity.



CVSS Threshold	Effort	Coverage	Efficiency
10+	0.1%	0.4%	10.0%
9.8+	19.4%	36.1%	6.7%
9+	20.9%	37.4%	6.5%
8.8+	30.9%	55.3%	6.5%
8+	34.6%	58.3%	6.1%
7.8+	45.0%	67.4%	5.4%
7.5+	59.5%	84.9%	5.1%
7+	63.1%	86.8%	5.0%
6.5+	70.3%	90.4%	4.6%
6+	79.4%	94.3%	4.3%
5.5+	86.5%	96.2%	4.0%
5+	93.4%	98.8%	3.8%

Many assume high CVSS scores indicate a high likelihood of exploitation.

There's little correlation: just ~37% of vulnerabilities with a CVSS score of 9+ have known exploits.

HOW DOES THE KEV PERFORM?

Another popular resource for prioritizing remediation is the Known Exploited Vulnerabilities Catalog (KEV) maintained by the Cybersecurity and Infrastructure Security Agency (CISA). Although created to guide U.S. government agencies, CISA recommends that all organizations monitor the KEV to reduce the likelihood of compromise by known threat actors (and we do too, for what it's worth). We'll briefly review how the KEV performs relative to that goal.

The Venn diagram makes it clear that EPSS data sources contain evidence of exploitation for many vulnerabilities that are not on the KEV. That's not a knock; the KEV is relatively new and has a particular focus. It is also apparent that about a third of CVEs in the KEV are NOT among those observed by EPSS data sources. That alone makes the KEV useful for VM teams to help prioritize remediation.

But the KEV's real strength is its performance on the efficiency scale. It's a great (and FREE!) resource for vulnerability remediation that, unlike CVSS, will minimize wasted effort. It shouldn't be the totality of your prioritization strategy, but it's a strong indicator for VM teams to build on.

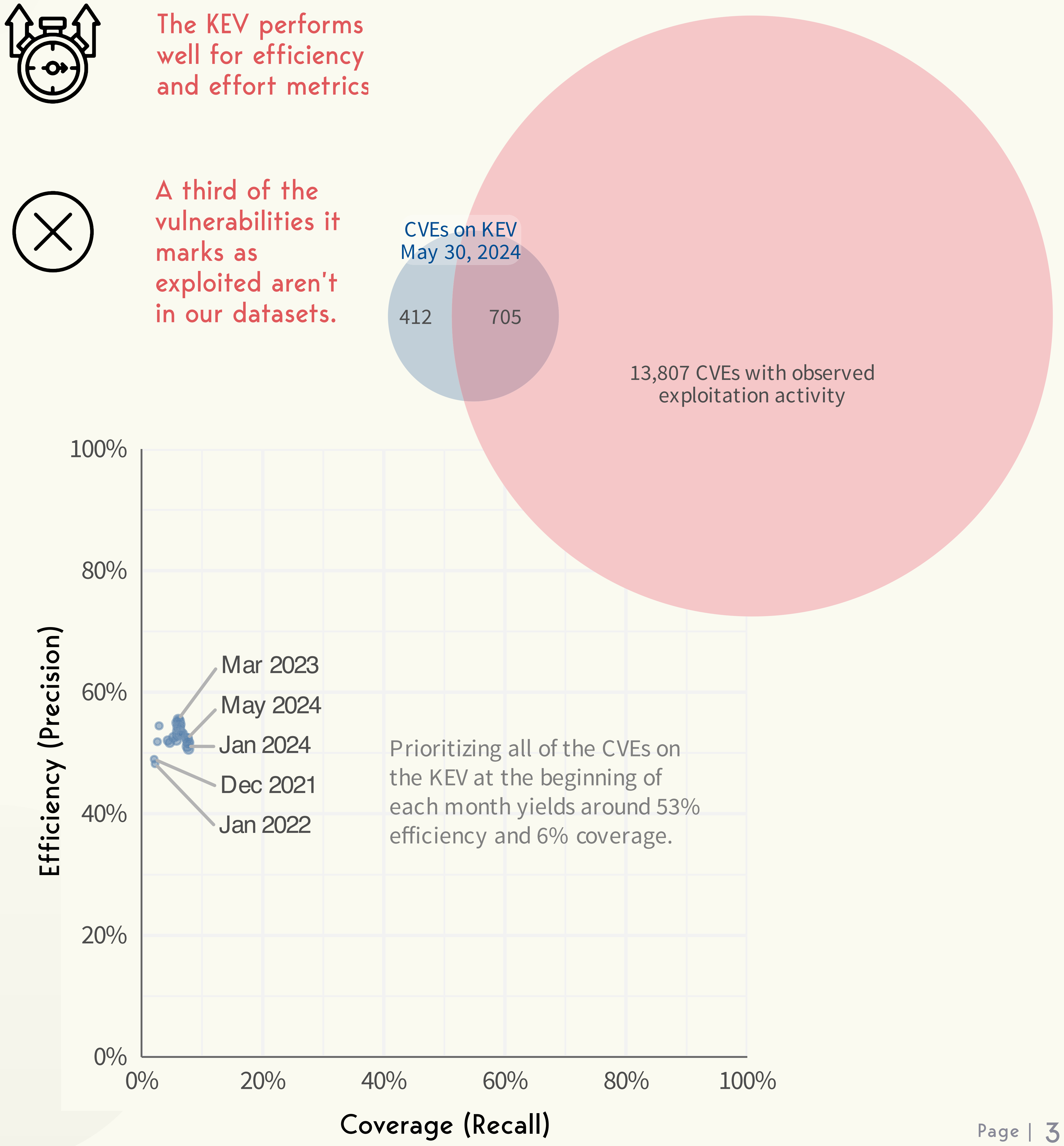
TAKEAWAY: The KEV is a good starting point for prioritizing remediation with little wasted effort.

