

sysdig

2024 Cloud-Native Security and Usage Report

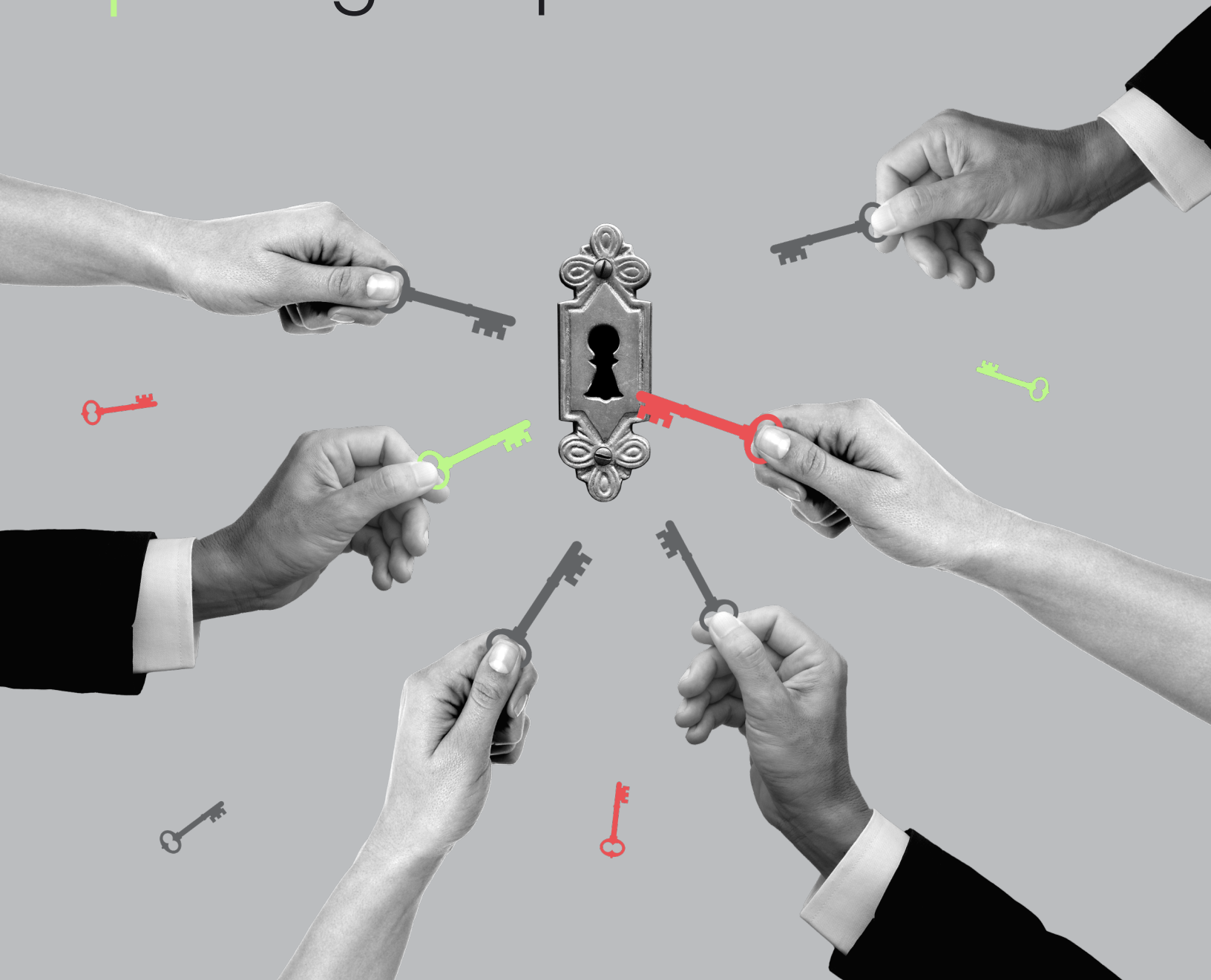


Table of Contents

Key Trends	03
Executive Summary	04
Vulnerability Management is a Priority	05
Fortifying Threat Detections	08
Neglected Foundations: The Overlooked Risks of Identity Management	11
Secure Delivery and Developer Habits	14
AI Adoption is Growing, But Not in the Ways You Might Think	18
Methodology	20
Conclusion	20

Key Trends

Identity management is the most overlooked cloud attack risk

98% of permissions are going unused and only 20% of CNAPP users are prioritizing CIEM.



Short-lived containers aren't stopping attackers

70% of containers live five minutes or less, but cloud attacks take only 10 minutes and leverage automation to work quickly.

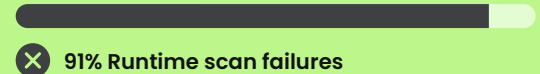


01

Shift-left is still a goal, not a reality

Runtime scan failures are at 91%, superseding CI/CD pipeline scan failures, but runtime prioritization has reduced critical and high vulnerabilities in use by nearly 50%.

