TAG

ANALYST REPORT: UNDERSTANDING RUNTIME CLOUD WORKLOAD SECURITY WITH INLINE MITIGATION

DR. EDWARD AMOROSO
CHIEF EXECUTIVE OFFICER, TAG

ANALYST REPORT: UNDERSTANDING RUNTIME CLOUD WORKLOAD SECURITY WITH INLINE MITIGATION

DR. EDWARD AMOROSO, CHIEF EXECUTIVE OFFICER, TAG

This report explains the benefits of runtime security with inline mitigation for cloud-native workload protection. Commercial vendor AccuKnox[1] is used to demonstrate the practical deployment and use of this powerful zero trust approach to cybersecurity.

INTRODUCTION

The need to innovate in cybersecurity has never been greater, given the threats that continue to target enterprise and government teams around the world.[2] Ironically, this need exists despite on-going design and deployment of new cybersecurity products and services from startups, usually funded by venture capital teams.[3] Our observation, sadly, is that too many of these new solutions mimic other products, presumably based on revenue objectives.

From time to time, however, new commercially available cybersecurity solutions do emerge that show promise in addressing the practical threats that exist for modern organizations. Such new defensive solutions usually require advanced technology since offensive attacks have become so automated. It stands to reason that manual-oriented controls will be ineffective in reducing risk. New security solutions must be fully automated.

This report focuses on a new method known as inline mitigation which is deployed in the context of runtime cloud workload protection. We will explain how this technique addresses modern threats to multi-cloud workloads as well as how it provides an excellent foundation for runtime security. Finally, we provide a case study of commercial vendor AccuKnox and how they deploy and support a security platform that operates effectively in this space.[4]

BROAD THREATS TO MULTI-CLOUD WORKLOADS

Cybersecurity practitioners tend to rely on taxonomies to ensure completeness of coverage in their defensive activities against threats. For example, the NIST Cybersecurity Framework (NIST CSF) offers a useful multi-phase lifecycle that covers the various phases of a cyber defense. Informally, one can view the earlier phases as the left portion (hence, shift-left references) and the latter phases as the right (hence, shift-right references).

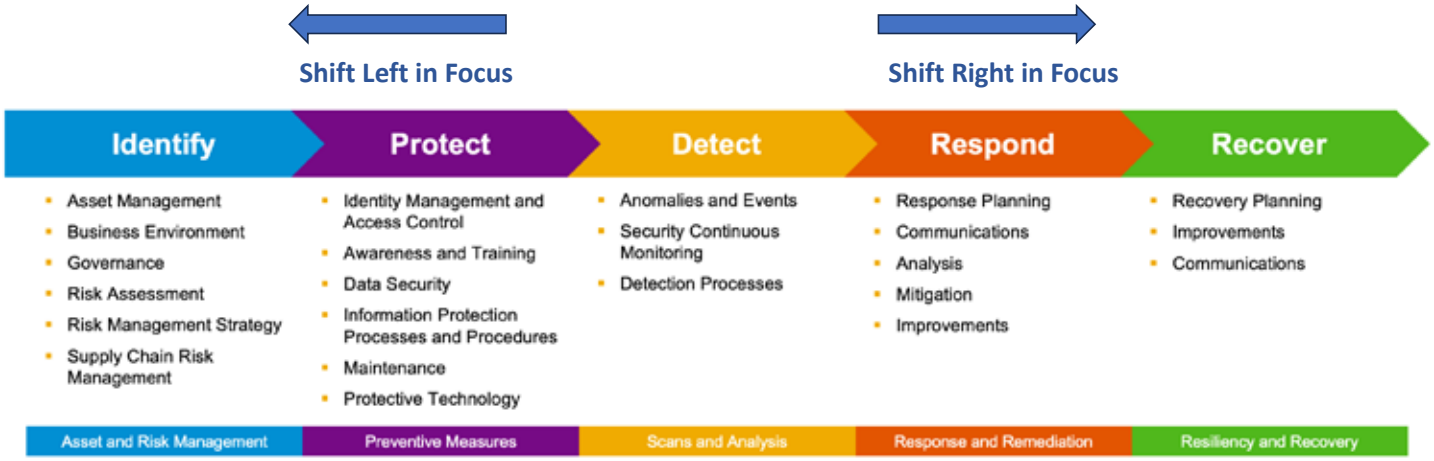

Figure 1. NIST CSF Shift-Left and Shift-Right Focus

To model cyber offense, the NIST CSF is perhaps less useful. Instead, practitioners have turned to models that explain how vulnerabilities can be exploited to create effective attack strategies. The MITRE ATT&CK Framework, for example, offers comprehensive coverage, albeit using a model that is complex and requiring of maintenance as new vulnerabilities arise.[6] Nevertheless, the idea that offense and defense use models differently should be evident.

