EMPIRICAL

Empirical Observations:

How to Deal with Speed

EMPIRICAL
OBSERVATIONS

> "It don't matter if you win by an inch or a mile, winning's winning."
>
> — Dominic Toretto, 2001.

A reasonable reaction to the current exposure management landscape is to look at the numbers and say, "Well, that seems bad." CVE volume keeps climbing, and AI is making it easier to find flaws, analyze them, and in some cases move from disclosure to exploit faster than ever. The gap between "this exists" and "someone is trying to use it" keeps compressing.

The second reaction is to look at the same numbers and ask a better question: "How much of this actually changes what we should do first?"

A lot of the anxiety around AI-discovered vulnerabilities assumes that every new disclosure adds an equal amount of work to the queue: more patching, shorter SLAs, and the greater emphasis on a speed contest against a machine that doesn't sleep.

That's not a good game to play. But the good news is, you don't need to.

Using research that we created in our contribution to this year's Verizon DBIR, Empirical produced this paper to discuss the pointlessness of faster patching and the risks of automated remediation. Exposure management doesn't get better with fast patching; it gets rescued by applying the proper prioritization. If the number of disclosures goes up but the number of truly actionable exploit risks stays relatively flat, the problem is not raw volume but rather a broken process for separating true signal from pure noise. Let's be clear: both approaches are needed and useful, but a clear Foundation for making decisions comes first.

In doing so, we'll push back on a few comfortable myths:

- **That vulnerability volume is the same thing as exploitable risk**
- **That "known exploited" is a permanent and equally urgent state.**
- **That severity is a good proxy for attacker behavior.**
- **That a faster world requires nothing more sophisticated than shorter patch SLAs.**
- **And that binary lists can keep up with a threat landscape where timing, probability, and local context are doing most of the actual work.**

The data points in a different direction. The future of exposure management is not panic at scale, but rather calibrated, time-aware prediction applied to prioritization. The disclosure curve can keep rising and exploit development can keep accelerating, but a program that understands odds, recency, and local relevance can still stay ahead.

Here comes the Flood

In April of 2026, FIRST counted 6,420 more CVEs than it had forecast for the same period two months earlier, some 46 percent higher than expected. By June the full-year projection had been revised from 59,000 to roughly 66,000. Then FIRST said the part that matters to enterprises – "actionable exploitability remains flat". That is, if you filter for real exploitability: a CISA KEV entry or an EPSS score above 10 percent then the patching burden has not moved. The denominator exploded while the numerator that actually threatens most organizations sat still.

We'll spend the rest of this document agreeing and diving deeper into richer data. More flaws are not what makes the work harder. The work gets harder because the signal-to-noise ratio is collapsing and because of the acceleration of just about everything, and mostly, because most programs are still triaging on the noise.

This year another datapoint raised eyebrows. Verizon's storied Data Breach Investigations Report, for the first time since its inception, raised vulnerability exploitation to the top of the threat stack.

"Exploitation of vulnerabilities is now the most common initial access vector for breaches. It has risen to 31% in this year's reporting dataset, while credential abuse—the previous leader—is down to 13%."

We at Empirical Security contributed the data and analysis for pages 19 and 20 of the DBIR this year, the section titled "Recency Bias for the Win?" The main conclusion here is that the longer it's been since a vulnerability has been exploited, the less likely it is to be exploited again soon. As those in the know already know, it sometimes takes months of work and dozens of charts to get down to one insight in the DBIR, and we want to release all the work that went into those two pages.