02

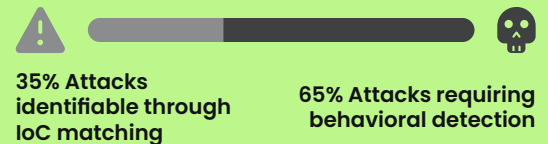


03

Threat detection programs are maturing

35% of cloud attacks are identifiable by IoCs, but 65% require additional nuanced behavioral detection and response mechanisms.

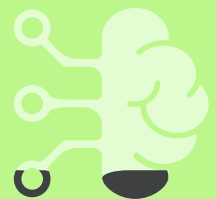
04



05

Enterprise GenAI adoption is growing slower than expected

31% of cloud users have integrated a variety of AI frameworks and packages, but only 15% of these integrations are generative AI.



Executive Summary

The Sysdig 2024 Cloud-Native Security and Usage Report comes at an exciting time after a year of cybersecurity making headlines worldwide. This is indicative of how broad the security landscape has grown in a short amount of time, thanks to the cloud.

As we have done in the past, this report looks at real-world data to draw conclusions about the state of cloud security. From our perspective, we see that organizations continue to struggle with the shift-left concept. Although runtime threat prioritization has greatly reduced vulnerabilities, there remains an urgency for powerful and speedy cloud threat detection and response (TDR).

Likewise, identity management remains an opportunity for maturity, as excessive permissions – also noted in last year's report – continue to be granted for both human and machine identities. As we reflect on the impactful and well-known identity attacks targeting companies, from casinos to consumer goods, it's clear that excessive permissions reveal a risky landscape demanding attention and action.

Short-lived workloads are no match for the speed of cloud attacks, and while many organizations are automating data-processing efforts, few have embraced applying generative AI (GenAI) to security practices.

In addition to highlighting a reduction in runtime vulnerabilities and the extent to which security teams currently accept risks associated with overly permissive privileges, this year's report also explores resource consumption trends, wherein unused and unrestricted resources increase both the chance of an attack and the likelihood of an attacker's success. The trends related to resource consumption further illustrate the significant financial and material impacts threat actors are likely to realize when they take advantage of these access control gaps.

Automation, DevOps cycles, and security tooling are all changing quickly. The ever-dynamic landscape inherent to the cloud poses a challenge for security leaders and practitioners alike – where every innovation is met with an exploit. **In the great move to the cloud, attacks can happen in 10 minutes, and speed is of the essence.** As an industry, we are innovating and maturing at the necessary speed of the cloud; in security, we are racing attackers at a breakneck pace, and every second counts.

This year's trends show that organizations are still choosing speed and convenience over security best practices in favor of more rapid development and innovation.


Vulnerability Management is a Priority

Organizations are seeking the most productive and time-efficient ways to minimize vulnerabilities and reduce their attack surface. 88% of Sysdig customers engage with vulnerability data weekly and, as a result, are successfully reducing vulnerabilities in use at runtime.

Shift-left is still a goal, but not yet fully realized

Vulnerability management is fundamental to organizational security programs and risk prioritization. The goal is to identify known vulnerabilities in workloads before exploitation and remediate them by patching code, updating dependencies, or mitigating the risk with some other security control. Ideally, vulnerabilities are identified in pre-delivery pipeline scanning, often labeled as a shift-left approach.

“ We’ve automated our manual reviews and now execute container vulnerability and compliance checks on containers as they’re promoted into our production environment. Those automated checks allow us to move faster.

SAP Concur  Director of Engineering

Scanning early – before production delivery – reduces opportunities for attackers but can lead to higher false-positive rates. Many organizations still approach the problem by scanning production environments continuously, referred to as a runtime security or a shield-right approach. Scanning in runtime, as part of a complete system, provides improved accuracy over shifting left, but the reality of having an exploitable issue in production is inescapable. The highest level of a mature security program will use both approaches to reduce false positives and attack surfaces.

During vulnerability data analysis, we looked at nearly 6 million runtime image scans and over 500,000 continuous integration and continuous delivery (CI/CD) build pipeline scans to review the policy failure rates. Runtime scans had a 91% vulnerability policy failure rate and, surprisingly, CI/CD build pipelines had a lower failure rate of 71%.

In following the shift-left mantra, we would expect these numbers to be flipped. Organizations should be scanning early and often, recognizing the failed builds, correcting the code, and then redeploying. **With this approach, a high runtime failure rate is unexpected, because issues should be caught before delivery and before they become exploitable conditions for attackers.**

One possible explanation for this data is that additional dependencies are being referenced that aren’t in scope for pipeline scans. Another reason may be that organizations are simply forgoing pipeline scans in favor of runtime checks for better accuracy, or to reduce the burden on development teams. Finally, not all packages are being checked all the time, which is often the case with middleware components such as NGINX, load balancers, and proxies, since the source may be considered vetted and assumed to be reasonably secure.