One particularly popular and common designation is the so-called *known/unknown* delineation for vulnerabilities. That is, when vulnerabilities become known, they can be categorized into useful databases such as the Common Vulnerabilities and Exposures (CVE) program which offers security experts useful data on exploits.[7] Cyber controls are often developed to address attack entries included in the CVE database.

In contrast, when exploits are identified locally, perhaps by some hacking group or organized cyber offensive team, such as a nation-state sponsored command, then we refer to this exploit as being essentially *unknown*, as least from the perspective of everyone not included in the initial discovery. When an organization is informed of a previously unknown exploitable condition, we refer to this as a *zero-day vulnerability*.[8]

It is the *unknown zero-day vulnerability* that is particularly worrisome from a threat perspective, because it cannot be easily identified either by signature or behavioral patterns. There is some hope that artificial intelligence can be used to predict the presence of zero-day exposures, and products have been available for years that work in this manner.[9] Nevertheless, deployment of such tools has not stemmed the continued problem of zero-day vulnerabilities.

MORE DETAILED THREATS TO MULTI-CLOUD WORKLOADS

Given the concern that exists for zero-day exploits, we can now turn our attention to software organizations running their apps in cloud environments. Note that cloud environments usually involve workloads or applications being deployed into public cloud infrastructure such as from Amazon Web Service (AWS), Microsoft, or Google, but they can also consist of virtualized environments in private data centers running software from companies such as VMware.[10]

The security problem that has emerged in these environments is two-fold: First, security teams have clearly recognized that when public cloud services are used, the day-to-day control around infrastructure, services, software, and day-to-day support and maintenance is outsourced. This can be welcomed, especially for smaller teams, but there is no debate that security control for these infrastructure tasks has shifted externally.

Second, and definitely more troubling, is that just as cloud infrastructure has made it so much easier to deploy and use powerful tools for search, recommendation, delivery, and visualization, attackers have also noticed the increased power of cloud – and their associated attacks benefit from such flexibility as well. What they look for is either sloppy configuration of cloud settings or errors in the code deployed into cloud.

MITIGATING THREATS TO CLOUD

To address these threats, an entire industry has emerged roughly around static and runtime methods. The first method involves doing a static scan of cloud infrastructure in order to identify where a given workload or application – or the underlying infrastructure supporting such software, has been set up improperly. Misconfigurations are all too easy to occur, especially since cloud services include so many different set-up and delivery options.

The most popular designation in the industry to find static problem in cloud is the so-called *cloud security posture management (CSPM)* category of commercial (and open source) platform and tooling. Nearly every organization in the world now uses some form of CSPM to locate whether their cloud capabilities are vulnerable to external exploits, often based on sloppy account management.

Two vendors who have benefitted from this CSPM focus are Wiz and Palo Alto Networks.[11] Both companies have grown their CSPM solutions considerably in recent years, and it is expected that every cloud service provider will soon also increase their own focus in this area with native capabilities. TAG predicts that the cloud security vendors competing with Wiz and Palo Alto Networks will experience increased acquisition activity from the major cloud providers.

A problem with static assessment of cloud workloads is that it represents a snapshot of configuration issues – and even in the presence of a dashboard with alerting, the process still relies on human intervention to notice an issue and find ways to re-design, re-code, or reconfigure the cloud application or workload. In the most complex environments, this will be useful but insufficient to address cyber risks.

BENEFITS OF RUNTIME SECURITY

In contrast to static scans, our industry has come to recognize the benefits of so-called *runtime security*, which involves a much more dynamic assessment of how vulnerabilities might be exploited. Runtime security is certainly not a new concept, with methods such as runtime application self-protection (RASP) being common. Advanced zero-day attacks cannot be thwarted by basic signature-based defenses because they require run-time security strategies.

Two strategies exist for runtime security. First, the controls can wait and watch to see if behaviors are suggestive of some sort of problem. The entire behavioral analytics community works essentially on this notion and even artificial intelligence models train on data in this manner. Such an approach is useful, and recommended, but it does suffer from the problem of waiting for problems to occur, which can lead to consequences.

The second strategy for runtime security involves more involved, proactive inspection and control of the application and workload execution. The goal with such controls is to prevent security issues from occurring in the first place – and the advantages of such an approach should be immediately evident to any practitioner. Of course, this approach is also the most challenging, in terms of design and implementation.