THE PERFORMANCE OF CISA'S KNOWN EXPLOITED VULNERABILITY (KEV) LIST

As with many sources leveraging expertise or threat intel, the KEV list is quite efficient. Out of the 1,117 CVEs on the KEV, we have observed exploitation activity on 705 (63%) at some point, but that rather high efficiency drops off by 10% (to 53% on average) as we measure month-to-month.

The KEV performs well for efficiency and effort metrics

A third of the vulnerabilities it marks as exploited aren't in our datasets.



Prioritizing all of the CVEs on the KEV at the beginning of each month yields around 53% efficiency and 6% coverage.

CAN METADATA HELP PREDICT EXPLOITATION?

One thing security researchers do when assessing vulnerabilities is parse the descriptive details included with the published CVE. This gives rise to inferences like “This enables remote code execution; it’s gonna be bad.”

Can such inferences form the basis of reliable predictions? This series of charts plot the performance of CVSS metrics, Common Weakness Enumeration (CWE) types, various attributes derived from the description, and the associated vendor(s).

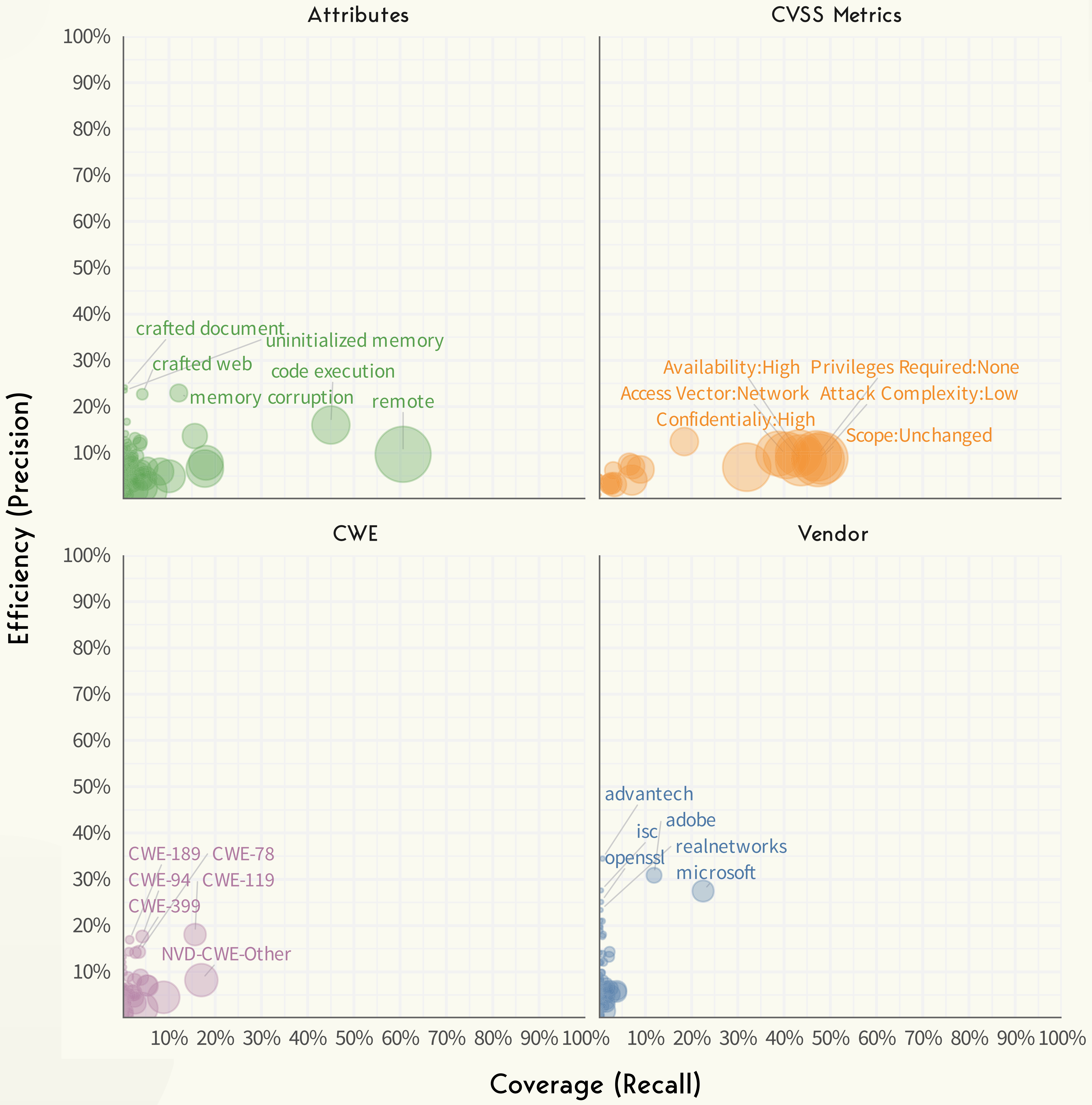
It is indeed true that a large proportion of exploited vulns enable remote code execution (high coverage). But so do many more that haven’t been exploited (leading to low efficiency and high effort). There are some decent indicators here, but on the whole, these don’t perform very well as individual predictors of exploitation.