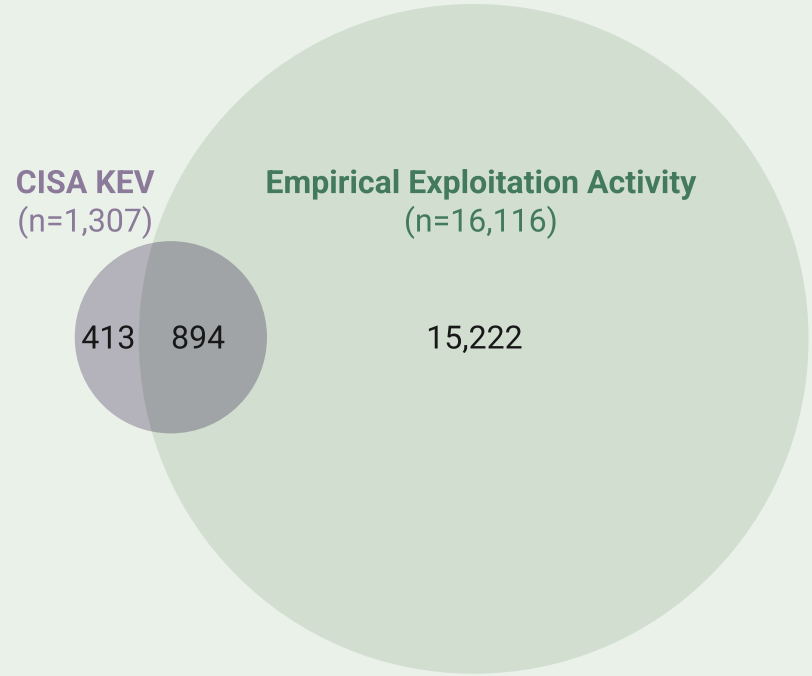
What the data says about which vulnerabilities attackers actually use, and why most programs are sorting the wrong pile

Most vulnerability programs run on a set of binary flags: is it exploitable? Is it on a KEV list? Is it "critical", whatever that may mean. The CISA KEV catalog enshrines this, and it does so honestly: a vulnerability gets added, it stays added and there is no field for when exploitation last occurred. There is no decay. A flaw that a botnet hammered in 2021 and that has been silent for four years carries the same label as one being weaponized this morning for which no patch exists yet.

When compared with actual exploitation telemetry the binary decisions don't survive contact with the data. When pulling data for the DBIR (in 2025), across all observed history (about 6 years in our case) we counted 16,116 CVEs with exploitation activity. The CISA KEV catalog holds 1,307 of them, and our telemetry corroborates 894 of those. That leaves 413 KEV entries we cannot see at all, and 15,222 actively exploited CVEs that never reach the catalog.

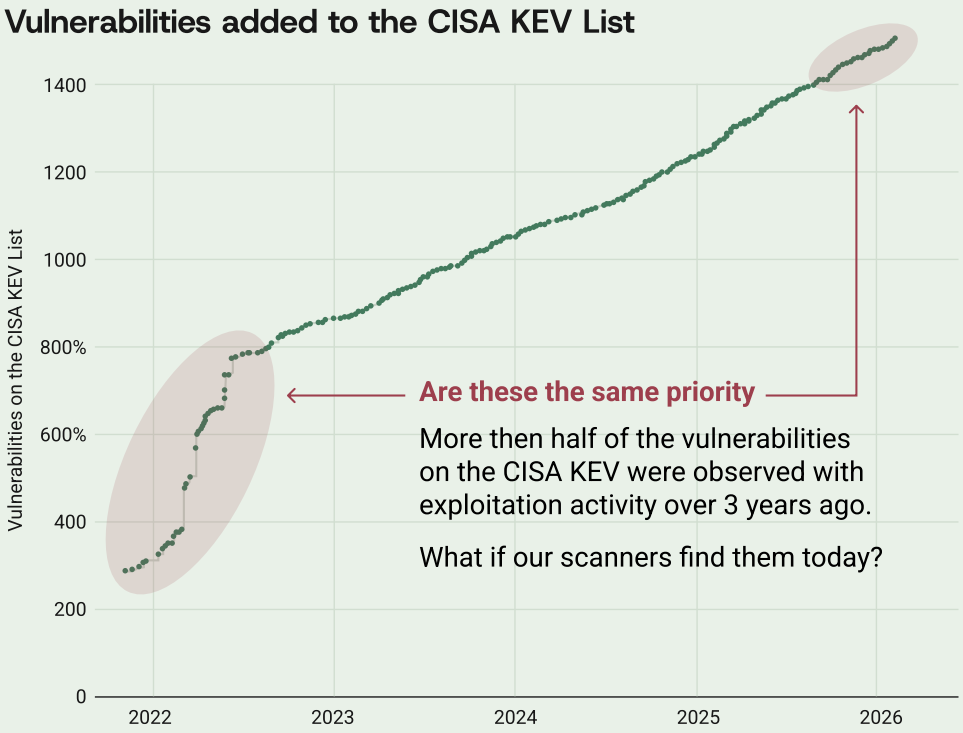
Exploitation Activity by Source (All time)

The lack of overlap indicates there are more vulnerabilities being exploited and not being observed or captured for prioritization.



