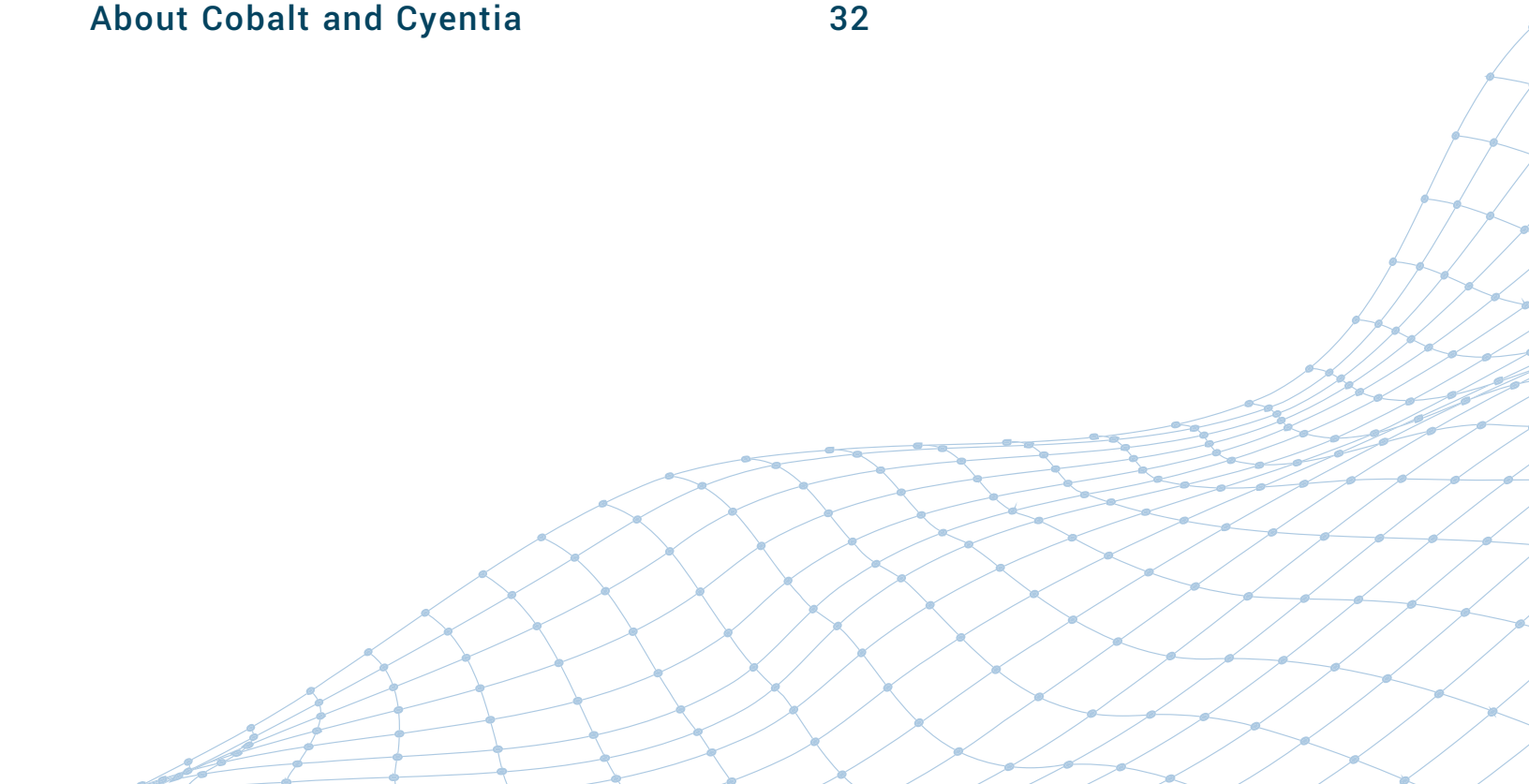
Cobalt | Cyentia Institute

State of Pentesting Report 2025

TABLE OF CONTENTS



FOREWORD

Thank you for reading the State of Pentesting Report 2025, our assessment of the results of thousands of pentests conducted via the Cobalt Offensive Security Platform. Looking back at the very first report in this series, published in 2019, it's remarkable how much the security landscape has changed, and how we've changed with it.

Cobalt, the pioneer in Pentesting as a Service, is now at the forefront of testing AI models and applications. We've found that AI security is lagging far behind the pace of AI adoption. Our LLM testing finds more vulnerabilities than any other type of test, and only 21% of the highest risk LLM vulnerabilities are resolved.

At the same time, security leaders are unfazed, but perhaps overconfident. While 81% are certain they meet security compliance, pentesting data tells a more complicated story. Although SLAs aim for two-week remediation windows, the real time to resolution often stretches to months or even years. It takes over three months for just half of the most serious issues to be resolved.

Development and security teams have made strides in reducing high-risk vulnerabilities, and the time to resolve these findings dropped by two-thirds over the past decade. These improvements were likely driven by increased adoption of structured pentesting programs over ad hoc testing, and more rigorous security standards during software development.

However, the persistence of unaddressed, exploitable vulnerabilities underscores the need for programmatic approaches that go beyond meeting SLA timelines to ensure true risk reduction.

At Cobalt, we understand the dangers of hidden security vulnerabilities, including ones frequently missed by automated scanners. As our industry evolves alongside growing investment in AI solutions, we continue to innovate. Yet our identity remains unchanged, centered around the expertise of the Cobalt Core, our team of over 450 pentesters.

Within these pages, you'll find an in-depth examination of pentesting results, and insights about the challenges security teams are facing. We expect these insights to help security leaders guide their organizations towards better offensive security programs, for greater assurance and reduced risk.

Gunter Ollman
Chief Technology Officer | Cobalt

Jason Lamar
SVP of Product | Cobalt

EXECUTIVE SUMMARY

Knowledge is power—and in security, that knowledge must come from the right insights.

Security leaders feel confident in their posture, but pentest data tells a more complex story: Critical vulnerabilities often remain unresolved, hidden beneath the surface of automated scans and service-level agreement (SLA) checkboxes. Even as remediation speeds improve, one-third of serious issues still slip through the cracks—and with genAI introducing new, high-impact risks, traditional approaches fall short.

That's why pentesting is essential. It transforms assumptions into evidence, surface-level confidence into actionable clarity. Structured, expert-led pentesting delivers the knowledge security teams need to understand their true risk—and the power to reduce it.

Key Findings

THE PERCEPTION AND REALITY OF SECURITY DON'T ALWAYS AGREE

Most organizations (81%) are quite confident that their security posture is up to snuff. Yet pentesting never fails to reveal hidden vulnerabilities.

14 DAYS

Three-quarters of organizations have set SLAs specifying that pentest findings should be fixed in two weeks or less. Few meet this goal.

67 DAYS

The median time to resolve (MTTR) stands at 67 days for all pentest findings. That's five times longer than the two-week SLA set by most organizations.

Less than half (48%) of all findings ever get resolved. That jumps to 69% for the riskiest findings, but that still leaves many open vulnerabilities.

AI IS QUICKLY EMERGING AS A MAJOR SECURITY RISK

Nearly all firms say they're integrating genAI into their products; just 66% are actively assessing the security (including pentesting) of these solutions.

32% of all LLM pentest findings are rated high risk. That's 2.5 times the overall proportion of 13% and the highest level among all types of pentests we conduct.

Only 21% of serious findings from LLM pentests get resolved. That's understandable given the newness of the tech. But that doesn't negate the risk.

PROGRESS IS BEING MADE TO REDUCE RISK, BUT THERE'S MUCH MORE TO DO

94% of firms view pentesting as essential to their program. But evidence suggests many aren't taking a programmatic approach to offensive security.

2X

In 2017, only 27% of serious pentest findings were resolved. That proportion soon doubled to 55% but has hovered at that level ever since.

-75 DAYS

In 2024, serious findings were fixed in one-third of the time it took back in 2017 (37 versus 112 days) That's shaving 75 days off the exposure window!

+34 DAYS

The largest organizations take over a month longer to resolve serious findings than the smallest firms (61 versus 27 days). Managing risk gets harder as you grow.

PENTESTING PERSPECTIVES

Before diving into all the issues Cobalt pentesters identified, let's review what we learned from our survey of 450 security leaders and practitioners about how penetration tests fit into their security programs.

Let's start with the most common internal justifications for conducting pentests. The top reason selected by 94% of respondents was that pentests are foundational to ensuring a strong security posture. This captures the assurance role of pentesting and reflects the reality that most breaches don't occur because the victim had no defenses. Rather, the defenses they had weren't as solid as they thought.

It's probably no surprise to learn that most respondents (91%) chose compliance as a major reason why they do pentests. What may surprise some is that even more of them (92%) say pentests are important to their organization's strategy and senior leadership. Over three-quarters of firms (and 92% of retailers) claim that pentesting improves customer trust.

Let's keep pulling on the "improves customer trust" thread. We asked participants about the types of security assurance commonly requested by their customers and regulators. Third-party pentest reports were the most common selection, at 59% of respondents. It's noteworthy that pentests rate higher than vulnerability scans and compliance certifications.

FIGURE 1

Why does your organization conduct pentests?

FOUNDATIONAL FOR MODERN PROGRAMS TO ENSURE A STRONG SECURITY POSTURE

IMPORTANT TO OUR ORGANIZATION'S STRATEGY AND SENIOR LEADERSHIP

CRITICAL FOR MEETING OUR BUSINESS AND COMPLIANCE GOALS

IMPROVES CUSTOMER TRUST IN OUR SOFTWARE PRODUCTS

FIGURE 2

What types of security assurance do your customers or regulators most commonly request to validate your software's security?