EPSS includes all of these info sources (and more) as inputs for its predictions.

TAKEAWAY: None of these attributes are reliable individual predictors; they’re best modeled collectively.

THE PERFORMANCE OF INDIVIDUAL VULNERABILITY ATTRIBUTES

Nobody would base their prioritization on a single variable, but it’s informative to look at their perfomance. It can align our expectations and build our intuition about how different vulnerability features may help predict exploitation.



CAN EXPLOIT TOOLS HELP PREDICT EXPLOITATION?

This next chart brings together the sources of metadata from the previous page and adds some popular exploit tools and databases to the mix (in red). While Metasploit, Sn1per, ExploitDB and their ilk aren't intended to score severity like CVSS or to predict exploitation like EPSS, they do offer a window into which vulnerabilities have seemed interesting enough to be "weaponized" to some degree. Given that, it makes sense that the vulnerabilities included in them would correlate with those exploited in the wild. The results shown here bear that out.

There's an important pattern here. Note that everything listed tends to perform better for coverage OR efficiency. None does both very well. Perhaps a model that factors all of this into making exploit predictions can do better? We'll find out on the next page.

The exploit tools and databases shown here are limited in scope (low coverage).

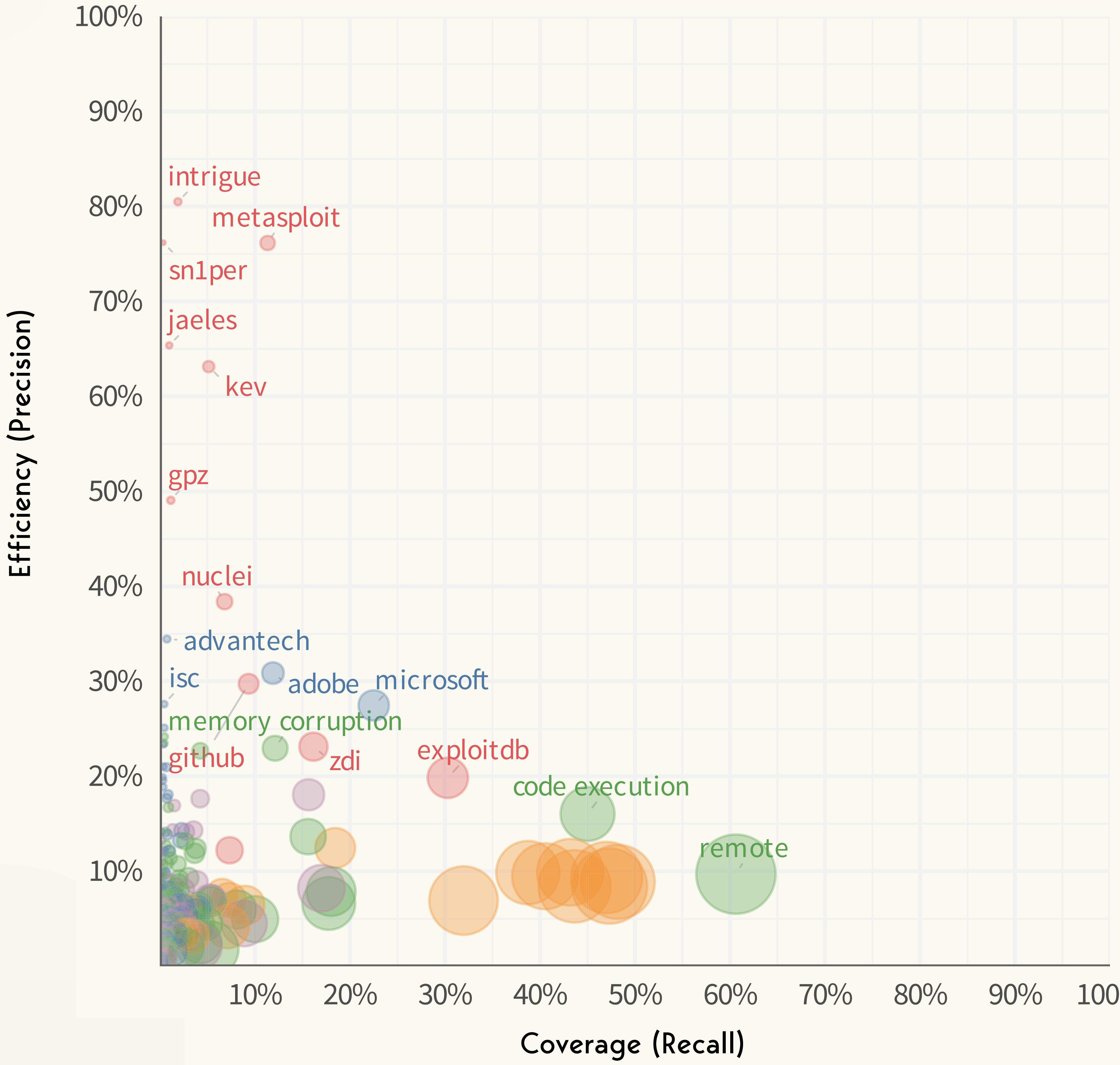
But a high proportion of the vulnerabilities they contain have known exploitation in the wild (high efficiency).