More than half of the entries on the KEV were last observed with exploitation activity over three years ago, yet a scanner, vulnerability intelligence tool, or exposure management platform finding one of them today fires the same alarm as a flaw under attack this morning.

KEV is small because humans vet each entry by hand, and the people doing it are careful and clever and working with the methods available to them. The telemetry circle is larger because it draws on honeypots, malware analysis, and network tooling that watches automated activity at internet scale, some of which gets blocked before it lands. Each circle catches a slice of attacker behavior the other misses.



Even a national census, counting people who want to be counted, has to estimate the ones it misses. When two honest sensors sample the same population and each catches what the other drops, two conclusions follow and neither is comfortable. You will never know with certainty whether any given vulnerability is being exploited. And there is more exploitation happening than either source observes.

As we often remind ourselves, the datasets available in security are not about the existence or absence of exploitation. The best we can do in the cybersecurity domain is to say a vulnerability is (1) has exploitation activity or (2) we don't know yet. This is a perfect set of constraints for probabilistic reasoning, and so we start asking a different question: What is the probability a vulnerability gets exploited in the next 30 days, given what we have actually observed?

Recency beats history, and history still helps

Conventional wisdom in cybersecurity tells us that if a vulnerability is known to be exploited that everyone should patch it immediately. As a result we get helpful researchers and/or security companies (that tend to sell solutions, but set that aside) announcing that this or that vulnerability has been exploited. They usually do this through a blog or a special publication or maybe a post on social media and some will focus on the vulnerability while others mention it casually in passing.

Other helpful folks attempt to turn these mentions into a collected list of "known exploited vulnerabilities" or a KEV for short. That's what defenders get, a list of vulnerabilities that, at a single point in history, has (likely) been exploited in the wild. The lists aren't often updated and just keep growing… So how helpful are these lists?

1. Exploitation is not an on/off state - Past exploitation does not guarantee future exploitation and just because we saw something exploited in the past does not mean we will see it exploited again. In reality, most exploitation activity is sparse and bursty.

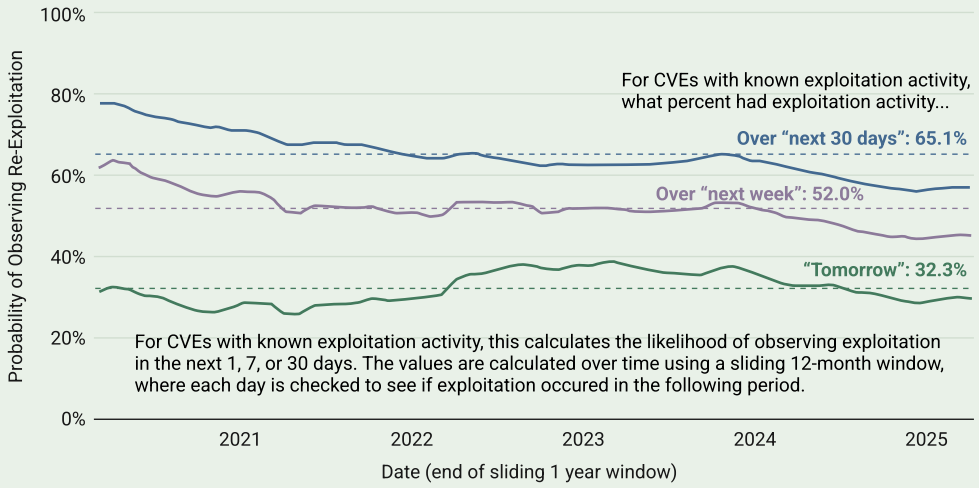
2. Exploitation is not widespread - only about 6% of vulnerabilities with exploitation activity were detected at more than 1% of data collection points. You are more likely to flip a coin and get heads 13 times in a row than observe most vulnerabilities being exploited.

We've been trying to figure out how to describe the (sparse) patterns in exploitation activity more accurately and succinctly. It can be deceiving to just treat exploitation (or even re-exploitation) as a binary and it's hard to compare a vulnerability with wide-spread, long-lasted activity to a relatively newcomer with sparse activity. There are no nice neat distributions here.

Question: What is the probability of observing future exploitation activity, given that there is known exploitation activity in the past?

Short answer: 32% - About one-third of the days recorded exploitation activity after the initial day with exploitation activity, but it fluctuates over time.