3RD PARTY PENETRATION TEST REPORT

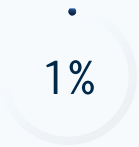
3RD PARTY CODE REVIEW

INTERNAL SECURITY TEST RESULTS

VULNERABILITY SCAN REPORTS

COMPLIANCE CERTIFICATION (SOC 2, ISO 27001, ETC)

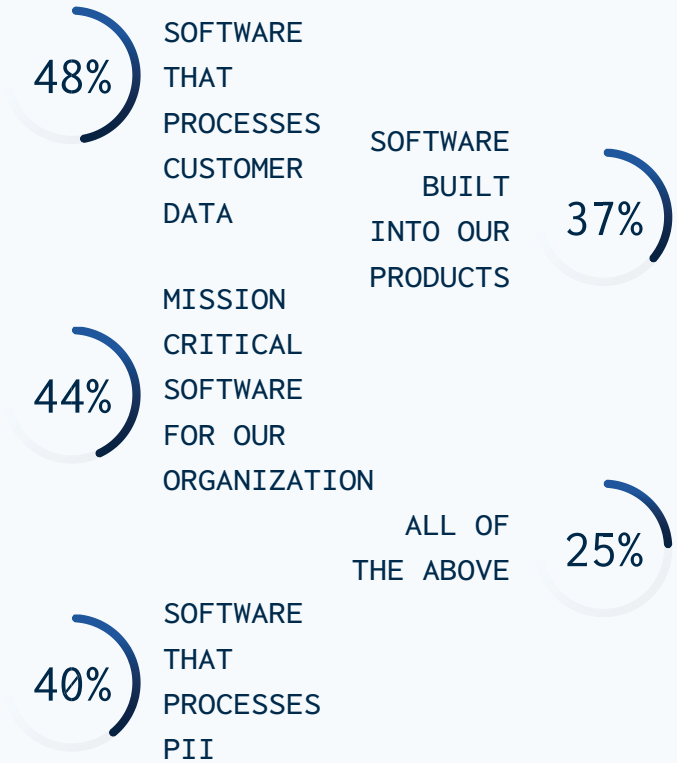
NONE OF THE ABOVE



There are several factors that tend to trigger requests to validate third-party software using pentests. Top choices include software that processes sensitive data or is critical to the organization's mission. Over a third of organizations test all vendor software that gets built into their products, though including those that answered "all of the above" raises that to over 60%.

FIGURE 3

What type(s) of commercial software do you require a pentest from vendors?



Pentests are also seen as a way to reduce liability for security issues. Per the responses below, that liability is seen at both the personal and corporate levels. Furthermore, a quarter of respondents believe security assurance to be so important that they've considered quitting when that obligation isn't taken seriously enough. Keep these (rather idealistic) responses in mind when we get to the section analyzing the reality of pentest remediation timelines.

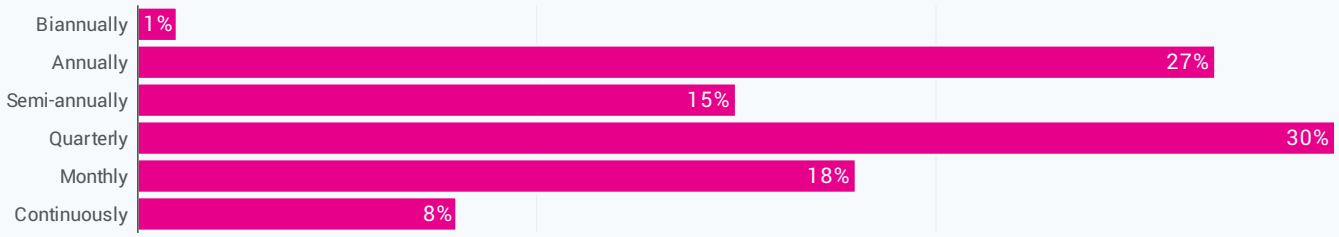
FIGURE 4

Concerns over liability related to unaddressed security issues

It's obvious from the responses above that pentests are viewed as an important part of modern security programs. But how often are they conducted? Just over a quarter of respondents do annual pentests, and another 15% say it's a semi-annual activity for their firms. That means more than half of organizations conduct pentests more than twice a year, with quarterly being the most common cadence (30%).

FIGURE 5

How often does your organization conduct pentests?



Taken together, these survey responses suggest that proactive pentesting is widely viewed as an essential part of cybersecurity programs as well as a major business driver.

The next section analyzes findings from penetration tests conducted via Cobalt over the last decade. You'll see right from the start that the perspectives shared here don't always align with reality.

TOP PENTESTING FINDINGS

According to most survey respondents, this should be a fairly short section. More than 8 in 10 of them expressed confidence that their organization's security posture meets all relevant regulatory requirements. The fact that this is not a short section but is rather filled with high-risk pentest findings suggests a widespread tendency for overconfidence.

FIGURE 6

I am confident that my organization's security posture meets all regulatory requirements we are subject to.

CONFIDENT IN REGULATORY REQUIREMENTS

Regardless of your view on whether compliance equals security or not, we can probably all agree that the feeling of being compliant or secure doesn't mean you've actually achieved either state. In fact, that's one of the main reasons why penetration tests exist. It's better to expose security issues now than turn a blind eye until attackers exploit them later.

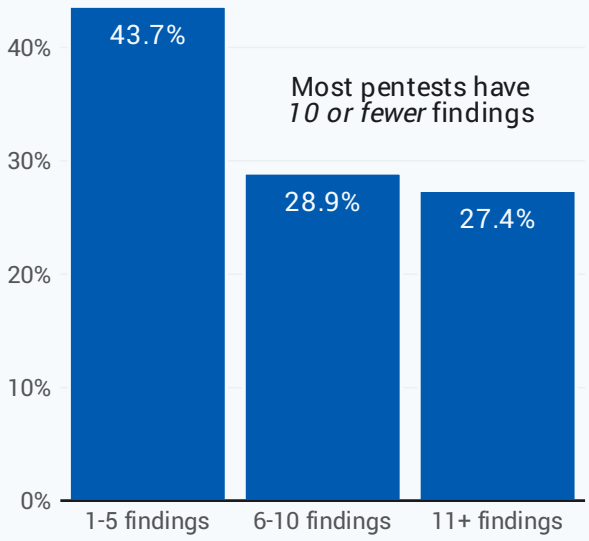
Cobalt pentesters follow specific methodologies depending on the type of test they're performing. By default, that includes industry-standard vulnerabilities from the Open Web Application Security Project (OWASP), which includes different top 10 lists for web, API, mobile, AI/LLM, and cloud systems. The Open Source Security Testing Methodology Manual (OSSTMM) is used for pentests of internal and external networks. More details on each of our pentesting methodologies can be found on the Cobalt website.

Pentest findings shared with the customer include exploitation details, impact assessment, and proof of concept with steps to reproduce.

Every pentest we conducted uncovered at least one reportable finding[1], but the median number was six. The majority of organizations had 10 or fewer findings. That said, over a quarter of pentests revealed more than that, and some resulted in a list of issues that stretched into the low hundreds. All sector and size groups showed this same overall distribution with only slight variation.

FIGURE 7

Number of findings in each pentest



We suspect the number of findings depicted here may strike some as low. It's much lower than, for example, the number of issues typically identified with a vulnerability scanner. But keep in mind that pentests are often focused on a particular asset or group of assets rather than the entirety of enterprise infrastructure.

[1]"Informational" findings have been removed from these statistics because they are not exploitable and therefore represent minimal risk.

Furthermore, findings uncovered by vulnerability scanners and code analysis are more theoretical in nature. They're weaknesses that, given the right circumstances (which usually aren't accounted for), could be exploited.

Pentest findings, on the other hand, do account for relevant circumstances (e.g., accessibility and mitigating controls) and are proven to be exploitable by the pentester. The outcome is a smaller set of real risks.

As previously mentioned, pentests are specific to certain types of assets. This section presents the most prevalent findings associated with three popular pentest methodologies, starting with web applications (which include APIs and microservices).

Pentest findings are grounded in human expertise, revealing how vulnerabilities can be actively exploited in real-world scenarios. Unlike scanner results, pentests go beyond the theoretical to uncover the full scope of exposures and highlight what truly poses risk to your organization.

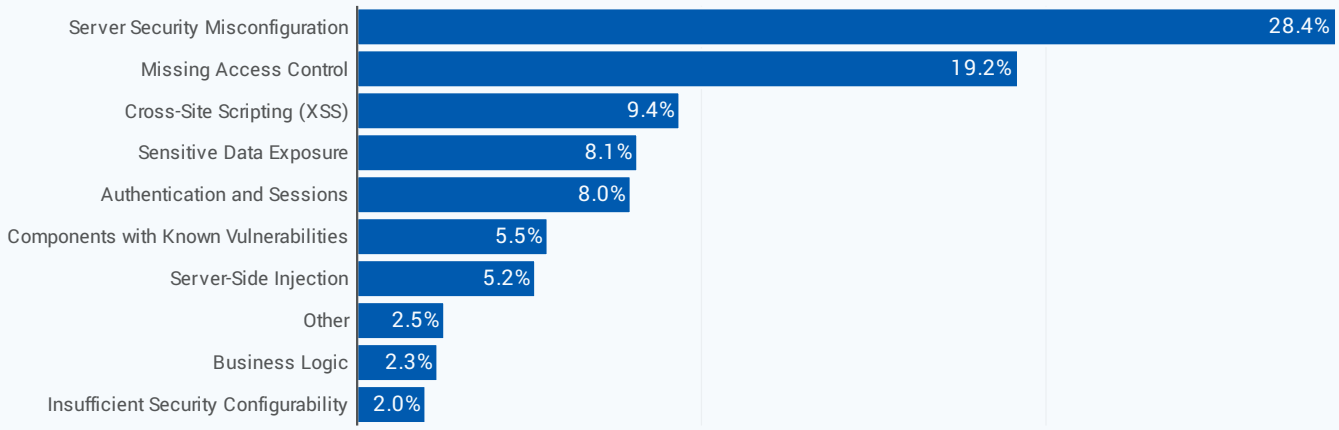
Web applications and APIs

Web applications serve as the public and/or customer-facing presence for many organizations, which means they're reachable by attackers too. That makes identifying and remediating vulnerabilities all the more important.

Figure 8 shows that server misconfigurations are the most common type of finding. This category has a broad scope from OWASP, referring to improper implementation of security controls on a server that leaves it vulnerable to exploitation. This can result from default settings, unnecessary services, misconfigured permissions, or missing security patches.