They include vulnerabilitiess likely to be on attackers' radar and thus worth prioritizing for remediation.

TAKEAWAY: Exploit tools and databases generally offer high signal-to-noise ratio (high efficiency) but with limited individual coverage.

THE FULL VIEW OF THE PERFORMANCE OF INDIVIDUAL VULNERABILITY ATTRIBUTES

Curated lists based on expertise (Metasploit and off sec scanners) increase in efficiency at the expense of coverage with less effort (smaller circles). Meanwhile, static attributes of vulnerabilities can drive a lot of effort to achieve coverage but at a much lower efficiency.



HOW DOES EPSS PERFORM?

At 20+ pages into a study that promises to evaluate the performance of EPSS, we are now ready to make good on that promise. Recall that the perfect predictive model will max out on the coverage and efficiency axes in the upper right. Nothing we've shown thus far comes close, but EPSS has moved closer to that coveted upper-right corner with each successive version.

This plot likely prompts the question "Why lines vs. dots?" That stems from EPSS producing scores ranging from 0 to 1, with each achieving different coverage and efficiency levels. Each line plots the daily results for each version's lifespan. The number bubbles indicate the performance of thresholds in that range. We discuss how to choose the ideal EPSS threshold for your team later.

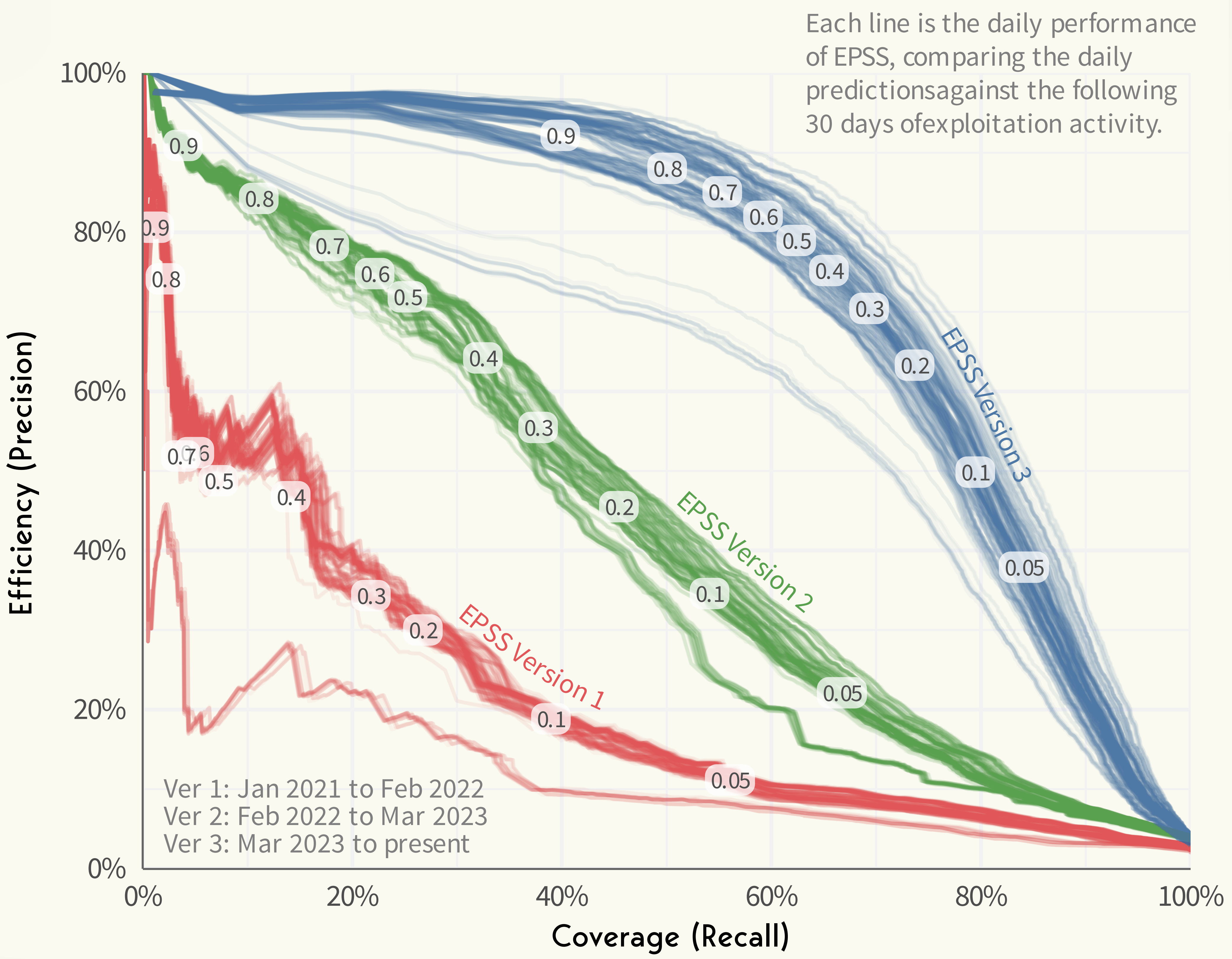
Remediating vulnerabilities with an EPSS score of 0.6+ achieves a coverage of ~60% with 80% efficiency.

At 0.1+, that changes to 80% coverage and 50% efficiency.

TAKEAWAY: Versions of EPSS show increasingly strong performance across the range of scores.

THE PERFORMANCE OF THE EXPLOIT PREDICTION SCORING SYSTEM (EPSS)

The output of EPSS is a probability (0%–100%) of exploitation activity being oberved in the next 30 days. Because it's a continuous value, the "point" slides across the plot, creating a line from high efficiency to high coverage.



Each line is the daily performance of EPSS, comparing the daily predictionsagainst the following 30 days ofexploitation activity.

Ver 1: Jan 2021 to Feb 2022
Ver 2: Feb 2022 to Mar 2023
Ver 3: Mar 2023 to present

HOW DO EPSS AND CVSS COMPARE?

Since we've now measured the predictive performance of EPSS and CVSS, we suspect readers may have this question. There are many ways to go about answering it, but we think the "apples to apples" comparison shown here is the most fair and useful.

We feel compelled to assert that we're not trying to pick on CVSS here. But it's important to understand which of these scoring systems is better suited to prioritizing remediation based on the probability of exploitation. EPSS clearly wins in that regard.