Probability of Exploitation Resurgence



We leveraged our dataset containing only CVEs (vulnerabilities) that had already been exploited at least once — no hypothetical risks, just real-world activity. In our data, we are approaching nearly 17,000 CVEs with exploitation activity over the past several years. For each CVE and every day after the first known exploitation, we answered:

- **Was there exploitation the next day?**
- **Did exploitation happen within the next 7 days?**
- **Was there any exploitation activity in the next 30 days?**

We rolled this analysis through the dataset using a 12-month sliding window — meaning, we always focused on just the most recent 12 months of data to calculate these frequencies. This helps capture and visualize how exploitation patterns change over time. So, what did we find?

Here are the numbers:

- **On any given day after a CVE has known exploitation activity, there's a 32.3% chance that exploitation happens again "tomorrow".**
- **There's a 52% chance that exploitation happens at least once in the next 7 days.**
- **And looking a bit further, there's a 65.1% chance that exploitation happens again within the next 30 days.**

The difference between these three forward-looking windows is highlighting the sparseness and burstiness of exploitation activity. If exploitation activity was a binary (once exploited, always exploited) these would be much closer together. But since activity is bursty, it's more likely to see exploitation activity over a 30-day window than a 7-day window.

Why does this matter? It's certainly good advice to prioritize anything with known exploitation activity, but our data shows the reality is a little more nuanced. Once a vulnerability has been exploited, it's likely not just a one-time event, but it's also not an immediate justification to panic. Additionally, just saying it's been exploited in the past is an absolute bare minimum of information.

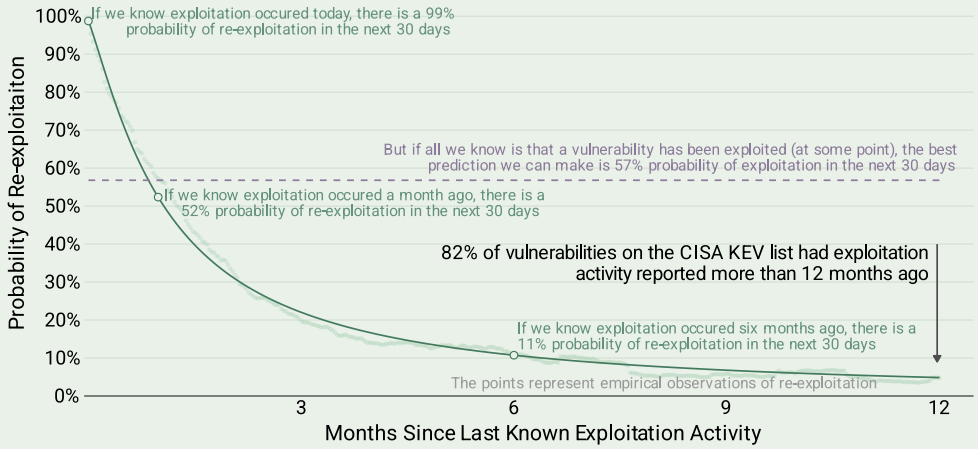
Past exploitation is not a guarantee of future exploitation

Another assumption within cybersecurity is that once a vulnerability is exploited in the wild, a switch has been flipped and that particular vulnerability is exploited everywhere and consistently. This assumption means that all we need to know is that exploitation has occurred at some point. But past exploitation is not a guarantee of future exploitation.

There isn't one pattern for exploitation activity, some vulnerabilities are exploited consistently and relatively widely, but the majority of activity is bursty and sparse and sometimes activity is only observed for a brief period and then stops completely. If we have a systematic and consistent method for observing exploitation activity we can attempt to capture and explain this pattern. By tracking and factoring in recent exploitation activity, we can build a rather strong model for future exploitation activity. We can certainly do better than any approach based on just knowing "exploited at some point".

How Recency of Exploitation Predicts Re-Exploitation

If all we know is that exploitation has occured at some point, the best predictive model of re-exploitation in the next 30 days is a flat 57%. But if we know how recently a vulnerability has exploitation activity we can make a much more informed and accurate prediction.



The above chart is saying quite a lot. It's constructed by tracking daily exploitation activity (through various sources such as IDS/IPS, malware detections and EDR data), and for each day and each vulnerability with activity at some point in the past, we observe how long since the last observed exploitation activity and if any activity was observed over the next 30 days. We ended up with tens of millions of CVE+day combinations and the proportion of observations with exploitation in the "next" 30 days are represented as the faint points behind the line. The dark green line represents a best fit logistic regression model. The dashed purple line represents the naïve prediction where we only know that exploitation has occurred, but we don't know exactly how recently.