In-use vulnerabilities are being vanquished

Many organizations prioritize the remediation of critical and high vulnerabilities according to the Common Vulnerability Scoring System (CVSS), but this only narrows the list from hundreds of thousands of vulnerabilities to tens of thousands, most of which don't pose a real risk to the business. Cloud security programs need more effective ways to narrow down the list.

Last year, we reported on using in-use vulnerability exposure as a way to prioritize risk. Filtering down to vulnerabilities that are exploitable and where the code is in use by the application makes to-do lists both manageable and actionable. We are excited to report that workloads with in-use packages containing fixable critical or high vulnerabilities have been reduced by nearly half, from 15% to 8.2% over the last year. This indicates that technical teams are both able and willing to rapidly pay down their high-risk vulnerability debt when presented with actionable, well-scoped remediation priorities.

There are other areas where we also saw improvement in vulnerability management. Workloads running code that contains fixable critical or high vulnerabilities with a known exploit, dropped from 2% to 1.2%. Running workloads with critical or high vulnerabilities but with no fix available dropped from 1% to 0.5%. We hope to see these trends continue – because running vulnerable images is still a massive security risk – but prioritization of those that matter the most to your organization and environment will undoubtedly reduce your attack risk.

Out of 100 workloads with critical or high vulnerabilities



**Critical and high
vulnerabilities in use down by**

half

over the past year!

This indicates that technical teams are both able and willing to rapidly pay down their high-risk vulnerability debt when presented with actionable, well-scoped remediation priorities.

Fortifying Threat Detections

Advanced cloud threat detection strategies

Cloud security is clearly maturing beyond prevention, driving more urgency around the need for cloud detection and response. To that end, nearly 90% of Sysdig customers leverage TDR insights weekly. Comprehensive threat detection requires multiple approaches, including threat intelligence, indicators of compromise (IoCs), and behavioral detections. In our data set, 35% of attacks were identifiable through IoC matching, while the remaining 65% required additional nuanced behavioral detection mechanisms. This shows that although threat intelligence feeds are incredibly useful, they don't come close to providing full detection coverage.