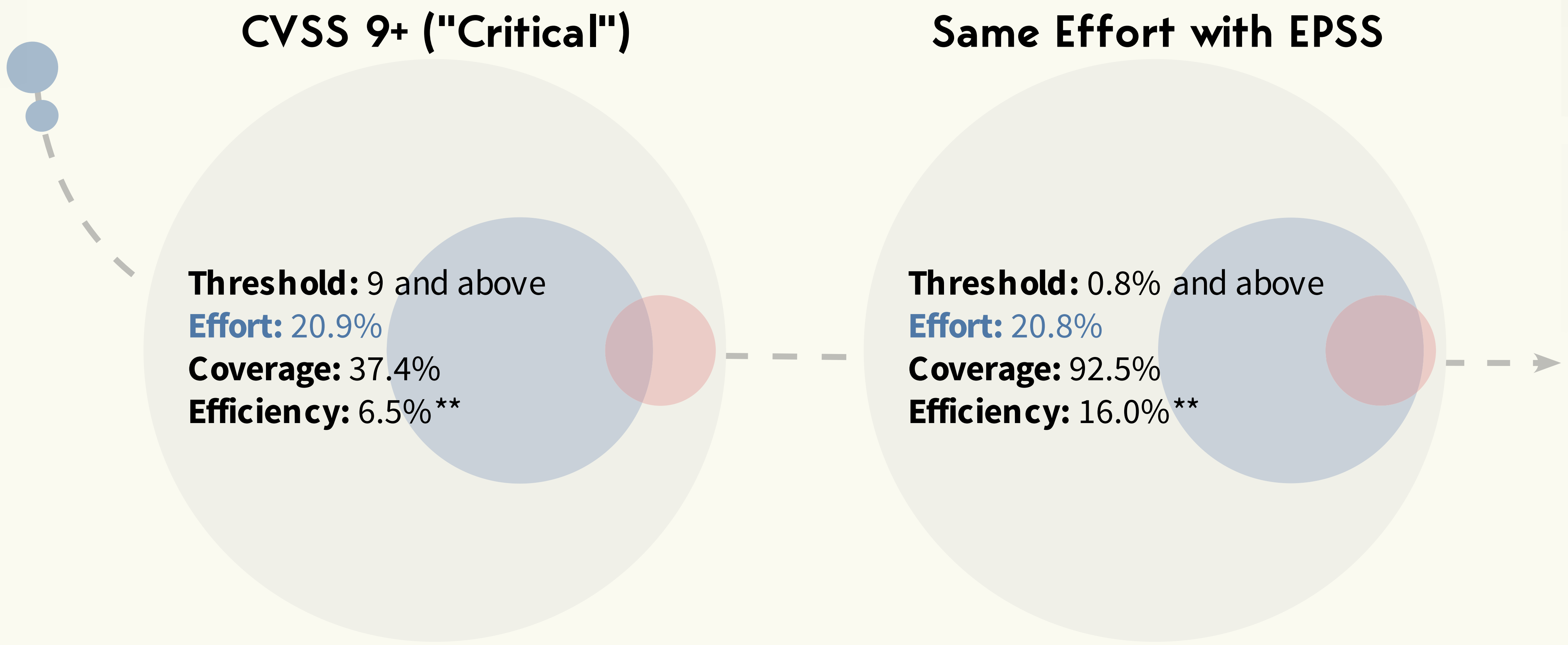
If we compare based on equivalent level of effort (remediating about 21% of vulnerabilities), EPSS achieves almost 3x more coverage (93% vs. 37%) and over twice the efficiency (16% vs. 7%) of CVSS.

Achieving equivalent coverage (87%) requires 6x more effort (63% vs. 10%) and is 6 times less efficient (5% vs. 30%) with CVSS than with EPSS.

TAKEAWAY: EPSS performs demonstrably better than CVSS in equivalent metric-based comparisons.

PERFORMANCE FROM CVSS TO EPSS

It is difficult to map direction from CVSS scores to EPSS scores. But if we hold one of the performance measures the same (such as effort or coverage) we can look at the changes in the other metrics as shown here.