To evaluate this approach, we can calculate a Brier score, it's a method to measure the accuracy of probability-based predictions. Brier scores range from 0 to 1, with 0 being absolutely perfect predictions and 1 being completely and consistently wrong predictions. As a point of reference, short-term weather forecasts (1-2 days out) typically have a Brier score in the 0.06 to 0.2 range. Forecasts of US presidential elections typically range from 0.1 to 0.2 and EPSS falls in a nice range of 0.017 to 0.02 (yay team!).

The Brier score of the naïve model (which assumes a flat 65% probability of re-exploitation in the next 30 days regardless of recent activity), is 0.387. Which is not that impressive, Clearly there is a pattern in re-exploitation activity that a simple model misses. It under-estimates recently exploited vulnerabilities and greatly over-estimates the probability of re-exploitation when exploitation activity hasn't been observed in more than 3 weeks. When exploitation activity is tracked systematically and consistently we can factor that into our estimations. The Brier score drops significantly, down to a rather impressive 00004.

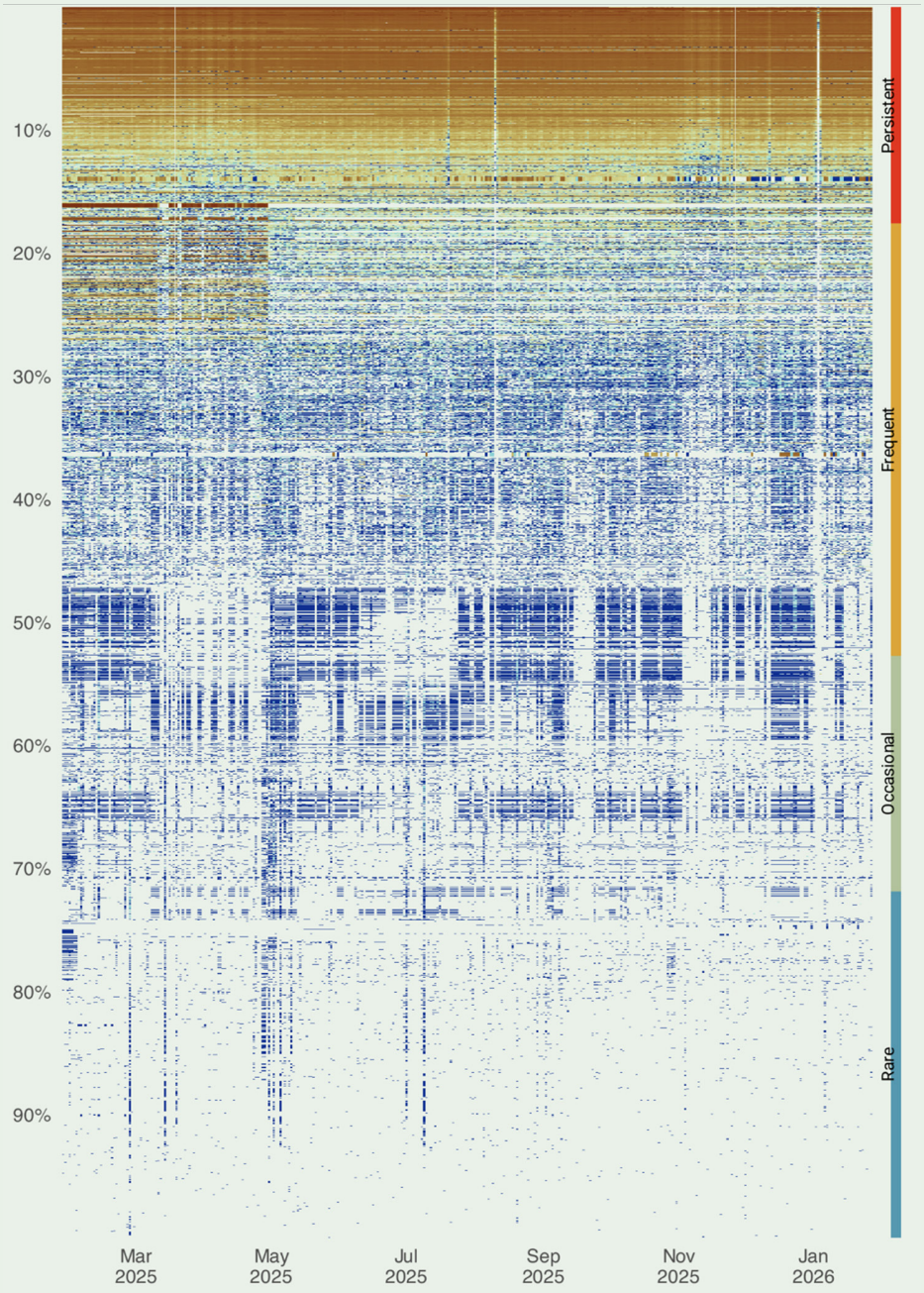
Prioritizing flaws that have had exploitation activity observed is a pretty good strategy in the grand scheme of things. But we must understand that exploitation activity is not a one-way door, and it is not binary. Vulnerability exploitation activity has a temporal quality and knowledge of that greatly improves prioritization. Lists of known exploited vulnerability lists are better than nothing, but they could be greatly improved with systemic detection capability so that recency could be factored into prioritization models.