35% Attacks identifiable through IoC matching



65% Attacks requiring behavioral detection

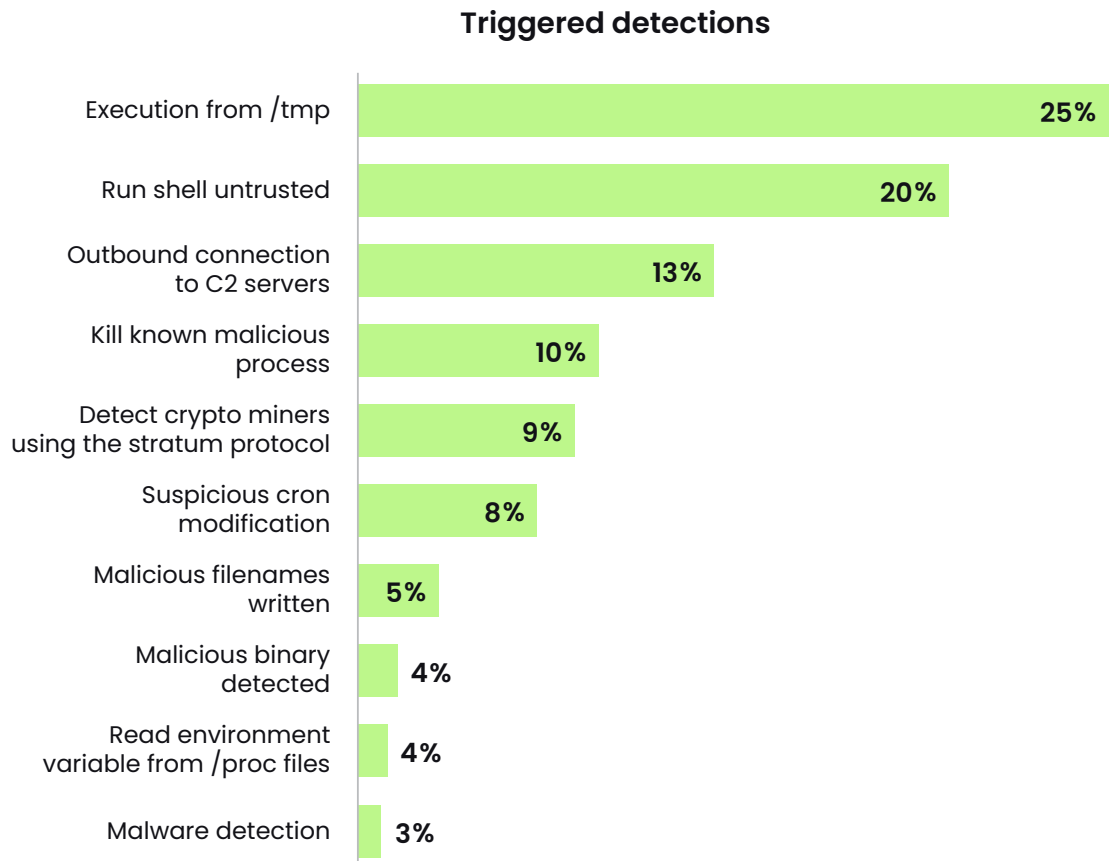


“ It's too tedious to go through logs and alerts manually to meet the standards we've established. It would require a full-time person and is just not a necessary or good use of time. We've automated the entire process and have reached a more accurate level of insight faster.

 BEEKEEPER Security Architect

With the high requirement for behavioral detections to catch unknown threats, detection engineering as a practice is becoming more commonplace within cloud security operations centers (SOCs). The continuous creation and testing of custom threat detections is a practice shared by nearly 65% of TDR users and is a positive indicator of a proactive and maturing threat detection program.

Last year, the most commonly triggered MITRE ATT&CK tactics were defense evasion and privilege escalation. These tactics are heavily used by threat actors, leading to a large number of techniques and rules dedicated to these kinds of attacker actions. This year, the most commonly triggered detections fall under initial access and execution tactics.



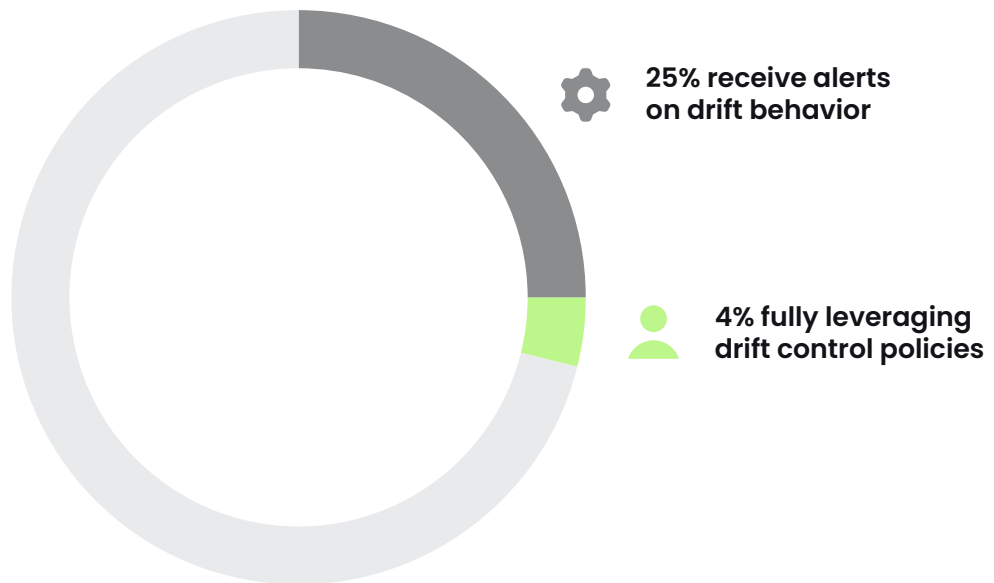
We still believe that security scans and testing cause a majority of these triggers, which indicates that these organizations are focused on detecting tactics that present themselves earlier in an attack chain in the hopes that they can respond before it's too late.

Getting ahead of drift in container security

Drift control can be a powerful security tool for teams that have embraced the distributed, immutable, ephemeral (DIE) operating philosophy. DIE is the notion that an immutable workload should not change during runtime; therefore, any observed change is potentially evident of malicious activity. It is possible to configure drift policies to either prevent all such activity completely or simply alert upon detection. In essence, the implementation of drift control exemplifies how more robust detection can support greater prevention – detecting and blocking container drift is a powerful preventive measure for container security.

Approximately 25% of cloud users receive alerts on drift behavior. On the other hand, about 4% of teams are fully leveraging drift control policies by automatically blocking unexpected executions. While these numbers may seem low, it is important to understand that drift control significantly reduces alert fatigue and attack risk when the organization has already embraced DIE.

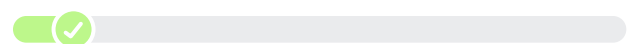
Drift control users



However, turning on drift control detection for nonimmutable workloads can generate false positives, whereas drift control prevention could impact availability and cause serious downtime. In cases where an engineer directly modifies a workload or makes a change outside of the established version control mechanism, the system may generate tens of thousands or even millions of drift detection alerts daily. Conversely, if individuals stick to their organization's formalized release processes and avoid changes to running workloads, the accuracy of the drift alerts will significantly improve. Activating preventive drift control measures in a mature developer environment will reduce the amount of potentially malicious events requiring incident response intervention by approximately 9%.