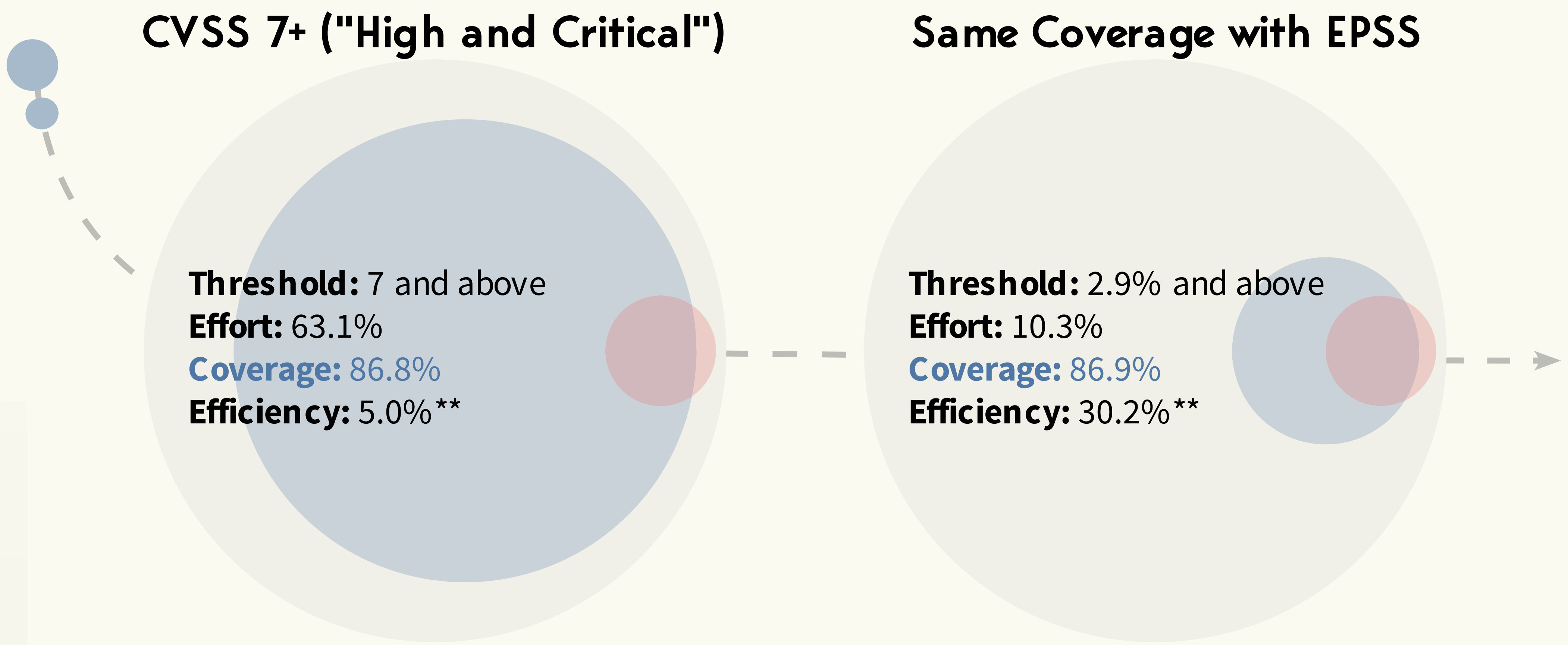


CVSS 9+ ("Critical")

Threshold: 9 and above
Effort: 20.9%
Coverage: 37.4%
Efficiency: 6.5%**

Same Effort with EPSS

Threshold: 0.8% and above
Effort: 20.8%
Coverage: 92.5%
Efficiency: 16.0%**



CVSS 7+ ("High and Critical")

Threshold: 7 and above
Effort: 63.1%
Coverage: 86.8%
Efficiency: 5.0%**

Same Coverage with EPSS

Threshold: 2.9% and above
Effort: 10.3%
Coverage: 86.9%
Efficiency: 30.2%**

WHAT EPSS SCORE WARRANTS PRIORITY REMEDIATION?

We showed previously that EPSS produces a range of scores that achieve different levels of coverage and efficiency. Because of that, many organizations attempting to use EPSS wonder what