FIGURE 8

Most common web and API pentest findings in 2024 (all criticalities)



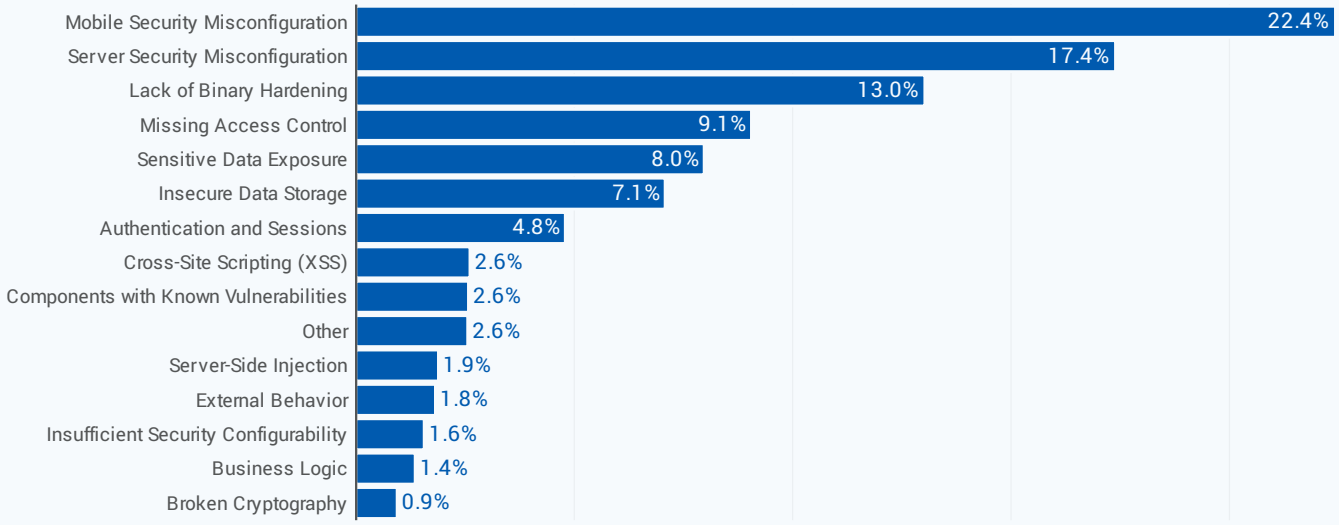
Missing access control lands at number two, indicating that improper enforcement of authentication and authorization measures is a significant issue, allowing unauthorized users to access sensitive data or functionalities. Identified in just shy of 1 in 10 pentests, cross-site scripting rounds out the top three. Along with sensitive data exposure (8.1%) and authentication and session vulnerabilities (8%), this highlights persistent weaknesses in handling user inputs securely and protecting confidential information.

Mobile applications

Top findings from mobile application pentests are revealed in Figure 9. Modern mobile apps are so tightly coupled with web services that it's not surprising that the findings are similar. Security misconfigurations once again take the lead. and missing access control remains among the more common findings. Issues stemming from a lack of binary hardening in mobile apps allow attackers to reverse engineer and modify code to enable nefarious functions.

FIGURE 9

Most common mobile application pentest findings in 2024 (all criticalities)



Mobile security misconfiguration is a broad category that includes findings such as a lack of SSL pinning, no jailbreak or root detection, unnecessary software components, and having the clipboard enabled. Many of these exist because the business has deemed them an accepted risk.

The presence of server misconfiguration issues may seem out of place here, but part of mobile testing is checking the API connections between the application and the mobile device. In general, secure enclave architectures have done a great job of protecting mobile devices and apps from vulnerabilities in other apps. Vulnerabilities in connected SaaS components, however, can be exploited and used as gateways to wider access. In that sense, it's not surprising that mobile pentest findings share similarities with application pentests above.

AI and LLMs

AI is the driving force behind business and technology trends today. For a growing number of companies, remaining competitive means integrating AI into their workflows and products. These trends are reflected in responses from survey participants. Nearly all of them (98%) say their organizations are currently integrating genAI into their products and services.

These capabilities, of course, introduce new and unique threats that must be managed. The organizations we surveyed seem to understand this well. Securing genAI tops their list of concerns, ranking above known exploited vulnerabilities, insider threats, and nation-state attackers. About 80% say they're increasing efforts to secure AI, and 66% are actively conducting regular security assessments (including pentesting) of their AI-powered solutions.

FIGURE 10

A risky combo: Breakneck adoption and lagging security of genAI

"WE'VE INTEGRATED / ARE IN THE PROCESS OF INTEGRATING GENAI INTO OUR PRODUCTS."

"WE'RE CONDUCTING REGULAR SECURITY ASSESSMENTS AND PENTESTS OF OUR AI-POWERED PRODUCTS."

As the race to integrate AI heats up, organizations are prioritizing securing genAI above securing known vulnerabilities.

What is it about genAI threats that so many organizations find concerning? We asked survey participants that very question, and their responses ranged from technical and operational issues (e.g., data exposure and integration) to governance and ethical challenges. The breadth of the concerns expressed is indicative of the uncertainty surrounding the security and safety of AI and LLM applications.

FIGURE 11

What concerns you most about AI/LLM-related threats for customer-facing applications?

DATA PRIVACY AND UNINTENDED EXPOSURE OF SENSITIVE DATA

PROMPT INJECTION ATTACKS AND ADVERSARIAL MANIPULATION

MODEL BIAS AND ETHICAL RISKS

INSECURE INTEGRATION WITH EXISTING APPLICATIONS

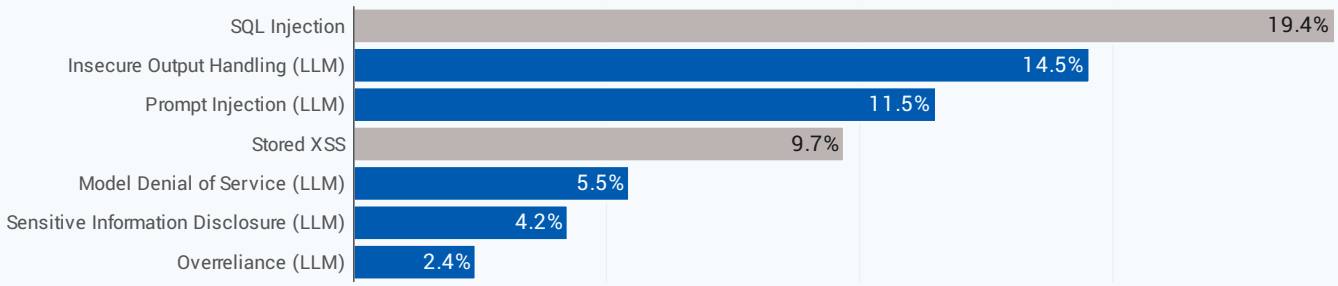
LACK OF VISIBILITY INTO AI-GENERATED OUTPUTS

Cobalt's AI testing follows the OWASP Top 10 for LLM Applications, focusing on sensitive data exposure, insecure output handling, and injection attacks that could compromise model integrity. Our pentesters conduct dynamic testing to detect prompt injections and model-based DoS vulnerabilities, while also assessing LLM production services and plugins for unauthorized data exfiltration or excessive system access. This comprehensive approach ensures robust security in LLM environments, addressing risks traditional models may overlook.

The top five findings from AI and LLM pentests in the last year are highlighted in Figure 12. We've included common non-AI issues identified by these tests as well (in gray). This is a reminder that AI applications are still prone to the same old findings that affect any application. But we'll highlight the AI and LLM findings here.

FIGURE 12

Most common AI and LLM pentest findings in 2024 (all criticalities)



The most prevalent LLM-specific findings concern insecure output handling, where improper validation of model-generated responses can lead to data leaks, injection attacks, or security misconfigurations. Prompt injection vulnerabilities follow closely, which could enable attackers to manipulate model inputs to bypass safeguards or alter system behavior.

AI is changing quickly, and so are the risks. OWASP understands this and updated the 2025 edition of the Top 10 for LLM and genAI to expand on DoS and other availability issues. The new category is called Unbounded Consumption, and includes threats like Denial of Wallet (DoW), which attackers can use to exploit the cost-per-use model of AI services.

Issues that open the LLM to DoS attacks come next, which means adversaries can overload or disrupt LLMs, impacting availability. Findings associated with sensitive information disclosure highlight the risk of LLMs inadvertently exposing confidential data due to improper access controls. Lastly, overreliance refers to the tendency for misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

These findings underscore the need for robust security measures, including input validation, response filtering, and strict access controls to mitigate emerging LLM-related threats.

Download Cobalt's whitepaper The Responsible AI Imperative: Why Secure AI Is the Only AI That Matters to explore AI risks, pentesting challenges and methodologies, and best practices for securing AI applications.

Severity of findings

As you likely suspect, not all security issues uncovered during a pentest carry the same risk for the organization. That's why we rate the severity of all findings on their likelihood of exploitation and the potential impact on technical and business operations.

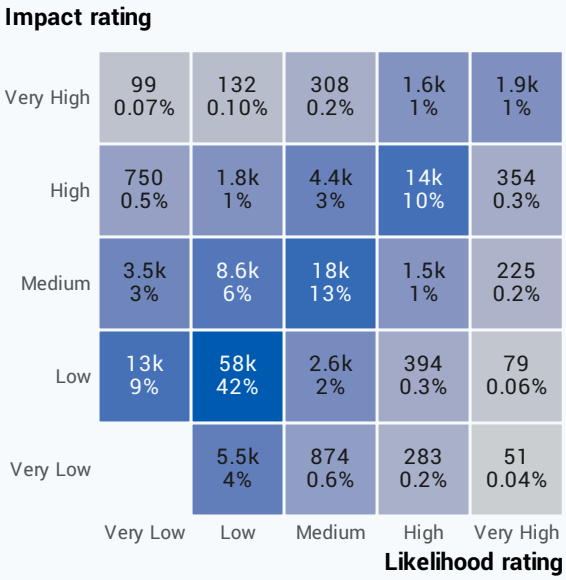
The product of these two ratings derives each finding's level of risk. A rating of five on both scales yields a critical risk (25). High-risk findings score from 16 to 24 and so on. A proportional breakdown of severity for all findings in our sample is portrayed in Figure 13.[2]

Much of our analysis going forward will focus on findings with a severity rating of high or very high (four boxes in the upper right), which we'll dub "serious." That designation represents about 13% of all non-informational findings from pentests conducted over the 10 years of our data. We'll now explore how that ratio varies across tests and organizations.

All types of pentests can identify serious security issues, but some tend to do so more than others. Case in point, about one-third (32%) of the findings uncovered during our AI and LLM pentests warrant a serious rating. The fact that organizations are integrating AI capabilities at breakneck speed, combined with the nascency of security practices for those capabilities, undoubtedly contributes to this (see Figure 10).

FIGURE 13

Severity rating of pentest findings and breakdown by severity



13% ABOUT 13% OF ALL FINDINGS RATED SERIOUS (HIGH LIKELIHOOD AND SEVERITY)

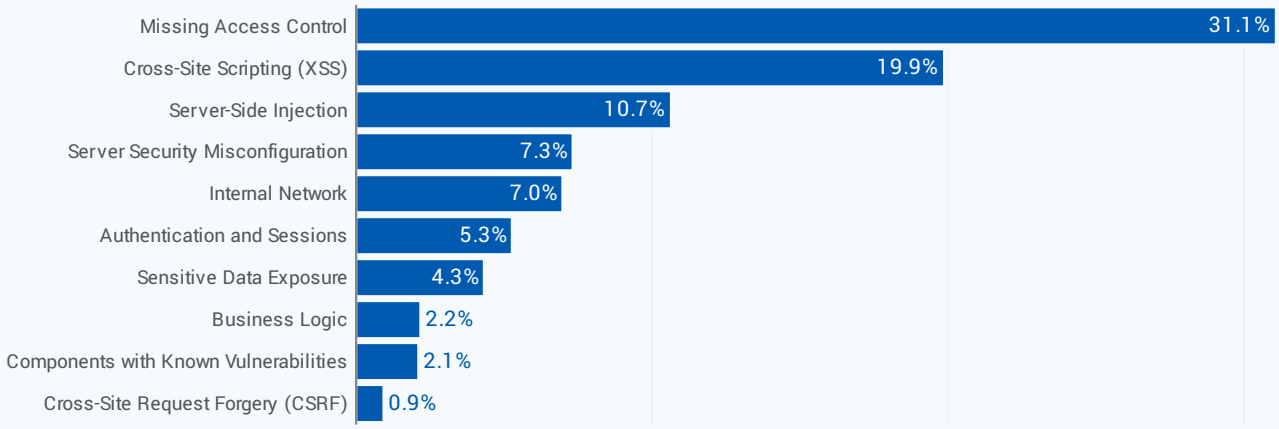
32% 32% OF AI/LLM FINDINGS RATED SERIOUS

[2]Informational findings (rating of very low on both scales) have been removed from the chart.