Exploitation is a mixture, not a monolith

Exploitation refuses to be one thing. Cluster the activity and four behaviors fall out.

It Boils Down to Bursts and Blast Radius

Exploitation cannot be represented with a single timestamp. Real exploitation activity is bursty and unevenly distributed.

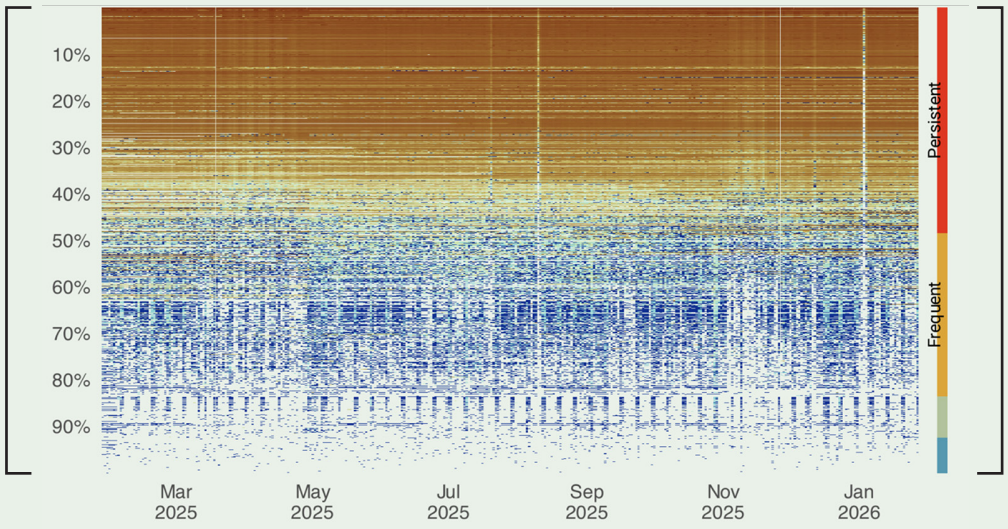


EMPIRICAL

It Boils Down to Bursts and Blast Radius

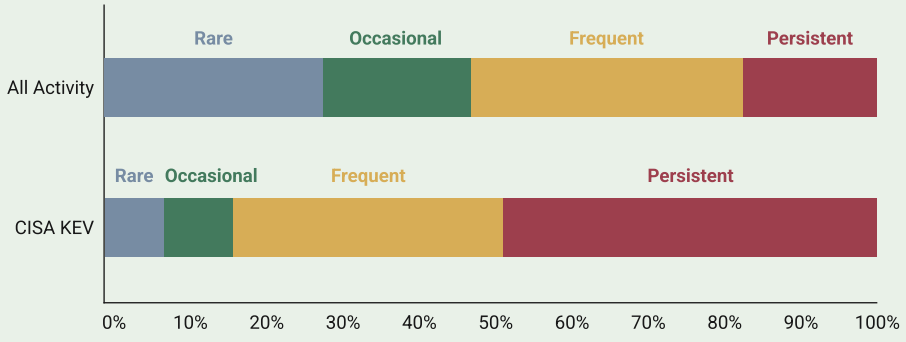
Exploitation cannot be represented with a single timestamp. Real exploitation activity is bursty and unevenly distributed.