score(s) should trigger priority remediation. EPSS doesn't come with that guidance because the answer is ultimately dependent upon your organization's risk tolerance and capabilities. The figure below should offer some insight that helps guide these decisions.

Maximizing coverage comes with the cost of lower efficiency and higher effort. Risk-averse firms may be willing to make that trade. Resource-strained or less mature organizations may wish to maximize efficiency first and work to broaden coverage over time.

Using EPSS to prioritize remediation is a balancing act of competing priorities.

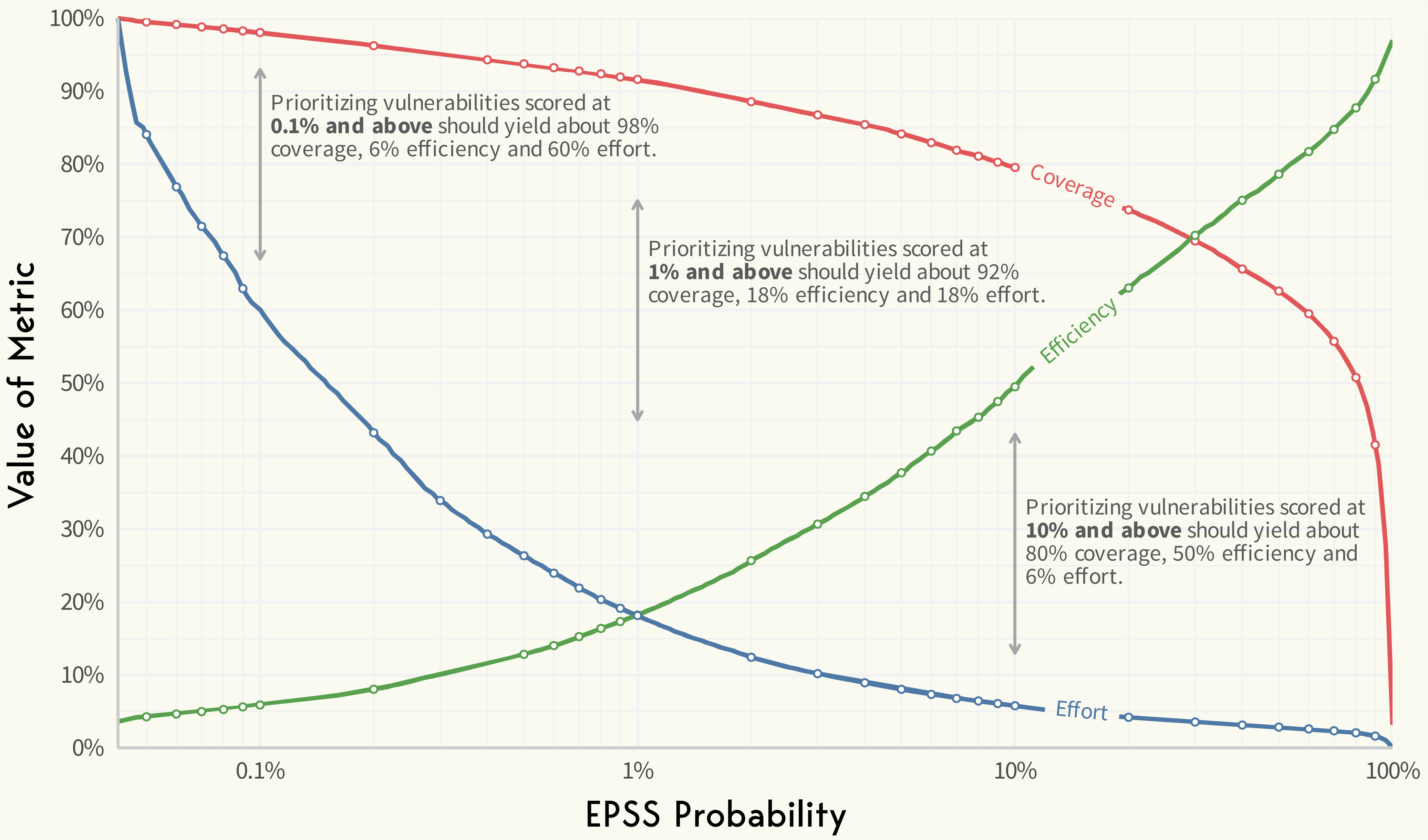
There's no "easy button" to achieve high coverage.