These lower numbers speak broadly to the state of security maturity regarding continuous delivery and infrastructure automation practices, and it appears that there is still a long road ahead for some organizations.

Preventive drift control reduces response efforts by 9%



Neglected Foundations: The Overlooked Risks of Identity Management

While organizations are making strides in reducing vulnerabilities and prioritizing threat detection efforts, identity management appears to have fallen by the wayside. Our data shows that permissions are still being granted in excess and not being properly maintained, leaving plenty of misconfiguration and privilege escalation opportunities for attackers. While detection will mitigate some of this risk, an organization can really improve its overall security posture and reduce its attack surface through the improvement of identity management efforts, specifically remediating excessive permissions.

Identity who?

Our data showed that, unlike the high priority given to threat detection and vulnerability management, **only 20% of cloud-native application protection platform (CNAPP) users are putting effort into reviewing and managing identities weekly.** This could be because this data is managed on a much less frequent basis, or because organizations are using other tools to see their data. Regardless, they are certainly not taking action to remediate excessive permissions.

Similar to shift-left approaches, enforcing least privilege is also still touted as a high priority for the security of any organization, and also underpins zero-trust design. Last year, we reported that permissions were being granted in excess and remained unused. In our analysis of this year's data, we found that the struggle with assigning and managing adequate permissions is only getting worse. We attribute this to a lack of regular and continuous identity management, and the fact that granting elevated permissions to identities is a convenience that saves a notable amount of time. Without permission restrictions, employees can work seamlessly without the impediment of having to request additional privileges during their projects, but this convenience and time savings of excessive permissions come with increased risk.

“ If you abide by the principle of least privilege, eliminating excessive permissions is a key priority. It's critical for us to understand where we have overly permissive identities and, due to the scale, we need an automated way to manage them.

Booking.com Senior Product Manager

Excessive permissions

Both human and machine identities use only 2% of the permissions they are granted.

Excessive permissions and administrative privileges are granted with every initiation of a tool or application as the default setup, and these are rarely modified. In some cases, nonhuman applications, tools, and services were granted access to tens of thousands of permissions upon initial implementation but were never disabled or deprovisioned. We discovered machine identities with hundreds of thousands of unnecessary permissions that went completely untouched for more than a year.

Excessive permissions

✘ **98% Unused permissions**



2% Granted permissions used ✔

The excessive scope of human identities is less surprising because overpermissioning has always been the easiest way to get employees working quickly and flawlessly. Machine identities, on the other hand, have no excuse for such treatment, because they should be created with a specific scope in mind and, unlike human users who may move between projects and roles, nonhuman identities should not require frequent changes in scope.

This practice creates undue risk when a majority of severe cloud security incidents with material impact are tied to the failed management of identities, access, and privileges. We've seen attackers exploit weak access controls and mispermissions countless times over the last few years. It's often the initial attack vector in an attack chain, and this identity compromise inevitably leads to application abuse, system compromise, or data exfiltration. If a security incident significantly affects an organization's financials or could cause concern among investors, there are now additional regulatory requirements to meet for materiality assessments and disclosures. Organizations already struggle to satisfy existing privacy and data security regulations, evident in areas such as protected health information with Health Insurance Portability and Accountability Act of 1996 (HIPAA) [violation trends](#). Gaps in access controls amplify the tsunami of potential impacts to an organization.

Machines remain on top

Nonhuman identities makeup 63% of cloud users and roles this year. After gathering data on the numbers of human and machine identities for three years, we are confident that machine identities will likely remain the majority identity type as organizations grow, scale, and automate their cloud services and tools. However, to reduce the vast attack surface that machine identities inadvertently create, the default permissions granted upon the creation of a new nonhuman user must first be reviewed following initial provisioning – and then continuously afterwards.

Cloud users and roles



Organizations quickly find themselves wrestling with offshoots of identity and access management (IAM) tooling that include cloud infrastructure entitlements management (CIEM), privileged access management (PAM), secrets management, and identity governance and administration (IGA). CIEM was specifically designed to address the complexity that arises with mixed-identity types in cloud environments, particularly multicloud approaches, where an identity fabric is normal. According to Gartner®, “By 2027, identity fabric immunity principles will prevent 85% of new attacks and thereby reduce the financial impact of breaches by 80%.”¹ Human-centric IAM approaches do not suffice for machine identities given the mixture of identity types, differing usage patterns, and the need for dynamic access control with modern technology stacks. Identity technologies must work in concert to power an identity threat detection and response (ITDR) strategy, and CIEM is an essential capability in addressing identity risks in the cloud.

In November 2023, an Iranian threat actor breached multiple U.S. organizations, including water authorities, because the default password on an industrial control device was never changed.