HOW INLINE MITIGATION WORKS

The method discussed here involved so-called inline mitigation, which has many of the benefits discussed above for proactive runtime security with the goal of not allowing security issues in cloud-hosted workloads and applications to occur. If done properly, inline mitigation can dramatically reduce the number of problems an organization experiences with cloud-deployed systems – and this will reduce cost and improve user satisfaction.

Inline mitigation works by reviewing and scanning the application or workload to be executed as part of the runtime environment. In the context of cloud workloads, this involves dynamically assessing the parameters of a container and everything that is needed to run that container – namely, the code, runtime, system tools, and any system libraries. Such inline investigation helps to provide continuous validation of a secure environment.

Inline mitigation coordinates with the observation of application behavior, usually through integration with existing standards or open-source utilities. For example, the extended Berkeley Packet Filter (eBPF) is a popular technology that can run programs in a privileged context such as the operating system kernel.[12] It is used to extend the capabilities of the kernel safely and efficiently at runtime without the need to load kernel modules or change kernel source code.

Inline mitigation approaches will typically utilize tools such as eBPF to perform security auditing, packet processing, network insight derivation, memory profiling, and intrusion detection.[13] The whole idea is to improve the security processing and analysis by being compiled into and running directly inside the runtime kernel. The result is efficient observability, tracing, and protection.

It is also important to note that the policy-based access control implementation implied by runtime mitigation is made possible through utilities such as Linux Security Modules (LSMs).[14] This involves use of a framework that allows Linux kernels (assumed for most cloud containers) to support mandatory access control (MAC) without the need for major changes to the underlying kernel.[15]

CASE STUDY: ACCUKNOX INLINE MITIGATION

A useful case study that ties together all of these technical concepts into a working solution that developers can deploy involves the popular commercial cybersecurity platform from a startup company called AccuKnox.[16] Their solution, as will be shown below, expertly integrates inline mitigation with use of popular utilities such as eBPF to accomplish the runtime security goals discussed above.

Understanding the different between inline mitigation, as implemented by AccuKnox, and conventional post-attack mitigation is best explained in the context of enforcement and observability. In the AccuKnox system, implemented in the context of the open-source KubeArmor environment (referenced below), an attacker process is thwarted at the operating system level using LSMs that deny the fork/execution of the unknown process.

Corresponding observability into the process comes from the eBPF functionality that includes probes into the underlying behavior to provide visibility. The resulting telemetry is provided and fed back via the KubeArmor environment to the LSMs to support security policy updates and host policies. The result is an inline security system that allows or denies processes, accesses, and other behavior based on policy (see Figure 2).

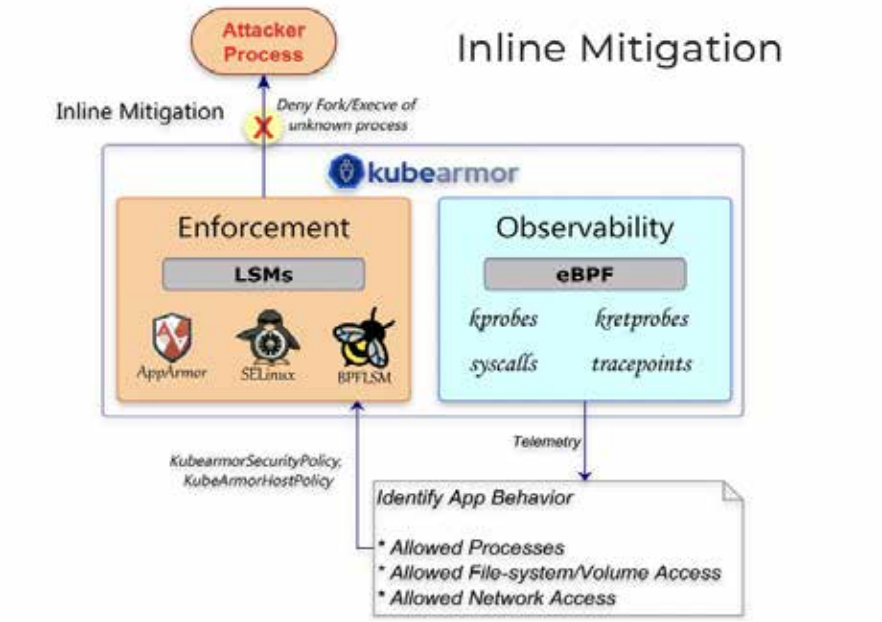

Figure 2. Inline Mitigation Using AccuKnox

In contrast, post-attack mitigation allows the unwanted behavior to occur, presumably some malicious intended execution. The eBPF can still provide visibility, but the telemetry would illustrate an on-going attack, which could then be deleted or killed by event handlers. This is certainly a useful response process, but it is clearly less desirable than stopping the attack in the first place.