Which types of findings tend to be the most serious? Figure 14 gives the relative frequency for the top 10 most common vulnerability types among serious findings. If you're looking to make big strides in reducing your risk exposure, these offer a great starting point. This blog post on top web applications vulns offers additional context on these findings.

FIGURE 14

Most common vulnerability types among serious findings (2024)



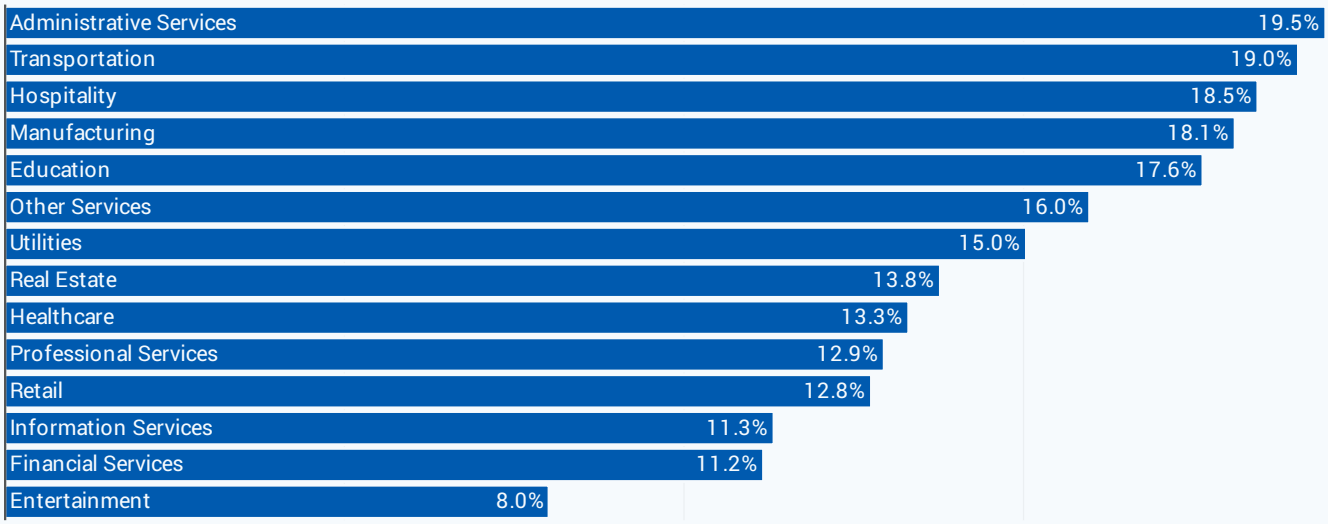
Let's close out this discussion with a brief look at how serious findings vary among different types and sizes of organizations. We'll start with size because there's not much to show or say on that aspect. The prevalence of serious findings is virtually the same across all size tiers. So, there's no evidence to conclude that SMBs or large corporations are less prone to risky issues.

The sectors in Figure 15 and other charts in the report use the North American Industry Classification System (NAICS). We've linked the first reference to each sector to the corresponding page on the NAICS website for those who want more detailed descriptions.

There is some evidence, however, that pentests are more likely to find serious issues in certain industries. Figure 15 reveals that the Administrative Services, Transportation, Hospitality, Manufacturing, and Education sectors have the highest proportion of serious findings, but several others are right up there with them.

FIGURE 15

Prevalence of serious findings by industry



The Entertainment industry (composed mainly of gaming, gambling companies, and video streaming companies) stands out with the lowest ratio. The Financial Services and Information Services sectors also show low rates, which suggests they manage issues well despite having high volumes of sensitive data and critical systems. Plus, stringent regulatory requirements may give them extra incentive to minimize risk exposures.

Resolution of findings

With many types of tests, the hard part is in preparing for the exam. Once it's done, you move on. Pentests are different in that the hard part comes after the exam results are posted. We're talking, of course, about fixing or otherwise resolving security issues identified by the pentest. That's where the real work begins.

There's a lot vying for the attention of security teams these days, and it's impossible to fix all the issues from all the things all the time. According to survey respondents, pentests rank high on the list of things that should get priority attention from security teams. It's tied for second with known exploited vulnerabilities and right behind alerts from threat detection and response tooling.

Despite the consensus that pentests demand attention, our data reveals that less than half (48%) of all pentest findings actually get resolved. That ratio does improve to 69% among serious findings, indicating intent to prioritize issues that represent the highest risk. The fact remains, however, that many security issues identified by pentests—and therefore exploitable by attackers—never get fixed.

If this resolution rate seems surprisingly low to you, you're not alone. We'd love to see 100% of pentesting findings resolved and work to support clients in pursuing that goal. The silver lining here is that most firms are doing better than these overall stats indicate. Over half of organizations have resolved 90% or more of their serious pentest findings, proving that the goal is practically achievable. That said, there are some organizations fixing very few of their risk issues, which pulls that overall ratio down.

FIGURE 16

Proportion of all and serious findings resolved

ALL FINDINGS RESOLVED

SERIOUS FINDINGS RESOLVED

57% OF ORGS RESOLVE 90% OR MORE OF SERIOUS FINDINGS

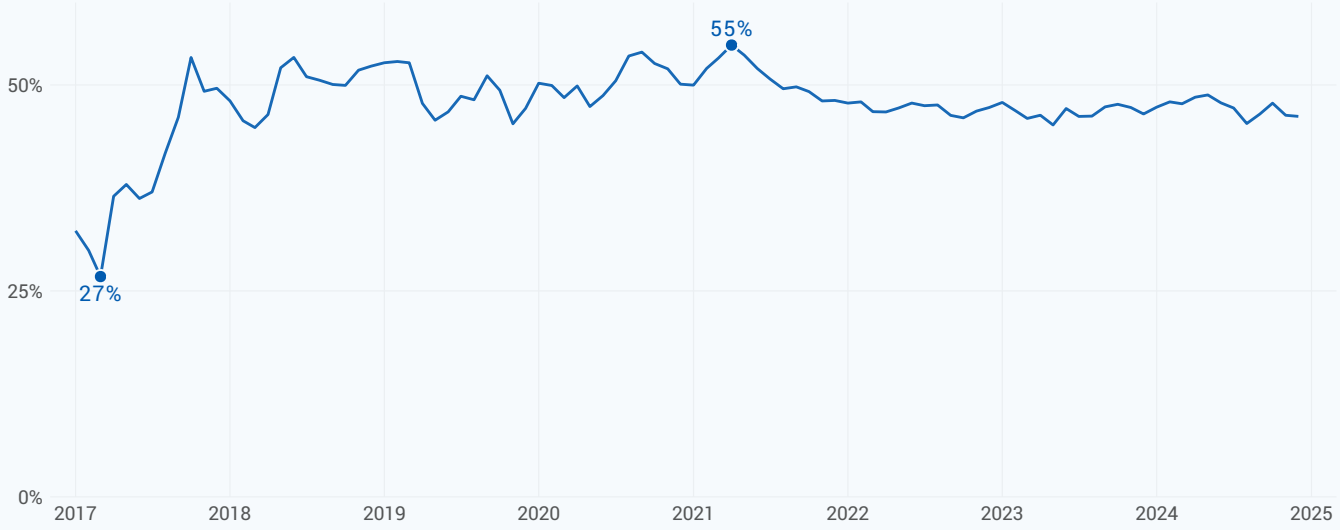
15% OF ORGS RESOLVE 10% OR LESS OF SERIOUS FINDINGS

Pentesting is the second most prioritized amongst security teams, however, many security issues identified by pentests never get fixed.

Figure 17 traces a rolling trend of the percentage of serious findings resolved in the prior 12-month period.[3] The rapid improvement in the first couple of years is largely attributable to a rapid expansion in the number and types of pentests conducted in the early days of our services. Minor fluctuations aside, the trend has been remarkably consistent over the past several years. The impression one gets is that a kind of stalemate has been reached, whereby organizations aren't making forward progress against high-risk findings in their environment.

FIGURE 17

Proportion of serious findings resolved over time



These facts prompt the obvious question, "Why?" We've observed many reasons over the years regarding why findings from our pentests aren't acted upon. Some organizations do only what they're required to do for compliance or third-party approval—get a pentest. Remediating risk is of less immediate concern.

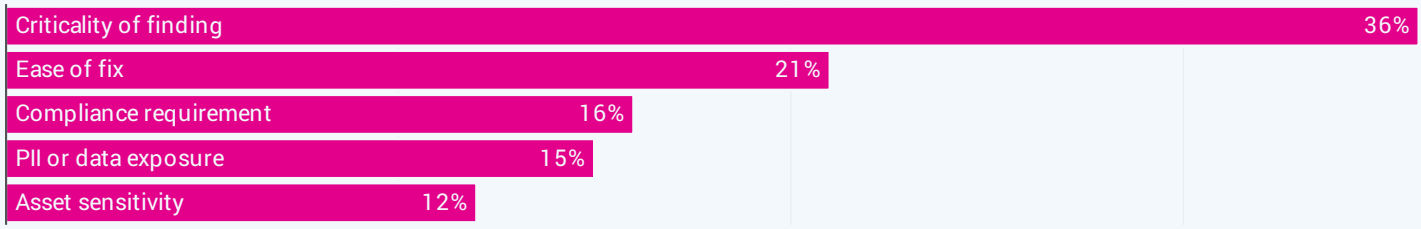
For the most part, though, it comes down to a host of organizational issues spanning people, processes, and technology. We've seen situations where those responsible for addressing findings are in another group with different priorities than the one that ordered the pentest (think developers and security teams). Since pentest vulnerabilities often fall outside of patch management, the more complex remediation processes challenge less mature teams. Technology roadblocks to remediation often come in the form of legacy or fragile systems that firms are loath to mess with. Resource constraints can impact all the above and more.