Subset to CISA KEV

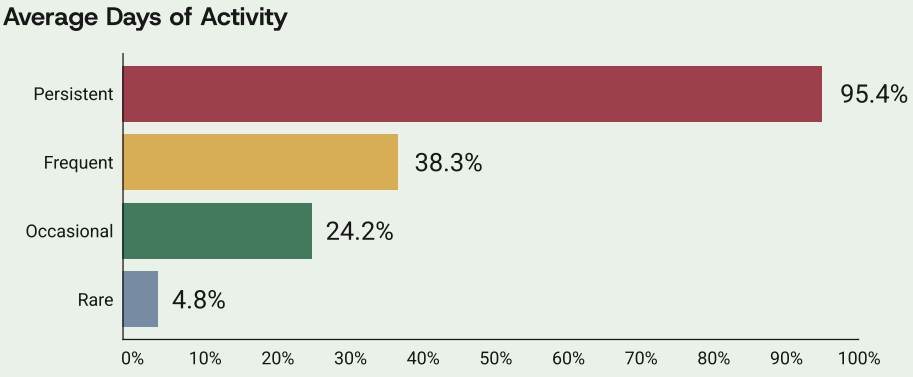


Some vulnerabilities are persistently exploited, a steady drumbeat that never lets up. Some are frequent but bursty. Some are occasional. Some are rare, a single spike and then silence. The proportions differ sharply by source. Among KEV entries with observed activity, 48 percent fall in the persistent cluster and 35 percent in the frequent one. Across all observed activity, persistent drops to 18 percent and the rare tail swells to 28 percent. The manually vetted list, in other words, is biased toward the loud cases, which is reasonable for a human process and misleading for an automated SLA.

All Activity vs CISA KEV



A “patch everything in seven days” posture is wasteful for a buried asset on the rare tail and far too slow for an internet-facing service in the persistent slice. A persistent vulnerability shows activity on 95.4 percent of days. A frequent one drops to 38.3 percent, an occasional one to 24.2 percent, and a rare one to 4.8 percent. That is a twentyfold spread between the top and bottom groups. A single SLA that treats all four as “known exploited” is averaging across populations that are very different in nature.



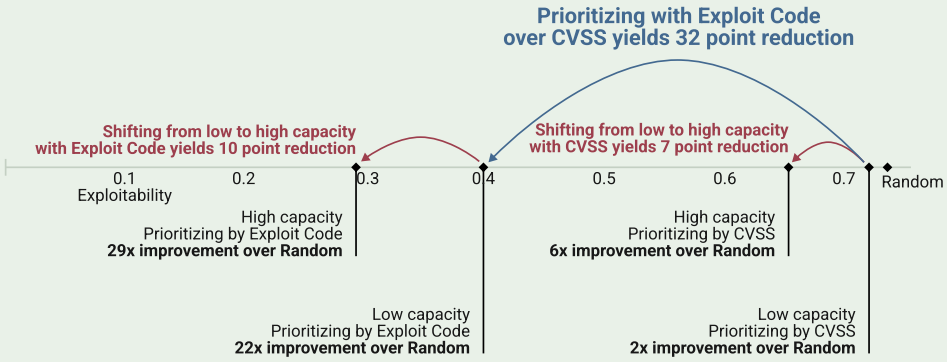
One detail still puzzles us, and it is worth stating plainly rather than dressing up. The activity matrices show vertical stripes, co-occurrence across unrelated CVEs on the same days, and an independent source that we did not include in the clustering shows the same streaks. They are too coordinated to be chance. Scanning botnets sweeping old payloads, common false-positive patterns in signatures, something else entirely: we have conjecture and not an answer. Honest models come with a list of the things they cannot yet explain.

Ignorance is Costly Bliss

There is nine years of evidence on what prioritization is worth, across more than 500 organizations and hundreds of millions of vulnerability instances. The headline number has not aged. Prioritizing by exploitation evidence rather than CVSS severity produces about a 29-fold improvement in efficiency for the same remediation effort. You fix far fewer things and prevent far more. That gap was already the central finding when the annual CVE total was a third of what it is now

The binding constraint was never discovery or remediation throughput. Prioritization is the constraint, and the volume curve makes that more true every quarter. When FIRST reports raw disclosures up 46 percent against forecast while exploitable risk holds flat, it is describing the exact shape of the problem that the Cyentia Prioritization to Prediction reports measured years ago: the problem has always been one of signal to noise rather than throughput.