Performance metrics can **help** dial in and maintain a balance that works for your firm.

TAKEAWAY: EPSS supports a remediation strategy tailored to your risk tolerance and capabilities.

PICKING THRESHOLDS FOR EPSS

Select a threshold for EPSS along the horizontal axis and trace it upwards to each metric to determine the coverage, efficiency, and level of effort. These represents the performance of EPSS from March 7, 2023 to to May 1, 2024.



Prioritizing vulnerabilities scored at **0.1% and above** should yield about 98% coverage, 6% efficiency and 60% effort.

Prioritizing vulnerabilities scored at **1% and above** should yield about 92% coverage, 18% efficiency and 18% effort.

Prioritizing vulnerabilities scored at **10% and above** should yield about 80% coverage, 50% efficiency and 6% effort.

Coverage The percent of vulnerabilities with observed exploitation activity in the following 30 days that had been prioritized.

Effort The percent of vulnerabilities being prioritized

Efficiency The percent of prioritized vulnerabilities with observed exploitation activity in the following 30 days.

CONCLUSION

What's next for EPSS

Thanks for sticking with us this far. If you are still hungry for more, please visit First.org and consider joining the EPSS Special Interest Group (SIG). The SIG discusses all things EPSS and is working on the adoption of EPSS and discussing ways that EPSS can and should be used in modern vulnerability management practices. If you still want more, much of the details behind EPSS are covered in the handful of publications we have published about EPSS (see the website).

The Exploit Prediction Scoring System is and always will be data-driven. Because of that we are continually working to expand the coverage of our data. Additionally, and with the help of sponsorships we are upgrading our data collection infrastructure this summer and will be releasing the next version of EPSS “real soon now” (watch the website!)

Having gone through all of that, the future for EPSS is simple: more of the same but better. We want to keep EPSS as simple as possible and to keep EPSS exactly what it is, a prediction scoring system that anyone can use. We hope that we can continue to improve and evolve EPSS, so please, join in the discussion, share your thoughts or better yet, share your data!

APPENDIX

EPSS Overview & History

May 2018

Cyentia released the first report in the Prioritization to Prediction series with Kenna Security. This research launched discussions that would lead to EPSS.

June 2019

First EPSS model and performance results presented at the Workshop on the Economics of Information Security (WEIS) conference in Boston, MA.

August 2019

Pre-publication paper “Exploit Prediction Scoring System” was presented at Blackhat, Las Vegas, NV and later published to Digital Threats: Research and Practice in July 2021.

February 2020

EPSS Special Interest Group formed at FIRST.org; first meeting held April 17th, 2020.

September 2020

First EPSS paper published in the Journal of Cybersecurity titled “Improving vulnerability remediation through better exploit prediction.”

January 2021

Cyentia began producing daily EPSS scores published via FIRST.org.

February 2022

EPSS version 2 published based on a more powerful machine learning model and more data sources.

February 2023

“Enhancing Vulnerability Prioritization: Data-Driven Exploit Predictions with Community-Driven Insights” posted to arxiv; presented at WEIS 2023 in July 2023.

March 2023

EPSS version 3 published with further improvements to the core ML model and even more data sources.

A VISUAL EXPLORATION OF

EXPLOITATION IN THE WILD

Cyentia INSTITUTE

FIRST

EPSS

tenable

JupiterOne

Nucleus

FOSSA

ACDS

SOFTRAMS

KYND

VULCAN.

APPSALUTE

edgescan