[3]The percentages of serious flaws resolved in this rolling trend are lower than the all-time stat (69%) because we're only looking at findings found and fixed in the prior 12 months. The all-time stat does not have a limit, so credit is given for any resolved finding no matter how far back it goes.

It's also useful to ask the opposite question: Why do pentest findings get resolved? We put that question to survey participants, and their responses are tallied in Figure 18. There was consensus on criticality as the top driver for remediation, which aligns with our observation that serious pentest findings are more likely to be resolved.

FIGURE 18

What is your top priority for addressing security findings?



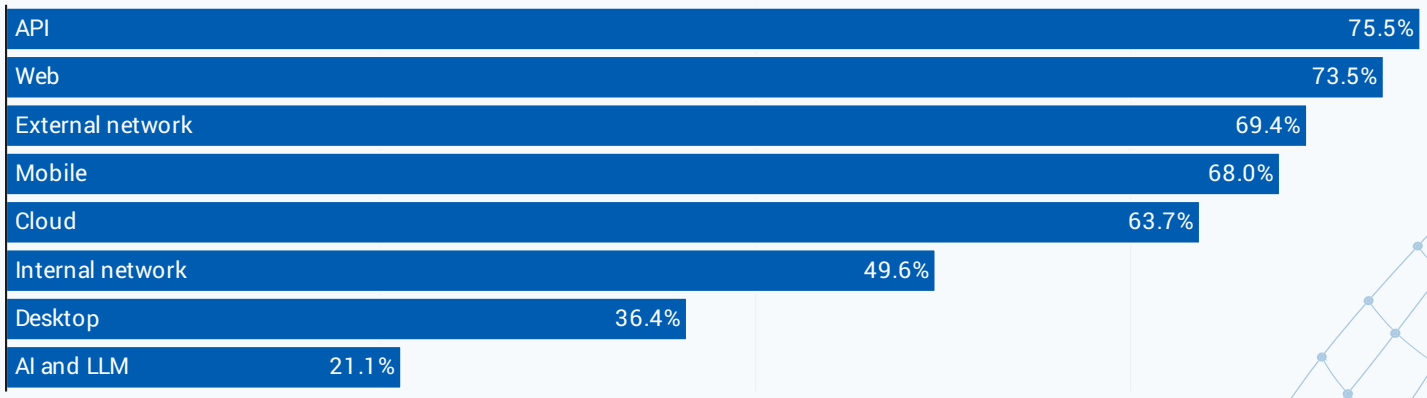
But criticality is far from the only reason that issues get prioritized. Fixing the easy stuff may at first seem like the lazy approach to remediation, but it's actually very practical. Why wouldn't you squash a moderate bug if it's quick and easy to do so? That delivers fast value to the customer.

The last three drivers are similar and, in many cases, intermingled. For example, a vulnerability that places PII at risk would need to be resolved to attain or maintain compliance. Or an asset would be considered sensitive because it stores very sensitive data. These have more to do with what's at risk than the finding itself.

Peering deeper into the resolution of serious findings, there are some major differences among pentest methods. Web and API having the highest resolution rate is no surprise. Product and development teams are often the ones ordering those tests and are generally motivated and capable of addressing issues. Remediating network and desktop findings, on the other hand, often requires other teams (e.g., IT) or service providers to fix them.

FIGURE 19

Resolution of serious findings by pentest method

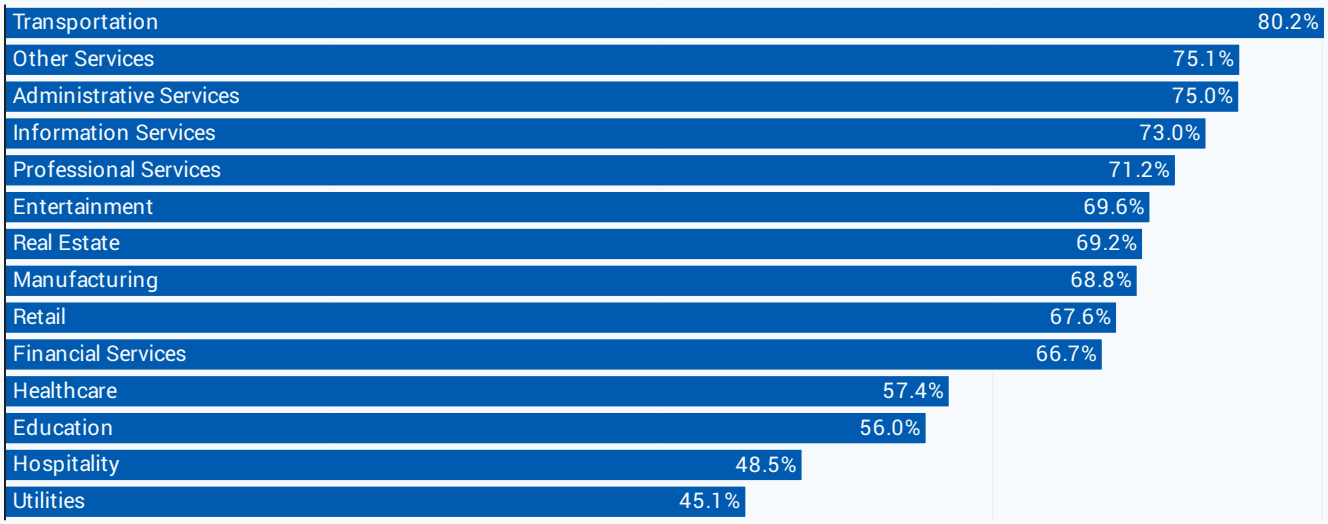


AI and LLM tests show the lowest proportion of serious findings resolved, which makes sense given the relative newness of that specialty area. Software security has decades of maturing under its belt, which has spawned policies and refined practices. GenAI hasn't been around long enough for that, and many teams lack the expertise to adequately address findings. Furthermore, some issues can only be fixed by the company that developed the model.

Organizational characteristics and capabilities also contribute to the scope of resolution. Figure 20 shows some degree of variation among industries in the proportion of serious findings that get resolved. The Utilities, Hospitality, Education, and Healthcare sectors exhibit the lowest rates. That's especially concerning in the case of Utilities and Healthcare, where unresolved exposures can put human wellness and safety at risk.

FIGURE 20

Resolution of serious findings by industry



Several other industries range in the vicinity of the overall resolution rate for serious findings of 69%. The Transportation, Other Services[4], and Administrative Services sectors are the only ones that manage to close out at least three-quarters of their serious findings. We suspect this may be more a reflection of comparatively smaller and less complex attack surfaces to manage, rather than of superior remediation capabilities.

In a twist that may surprise some, small firms are better at fixing pentest findings than larger organizations. Well, perhaps "better" isn't the right term because there's more going on here than just ability. SMBs tend to have a smaller scope of testing and fewer findings to address. Furthermore, many of our SMB clients focus their pentesting on a particular product that's core to their business to meet customer or third-party requirements.

FIGURE 21

Resolution of serious findings by organization size

[4]This NAICS designation is admittedly vague. It consists of everything from nonprofits to janitorial services to true "others" that simply don't fit elsewhere.

Enterprise security teams, despite having more resources, contend with complex environments and processes. They tend to have a larger share of legacy applications that either can't readily be updated or are updated at longer, more rigid intervals (i.e., a waterfall approach to development). This won't be the last time that we'll see this pattern of larger organizations struggling to keep up with pentest findings.

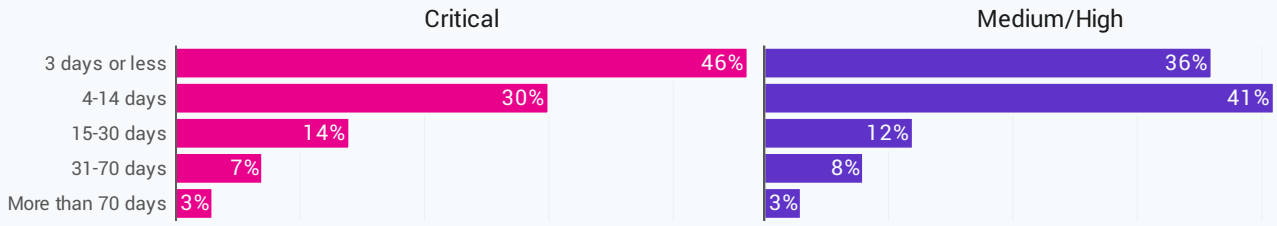
Time to resolution

To be fair, you could add "so far" to all the statements we just made regarding the proportion of pentest findings resolved. Some of the issues that persisted when we closed data collection for this report have been addressed by now or will be in the future.

This brings up the important topic of how long it takes to fix findings. There's a stark difference between perception and reality when it comes to resolution timelines. Let's start on the perception side by reviewing what survey respondents told us about their SLAs for pentest findings.

FIGURE 22

What is your SLA to fix vulnerabilities once identified?



There's a stark difference between perception and reality when it comes to resolution timelines

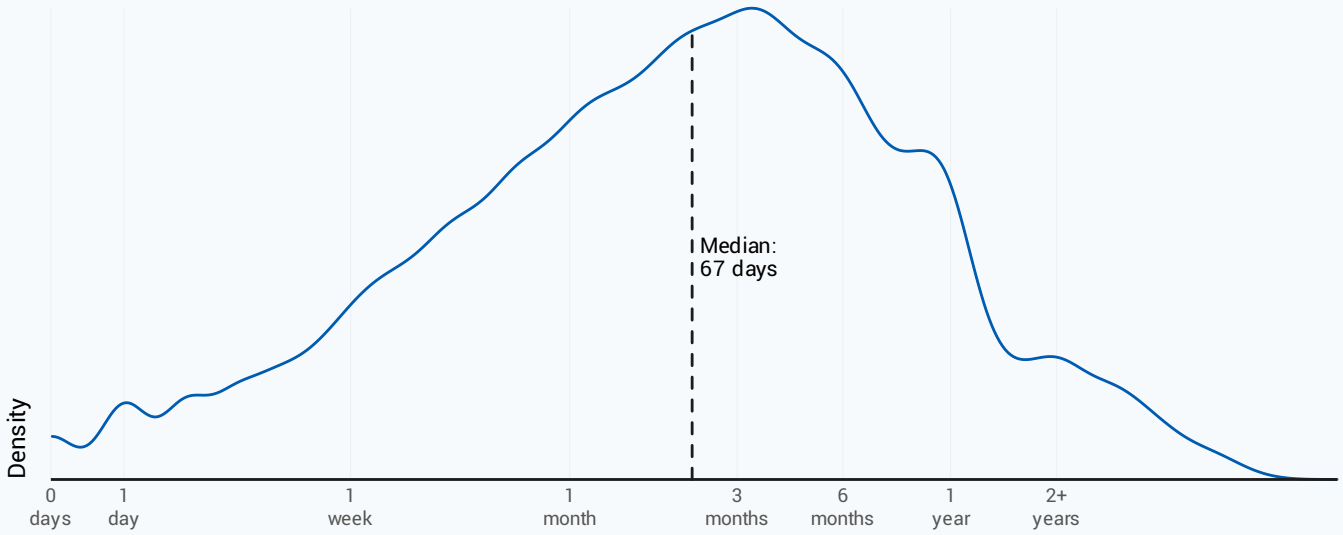
Figure 22 sets a fairly aggressive resolution timeline regardless of criticality. Three-quarters of organizations have SLAs stating that vulnerabilities should be fixed within two weeks of being identified. That bumps to 90% of firms if we extend to one month. So, that's the perception of how quickly pentest findings should be fixed.