Exploitability reduction achieved by improving remediation strategy and capacity



Dynamism

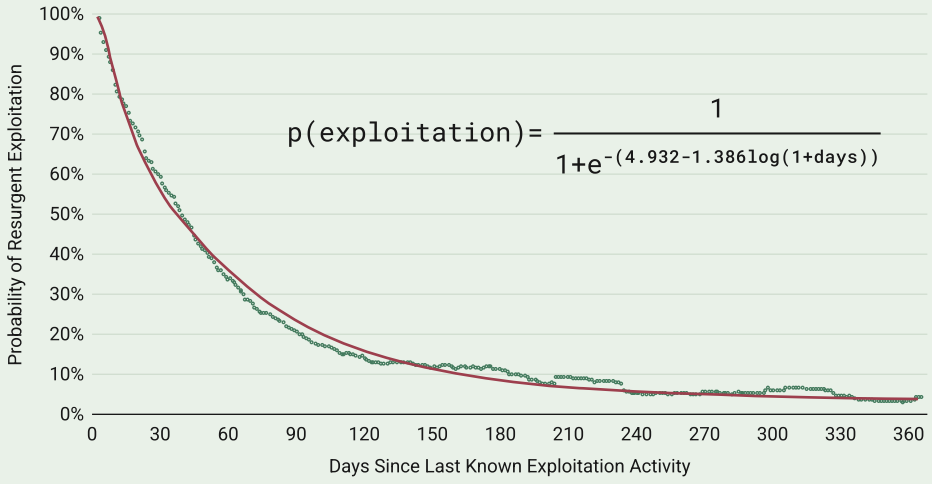
CVSS reports severity once and never updates, the way a thermometer reads a single moment. EPSS reads exploitation signal and moves with it.

KEV remains a list. None of them, on their own, answer the question a CISO is actually being asked, which is what is likely to happen to my organization in the next month, and what should my team touch first

That answer needs three things a static list cannot supply. It needs recency, because exploitation decays. It needs the local picture, because internet-wide probability and the realized risk to a specific environment are different numbers, and the gap between them depends on exposure, controls, and what an organization actually runs. And it needs calibration, because a probability that does not behave like a probability is just a vibe with a decimal point. A model that says 30 percent has to be right about 30 percent of the time to be useful operationally.

That said we can start with an incredibly simple model and build from there. If all we know about a vulnerability was when exploitation activity was last observed, a simple model of probability of re-exploitation would look like this (and be a pretty good fit to the data!):

Probability of Exploitation Resurgence



Winning by an inch

Speed comes in two forms, and both are getting worse. The first is volume. FIRST now expects something near 66,000 disclosures this year, and the count climbs every quarter with no ceiling in sight. The second is tempo. The window from disclosure to working exploit code keeps shortening, and for some classes of flaw the exploit arrives before the patch does. Every vendor selling a scanner will tell you these two trends are the emergency. They are not.

Speed only hurts you if you are racing on the wrong track. A program that patches by CVSS severity feels the volume increase as pure pain, because every new disclosure is another item in a queue that was already sorted badly. Doubling the disclosure count doubles the noise and obscures the signal, because “Actionable exploitability remains flat”.

So the answer to speed is not to patch faster. Very few organizations can win a throughput contest against a disclosure curve that has no upper bound, and the ones that try burn their remediation budget defending against ghosts. The answer is to sort better before you run. Tune your engine. A calibrated, time-aware probability turns the volume problem into a triage problem, and triage is a problem machine learning models are good at.

The tempo problem yields to the same move. A short disclosure-to-exploit window is dangerous only when you cannot tell in advance which disclosures will have exploitation activity at all. Most will not. The ones that will tend to announce themselves through the signals a global model already reads. A vulnerability that just went from quiet to active is exactly the case when a time-aware model escalates, and it escalates on the hour the signal appears rather than the day a human gets around to vetting it.

None of this requires patching everything in seven days. It requires knowing, on any given day, which few things are about to matter, and being willing to let the rest wait. It requires prediction, which is nontrivial. The volume can keep climbing and the exploit windows can keep shrinking, and a program built on prediction rather than binary flags will still be ahead. Winning is winning, whether you win by an inch or a mile. The trick is to stop measuring the distance and start measuring the odds.

EMPIRICAL

Empirical Security builds the foundational models for cybersecurity prioritization. Foundation, the global model, estimates exploitation probability across the internet. Radiant, the local model, estimates it for your environment. Both are calibrated, both are time-aware, and both plug into the stack you already run.

https://research.empiricalsecurity.com

EMPIRICAL
OBSERVATIONS