¹ Gartner, Invest Implications: Top Trends in Cybersecurity 2023, Frank Marsala, 23 March 2023. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

Secure Delivery and Developer Habits

Balancing the many types of risk is a tightrope act

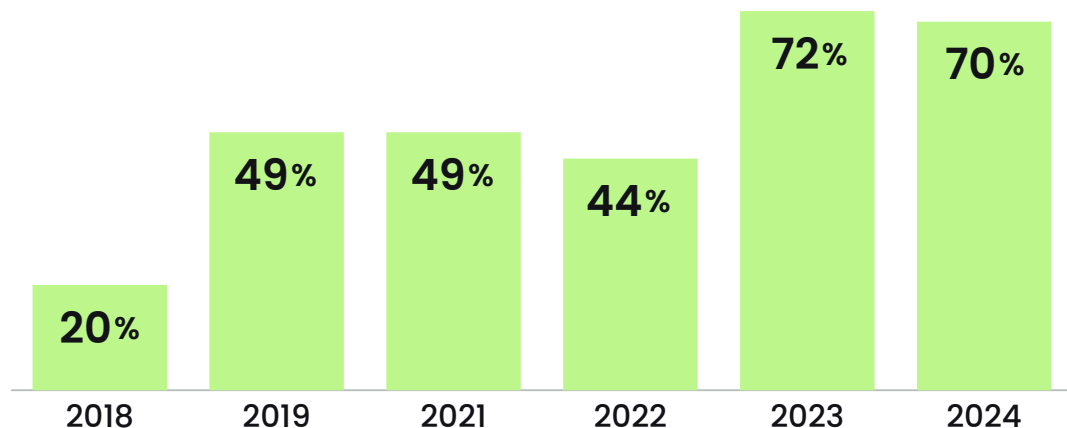
From the usage data we gathered this year and in previous years, there is one important fact to note that has been consistent: convenience is king. A majority of our customers use publicly available registries for images and do not limit or restrict their CPU and memory usage. The pace of development, code commits, and builds within open source projects is quick, and a lack of restrictions allows developers to work at the speed of the cloud. However, this creates governance challenges and security risks for organizations and elevates threats to operational resilience, a necessary ingredient for cybersecurity.

Attackers have container lifespans beat

Year-over-year we have seen the average lifespan of a container get shorter. This year, 70% of containers are short-lived and spun down in five minutes or less. The Sysdig Threat Research Team (TRT) reported in the [2023 Global Cloud Threat Report](#) that a cloud attack takes only 10 minutes. If an attacker has not moved laterally, they are booted once the container has been executed and killed. However, we know that it is fast and easy for an attacker to enter and move through an environment because, in the cloud, attackers automate their discovery and reconnaissance efforts. Almost in an instant, once an attacker is in an environment, they have the lay of the land and are ready to press forward. This highlights the importance of real-time security and continuous scanning. Running vulnerable workloads, no matter how short-lived, leaves an organization at risk for an attack.



Containers living less than 5 minutes



Consuming public sources is still the norm, despite best practices

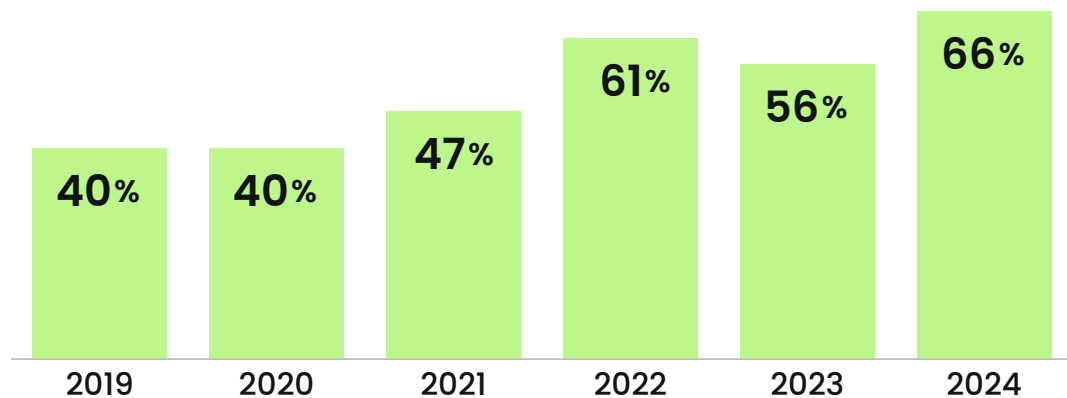
Year over year, we review the breakdown of public and private registry and repository use to discern from where images are most commonly pulled. This year, we analyzed over 1,400 uniquely named registries spanning more than 2.6 million containers. We found that 66% of all registries used to host and manage container images are public. The other 34% of registries are a combination of private instances of public registries, vendor-managed registries, or unique and customized registries — generally for our larger-sized and/or more security-savvy customers.