Figure 23 reveals the reality of resolution, and it's different, to say the least. The overall MTTR stands at 67 days across all types of findings for all organizations in our database of pentest results. That's five times longer than the two-week SLA set by three-quarters of survey participants.

The "M" in most measures of MTTR stands for "Mean" rather than "Median." We've opted to use the median because it's a better statistical measure of "typical" when extreme values are present (which definitely exist in resolution timelines).

FIGURE 23

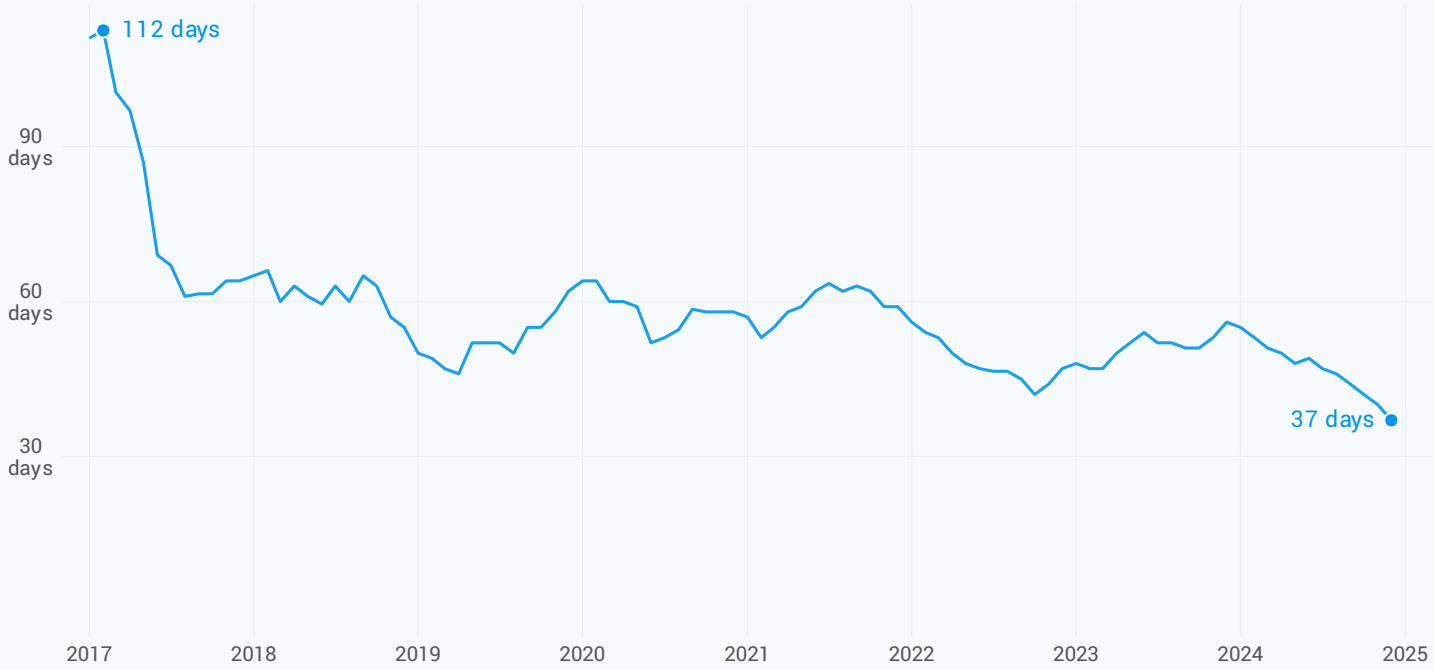
Overall distribution of time to resolve pentest findings



Serious findings tend to be resolved a couple of weeks faster (MTTR of 50 days), but that still falls well short of target SLAs. The good news from Figure 24 is that the gap is closing. In 2024, serious findings were fixed in one-third of the time it took back in 2017 (37 versus 112 days). Give yourself a pat on the back, everyone … but don't get complacent. Let's bust through the 30-day threshold by next year's report.

FIGURE 24

Median time to resolve serious findings (rolling 12 months)

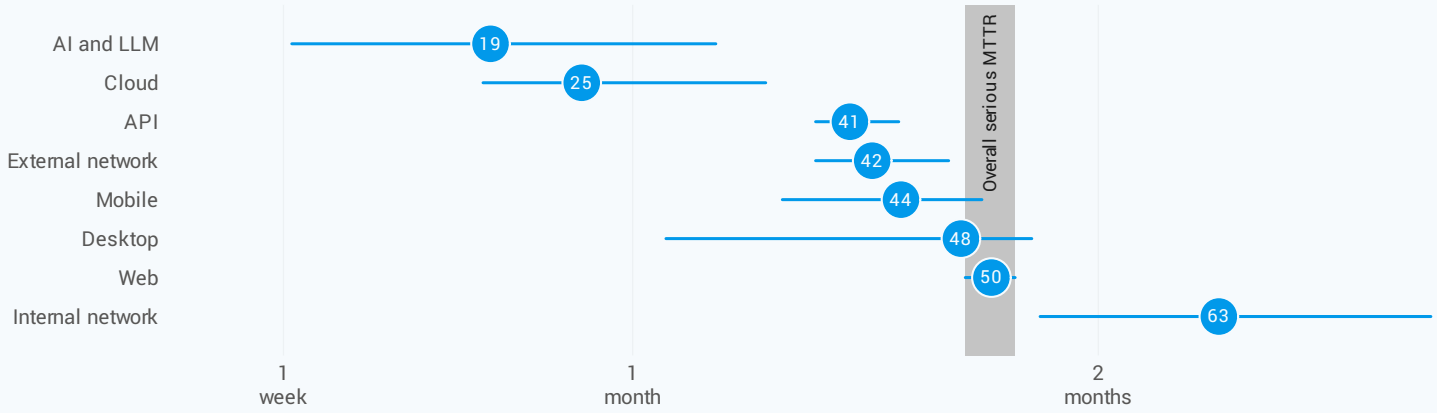


If resolution times speed up for serious findings, do they also vary based on other factors? Let's find out, starting with a comparison of serious findings among different types of pentests in Figure 25. The dots mark the MTTR for each method, while the lines extending from them establish a 95% confidence interval. The width of those intervals derives from the sample size and variance around MTTR within each type of test.
Even though some of the MTTRs appear to be quite different, the overlapping confidence intervals indicate those differences aren't statistically significant. So, while we can't claim that AI and LLM findings get resolved faster than those from cloud pentests, they definitely have a faster MTTR than internal network issues.

That statement may be a bit of a record scratch moment. "Wait a second—didn't you just tell me that AI and LLM pentests show the lowest rate of resolved findings?" Your memory is not failing you; that's correct.

FIGURE 25

Median time to resolve serious findings by pentest type

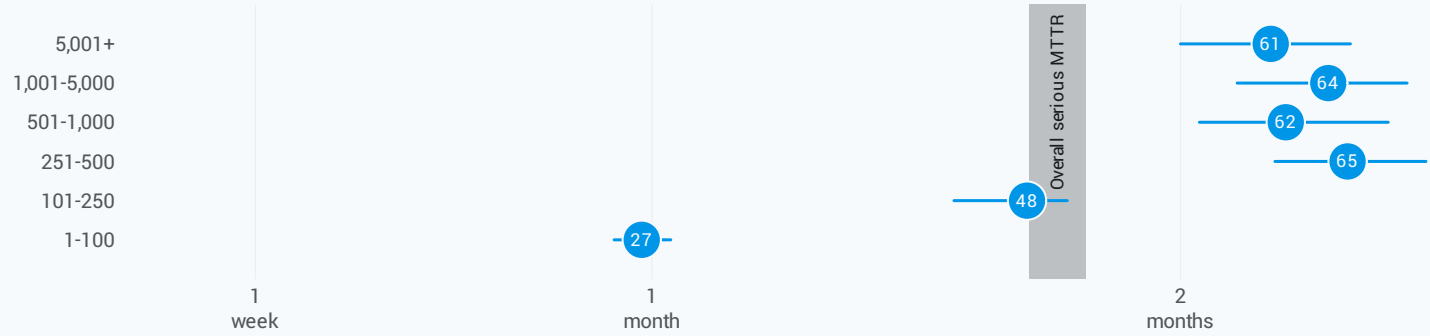


Here's our quick take on what's going on here. Keep in mind that MTTR only applies to what's fixed. So, the 21% of AI and LLM findings that are actually resolved (Figure 19) are at least fixed quickly. Since we learned that criticality and effort are big drivers of what gets resolved, we can infer that organizations quickly knock out the minority of serious AI and LLM findings that are easy and leave the rest. It's also likely that, in the case of SaaS applications, the vendor patches the vulnerabilities soon after disclosure.

Figure 26 compares resolution timelines for serious findings among organizations of different sizes. The overlapping confidence intervals of the four larger tiers indicate statistically similar resolution times among them. But the two smallest sizes boast significantly faster MTTRs. Thus, not only do larger organizations resolve a lower proportion of findings than their small counterparts (Figure 21), but they also take twice as long to do it.