Registry type



It's safe to say that after reviewing this data for several years, the convenience of public and managed registries is appealing to the majority of container users, regardless of their security maturity. Using public registries for images allows organizations to save time creating and maintaining their own registries and save money by not having to pay for vendor-managed registries. However, it is important to note that public registries come with reduced security control enforcement and create software supply chain risk.

Public registry use



During public image analysis for the 2023 Global Cloud Threat Report, the Sysdig TRT identified 13,000 suspicious container images on DockerHub with over 800,000 combined downloads. After conducting runtime analysis, roughly 6% of these images were deemed truly malicious. The security risks of pulling from public registries may seem trivial given this small percentage of malicious images, but it's still a significant risk related to container image hygiene. Compromised images can lead to big security problems for organizations if they are run without inspections or protections.



The ideal security approach is to use either vendor-managed registries or private instances of publicly available registries that teams can pull code or instantiate from. However, since these private registries contain copies of code pulled from public registries, they should also be stringently scanned based on organizational policies before they are implemented in an environment. Mature organizations store “golden images” or “golden source” within private registries, which also becomes a backbone of other practices, including GitOps and PlatformOps. This approach enables organizations to establish stronger governance over source packages and images that in turn help mitigate threats that arise when attackers target files in public repositories and registries.

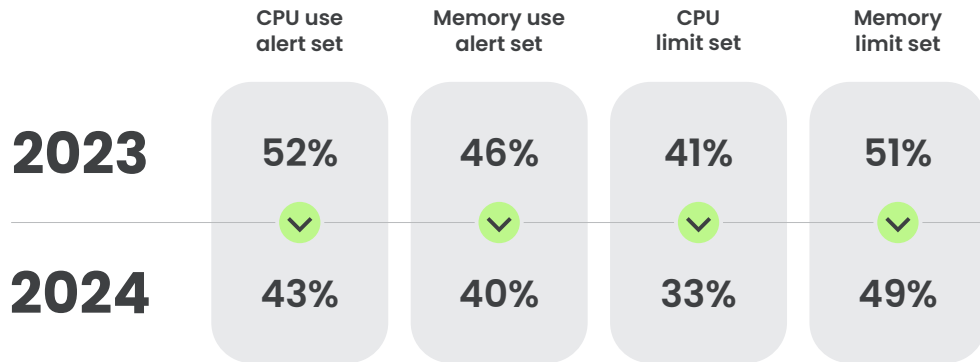
It stands to reason that some organizations are accepting the relative risk of pulling from public registries and perhaps mitigating elsewhere as part of their cybersecurity program, such as relying on runtime TDR. This may be for convenience, to reduce operational burden, as a byproduct of organizational silos, or as part of cost-cutting efforts. Regardless, it's in opposition to traditional defense-in-depth security thinking and raises concerns about secure delivery processes. This is not a security best practice, but with powerful TDR in place, the associated risk of public registries is being mitigated in some fashion.

Organizations are shielding right to mitigate the risk of using public registries with TDR, but a combined approach is ideal.

Resource constraints are still too lax

Less than 50% of environments have alerts set to trigger on CPU and memory use. Furthermore, a majority of users do not have maximum limits set on their CPU or memory use. It's likely perceived that alerts are too noisy, and organizations prefer the added capacity so as not to impact production applications. This is another case of trading off security risk for availability risk. Without setting alerts or limits, there is an increased risk of having to pay for resources used by attackers in your environment. The decision to prefer availability is about convenience and supporting the speed of development and elasticity. Reducing or disregarding resource limitations is a time-saving gamble.

Fewer organizations are concerned with CPU and memory use



The lack of resource constraints is an attack risk factor, and setting limits is considered a fundamental security best practice. Unlimited resources are a prime opportunity for attackers to latch on and take advantage of your environment, such as to perpetuate cryptojacking attacks or complex attack chains that use resources to target other systems within an organization's network (that is, lateral movement). This type of threat has proven to be a costly mistake if unchecked resources aren't found and removed quickly.

This is also a type of financial risk, and there's a business opportunity to see which processes are using memory and CPU and make some reductions to save costs. Given the current macroeconomic environment, most organizations are scrutinizing capital and operational expenses. Controlling resource consumption, certainly in cloud and container environments, is a way to achieve those financial goals.

AI Adoption is Growing, But Not in the Ways You Might Think

Data processing and comprehension are faster with automation

There is an incredible number of AI packages and frameworks being implemented, but 85% of these packages are being used to empower data analysis and enhance correlation and anomaly detection through machine learning (ML) rather than GenAI. Roughly one-third of our customers have integrated AI frameworks and packages. To keep up with the speed of cloud attacks, automating some security processes is necessary. Large language models (LLMs) can generate actionable insight, help teams get a better understanding of risks and security issues, and even make security operations and response times faster.

“ Using AI to provide relevant context during an attack or for day-to-day tasks has been extremely valuable to us. We anticipate that this will help break down silos in cloud domain knowledge, uncover hidden risks, and connect dots along the attack path.