FIGURE 26

Median time to resolve serious findings by organization size



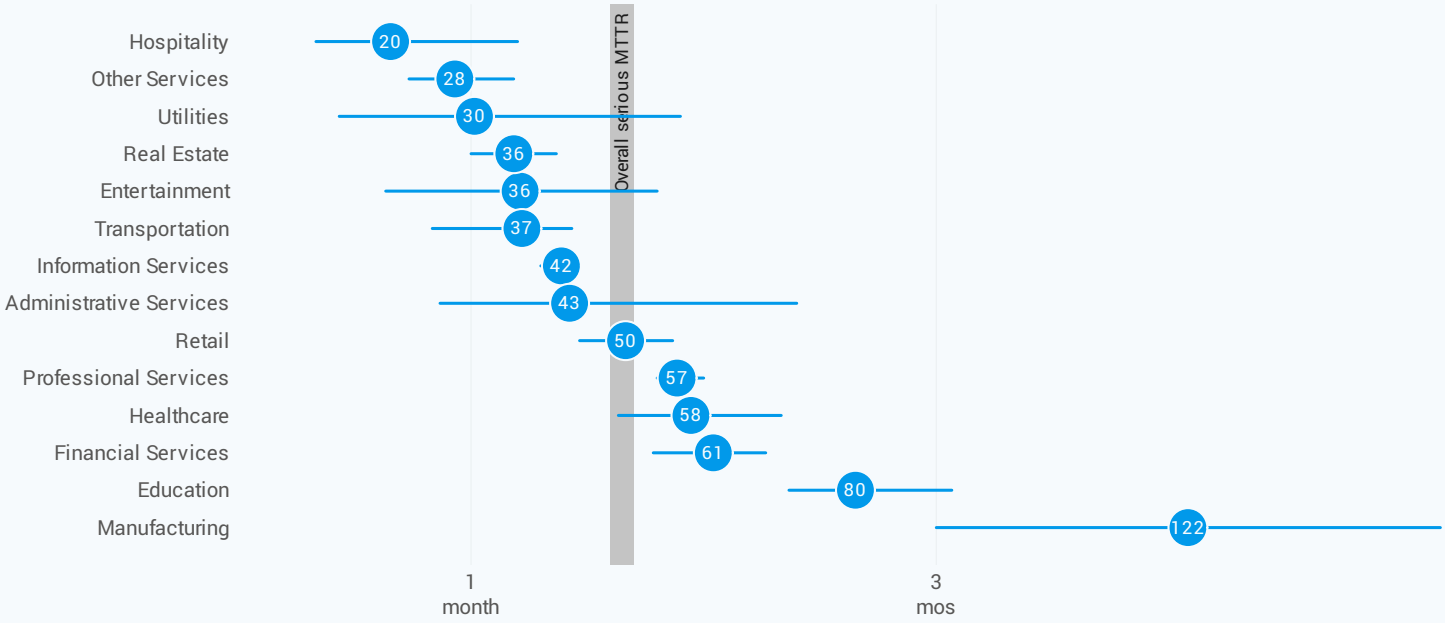
We'll echo what was said earlier about larger organizations contending with more complex infrastructure and processes. Sure, they have more resources to throw at fixing problems, but their problems tend to be bigger and interdependent with many others across the organization. Big corporations aren't known for moving fast, and that premise appears to carry over to resolving pentest findings.

Sector-specific MTTRs in Figure 27 range from 20 days (Hospitality) to 122 days (Manufacturing). Manufacturing and Education stand out from the pack with unusually long MTTRs. The challenges of managing IT in educational institutions are well documented. For manufacturers, part of the challenge stems from reliance on legacy systems that often require the device vendor to build and apply the patches. Plus, resolving some issues requires stopping production, which carries a higher business impact than system shutdowns in other sectors.

Our survey results corroborate these challenges for manufacturers. Representatives from that sector report longer SLAs for resolving pentest findings. They also put a higher priority on fixing the easy issues compared to other sectors.

FIGURE 27

Median time to resolve serious findings by industry



The hospitality sector's pole position in the resolution race is curious in light of its next-to-last-place finish for the overall proportion of findings closed (Figure 20). Is it really "winning" if you're super fast at resolving a minority of pentest findings but leave the majority open for exploitation?

MTTR can be misleading, it's inherently overly optimistic. It measures how long it takes organizations to resolve findings once they take action to do so. That's not the same as measuring the true progress toward remediating all findings. For that, we turn to measuring the half-life of findings using survival analysis in the next section.

Half-life of findings

Survival analysis is a statistical technique that measures how long it takes for an event to occur (e.g., decay of radioactive nuclides or failure of equipment). Specific to our use case, that event is the closure (death) of pentest findings. Survival analysis offers a more realistic view of remediation because, unlike MTTR, it accounts for both fixed and unfixed issues throughout their life cycle. Figure 28 helps demonstrate the concept.

MTTR measures speed but not progress. Half-life measures progress toward resolving all relevant pentest findings, making it the ultimate measure of remediation performance.

The survival clock starts once issues are identified (Day 0) and keeps on ticking until the last finding is fixed (which never happens). After one month, approximately 85% of findings remain alive (unfixed). The survival rate drops to about 60% at the one-year mark. And even after five years, 45% of issues haven't yet been fixed. That's a potent dose of reality, huh?

The median survival time, also known as the half-life, is the time it takes to resolve 50% of pentest findings. Figure 28 pegs that at 3.2 years across all criticality levels.

FIGURE 28

Survival analysis showing overall percentage of open findings over time

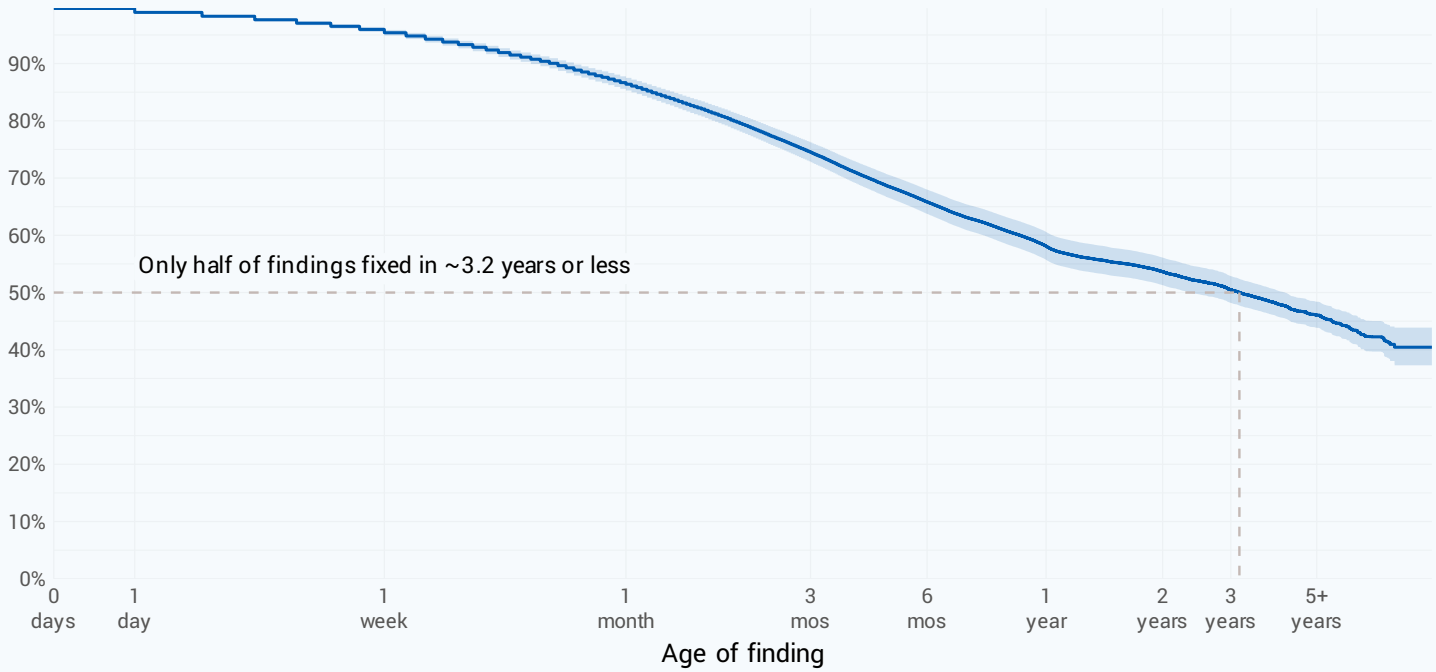
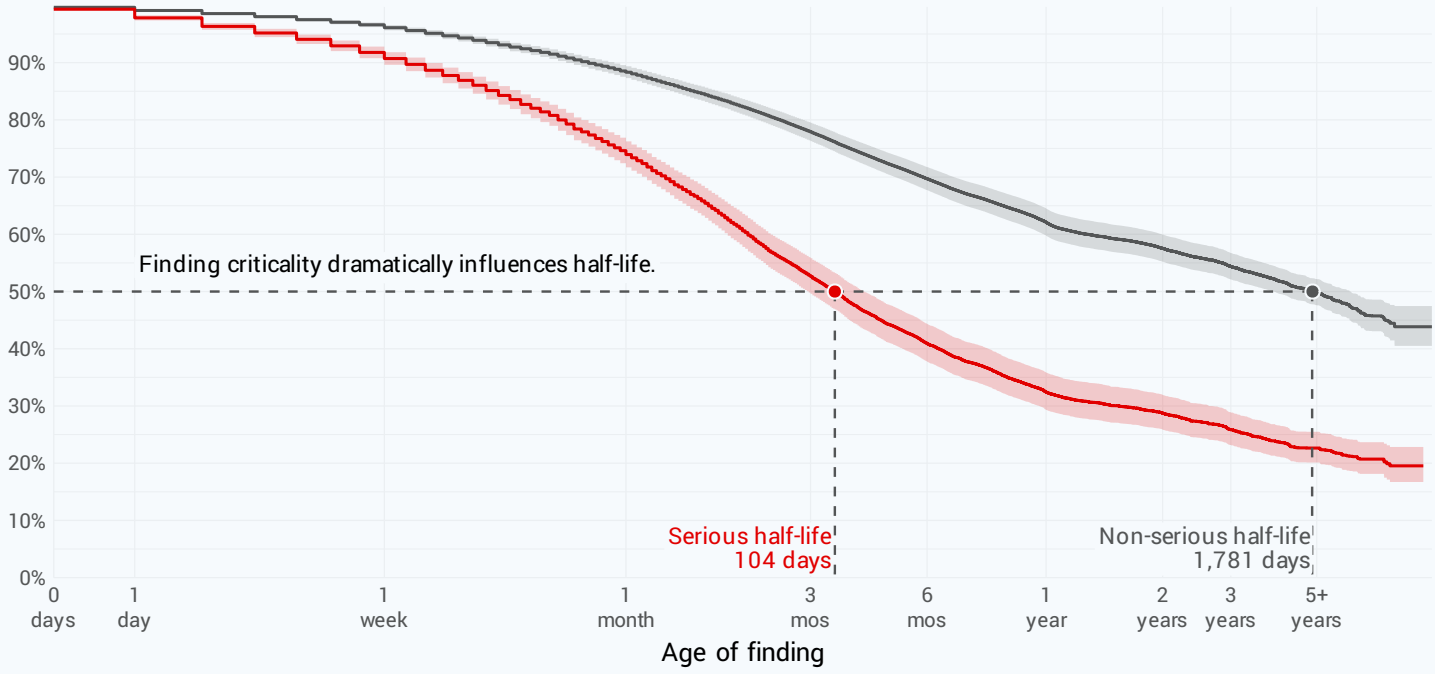


Figure 29 shows that serious issues have a much shorter half-life of 104 days, and half of less-critical issues are still around after five years.

Those keeping score may note that the half-life for serious findings is 43% longer than the MTTR (Figure 23). And it's not even close to the SLAs indicated back in Figure 22. Perception versus reality once again.

FIGURE 29

Survival analysis comparing percentage of serious and non-serious open findings over time



All together now

Speaking of keeping score, let's close out this section with a recap of key remediation stats. Figure 30 comes from a breed of charts we call "subway plots," and it's not hard to see why. They're useful for concise (albeit complex) rankings across multiple dimensions.

In this case, we're comparing the following for each sector:

Percentage of pentest findings that are serious (lower percentage = "better")

Percentage of serious pentest findings resolved (higher percentage = "better")

MTTR for serious pentest findings (fewer days = "better")

Half-life for serious pentest findings (fewer days = "better")

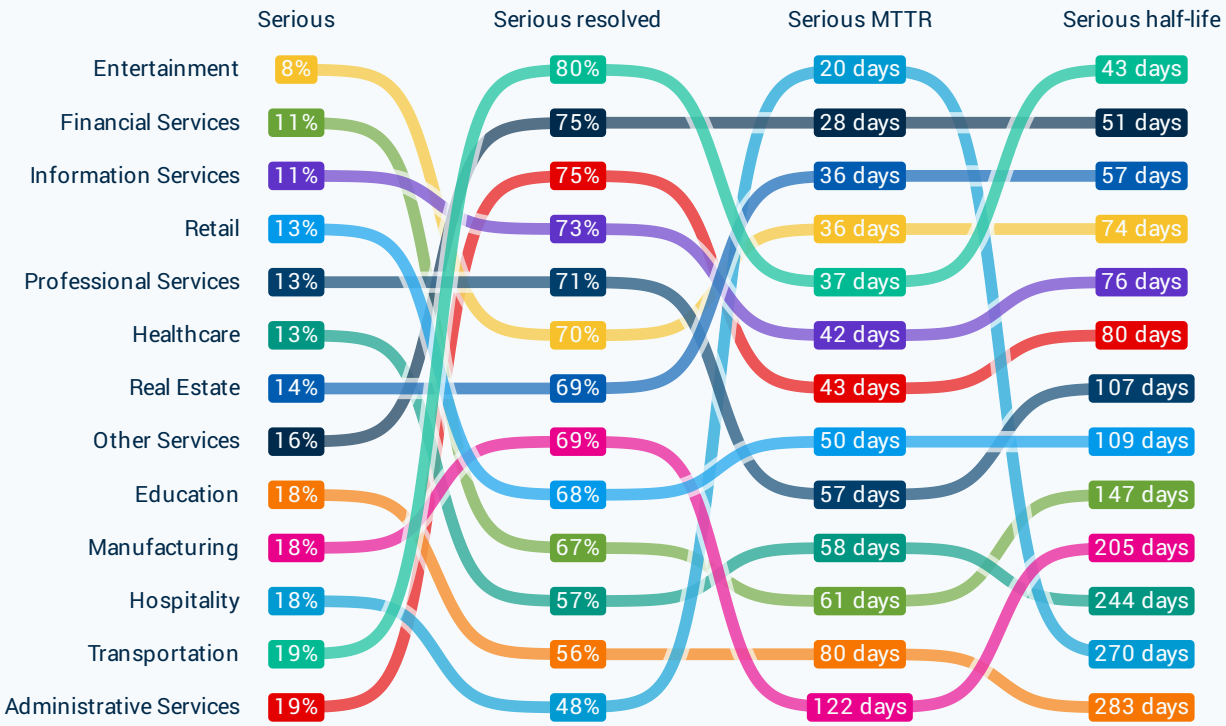
Jump on a line and ride it from left to right. For example, the Hospitality sector has the third-highest rate of serious findings and ranks last for resolving them. Hospitality's MTTR is outstanding, but its dead-last half-life reflects all those unresolved findings. Financial Services starts out with the advantage of a low prevalence of serious findings but falls to the bottom tier in the other three columns that measure the scope and speed of resolution. We'll leave you to navigate your own path through the subway and come up with ideas about what these findings suggest about sector-specific strengths and challenges.

The data and analysis in this report raise more questions than we can conclusively answer and leave room for future investigation.

At the end of the day, our study is but one contribution to the important conversation about how to solve the weighty challenges of securing our digital infrastructure. We invite you to join us as we keep pushing forward in a job that's never finished.

FIGURE 30

Comparison of key pentest remediation metrics by industry



RECOMMENDATIONS FROM THIS ANALYSIS

By implementing these best practices, organizations can create a structured, collaborative, and data-driven approach to offensive security and pentesting that enhances their overall security posture.

BUILD A PROGRAMMATIC APPROACH TO PENTESTING AND OFFENSIVE SECURITY.

Although 94% of security professionals agree that pentesting is foundational to their programs, those numbers paper over the large number of organizations with an ad hoc approach to security testing. Adopt a more strategic approach that goes beyond irregular pentesting and develop a structured offensive security program. Organizations should start by prioritizing applications based on risk exposure, focusing on the most critical assets first. Then, move beyond your web-facing applications to add testing of your network and cloud infrastructure. As you move quickly to adopt AI in your organization and products, don't ignore the associated risks—work with pentesters experienced in LLM testing.

ESTABLISH PROCESSES FOR RESOLVING PENTEST FINDINGS

Organizations are resolving less than 70% of serious findings, highlighting a big disconnect between testing and a process for addressing security issues. Establish an annual pentest calendar and align security testing with product development roadmaps to ensure continuous testing as part of your workflows. Tracking security findings in a centralized system is essential for visibility and smooth processes for remediation. Mature security teams should consider advanced testing techniques like secure code review (SCR) and digital risk assessments (DRA), followed by red teaming and threat modeling, to identify and mitigate emerging threats.

FOSTER COLLABORATION AND ALIGNMENT BETWEEN SECURITY AND CROSS-FUNCTIONAL TEAMS

Security and development teams must work together to prioritize remediation efforts. For organizations struggling with developer engagement, align security goals with engineering metrics to foster collaboration. Escalate security concerns to leadership, highlighting past security incidents by partnering with incident response to discover which vulnerabilities led to breaches.

Building strong cross-functional relationships is key—security teams should proactively engage with development peers to understand their challenges and priorities. Regular meetings, shared reporting, and public recognition of security champions can foster a positive security culture. Establishing documented processes, defining ownership of security fixes, and agreeing on SLAs for remediation ensures accountability and continuous improvement.

Schedule a meeting with a pentesting expert

RESEARCH METHODOLOGY

This report analyzes two different datasets. The majority of analysis is based on data collected during Cobalt pentests. This is supplemented by insights collected via a survey by a third-party research firm, Emerald Research. These are further described below.
Pentesting data

All penetration testing data analyzed in this report was collected by Cobalt pentesters. This spans over 16,000 pentests conducted on more than 2,700 organizations over a 10-year period. Metadata from these pentests was exported from Cobalt's platform, sanitized to remove client-identifying and other sensitive details, and provided to Cyentia Institute for independent analysis. All statistics and charts featured in this report were produced and validated by Cyentia.

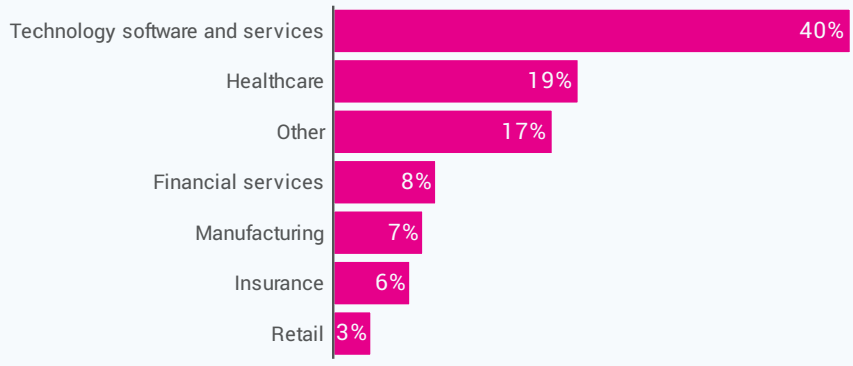
Cobalt pentesters follow specific methodologies depending on the type of test they're performing. By default, that includes industry-standard vulnerabilities from the Open Web Application Security Project (OWASP), which includes different top 10 lists for web, API, mobile, AI/LLM, and cloud systems. The Open Source Security Testing Methodology Manual (OSSTMM) is used for pentests of internal and external networks. More details on each of our pentesting methodologies can be found on the Cobalt website.

Survey sample

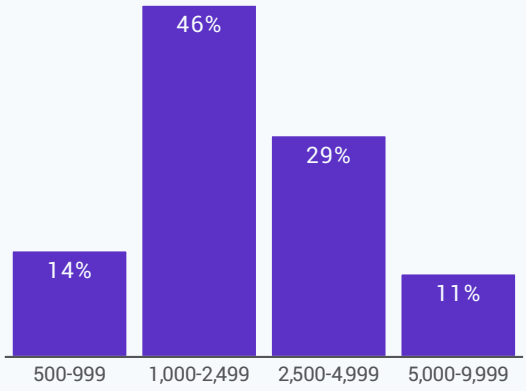
Cobalt contracted Emerald Research to administer a double-blind survey. About 1,600 people responded to the invitation, but over 70% were disqualified based on screener questions and various quality checks during and after the survey.

In the end, a total of 450 validated responses were collected. The sample consists of full-time information security leaders (50%) and practitioners (50%). These participants represent a range of industries and organization sizes.

Respondent industry



Organization size



About Cobalt and Cyentia

Cobalt is the pioneer in pentesting as a service (PTaaS) and a leader in offensive security services.

We are focused on combining talent and technology with speed, scalability, and expertise. Thousands of customers and hundreds of partners rely on the Cobalt Offensive Security Platform, along with the industry's largest exclusive community of 450+ trusted pentesters and security experts, to find and fix vulnerabilities across their environments. By enabling faster pentest launches, real-time collaboration with testers, continuous scanning, and seamless integration with remediation workflows, we help organizations identify critical issues and accelerate risk mitigation so they can operate fearlessly and innovate securely.

The Cyentia Institute is a widely respected research and data science firm; our experts are advancing the cyber security industry in knowledge and practice. We accomplish that goal by collaborating with security companies to publish data-driven reports on a range of topics and through analytic services that help organizations manage cyber risk.