 **onna** Principal Architect

AI adoption

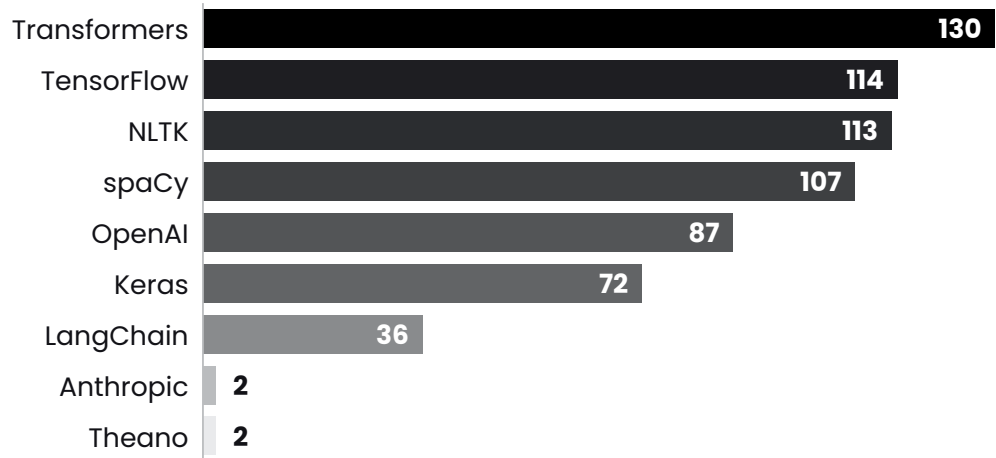


 **85% Data analysis and correlation**

15% GenAI 

The numbers we are reporting here are specifically those known packages within cloud workloads. Many organizations will not be building their own GenAI solutions at this point and instead will consume pre-built, pre-trained services such as those offered by Anthropic (Claude), OpenAI (ChatGPT), or cloud service providers. However, those organizations that are installing and using various types of AI packages in their data collection and comprehension efforts are taking advantage of the technology capability and reducing the data-crunching burden on staff.

GenAI package types



This table represents the 15% of GenAI package types seen in cloud environments

While individual use of browser-based GenAI tools and interfaces is almost certainly much higher, our data shows that enterprises appear to be slower to hosting AI explicitly in workloads running within their environments. It is less likely that organizations are using LLMs, like OpenAI's ChatGPT, for managing security practices; improving detections; or understanding and finding malicious actor tactics, techniques, and procedures (TTPs). Instead, GenAI is being used for tasks such as marketing campaign efforts, email composition, document writing, and code writing assistance. We anticipate an increased use of AI for security purposes across industries, and Gartner predicts that "by 2028, 75% of enterprise software engineers will use AI coding assistants, up from less than 10% in early 2023."²

² Gartner, Innovation Guide for AI Coding Assistants, Arun Batchu, Philip Wlash, Jim Scheibmeir, et. al 16 October 2023. Gartner is a registered trademark and servicemark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

Methodology

Sysdig is dedicated to sharing real-world customer data. Using this kind of data allows us to report on the actual changing aspects of container, cloud, and security trends, rather than opinionated or skewed survey results. Our analysis of container, cloud account, and application usage allows us to share unique insights, such as how many vulnerabilities are actually in use at runtime and exactly how compute resources are being used.

We derived the data in this report from an analysis of thousands of accounts and millions of containers that our customers run daily. Our security and container insights originate from a wide range of industries, with organizations ranging in size from tech startups to large enterprises. This anonymized customer data is hosted in regions across the globe and spans multinational organizations in North and South America, Australia, the EU, the U.K., and Asia.

Conclusion

Sysdig's 2024 Cloud-Native Security and Usage Report provides a comprehensive view of the evolving security landscape in cloud environments and organizations. It underscores the ongoing struggle with the shift-left concept despite significant advancements in runtime threat prioritization. The urgency for robust cloud threat detection and response mechanisms is evident, especially in the face of the identity management struggle.

The report sheds light on the reduction in runtime vulnerabilities but highlights a concerning acceptance of risks associated with excessive permissions granted to human and machine identities, posing a considerable threat to the security landscape. The trends in resource consumption emphasize the financial and material impacts that threat actors can exploit through access control gaps, underscoring the need for proactive measures. While organizations are moving fast in the cloud by ignoring best security practices, they are not comfortable implementing AI yet.

The key trends from this year's annual report highlight that realistic security practices put convenience before best security practices to keep pace with speedy cloud innovation, but there is a hesitancy to deploy AI packages.

We look forward to discovering and reporting on security advancements over the next year. See you then!



Check out Sysdig's past Cloud-Native
Security and Usage Reports

[READ NOW](#) →

REPORT

COPYRIGHT © 2024 SYSDIG, INC.
ALL RIGHTS RESERVED.
RP-009 REV. A 1/24