ACCUKNOX PLATFORM COVERAGE

The AccuKnox solution broadly aligns well with the static posture assessment for known and runtime control of unknown threats, along with the ability to provide on-going and continuous coverage in both areas using the automated platform. The diagram in Figure 1 below provides a schematic representation of how the AccuKnox commercial platform addresses these key cloud workload protection platform (CWPP) issues.

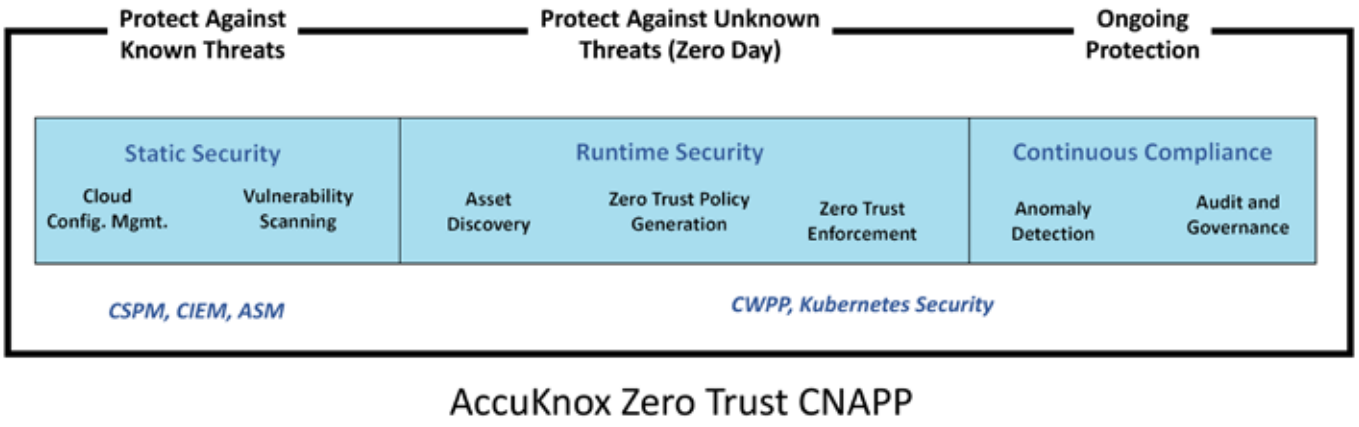

Figure 3. AccuKnox Threat Coverage for Cloud Workloads

ACCUKNOX PLATFORM OPERATION

The AccuKnox platform works by first profiling and creating a baseline of policies by observing application and network graphs. This results in an ongoing observability for security teams as the workload interacts with the host operating system and other workloads. The platform also includes the ability to enforce security policies using kernel primitives. Specific inline cloud workload security in the AccuKnox platform includes the following features:

1. **Inline Prevention** – This approach focuses on preventing attacks before they happen, rather than mitigating them post-attack.
2. **Runtime Container Image Scan** – This involves scanning container images during runtime to detect vulnerabilities or malicious activities.
3. **Audit and Forensics** – AccuKnox provides detailed audit trails and forensics capabilities for deep analysis and understanding of security incidents.
4. **Zero Trust Security** - This principle ensures that trust is never assumed, and that verification is required from everyone trying to access resources in the network.
5. **Runtime Application Security** – Monitoring and hardening applications at runtime to detect and prevent anomalous behaviors.
6. **Microsegmentation** - This involves dividing the network into smaller segments to control traffic and enhance security.
7. **Secrets Management** – Protecting the management of sensitive data like passwords, keys, and tokens.

TKey aspects of the AccuKnox platform include its ability to be non-intrusive to the applications it is running to protect, which is a major advantage in terms of performance without compromising on effectiveness. Most cybersecurity teams prefer to avoid the need to embed agents into the operating environment to minimize performance impacts and reduce the overhead of communication to and from the agent for management.

Integration of the AccuKnox inline solution is also an important benefit – and this includes, of course, ease of deployment into multi-cloud service infrastructure from Google. AWS, Microsoft, Oracle, and IBM. Partnerships are also in place with technology organizations such as the Cloud Native Computing Foundation, OpenSSF, and AICPA SOC 2.

ACCUKNOX ARCHITECTURE

A more detailed view of the architecture can be shown as a multi-layer approach to zero trust. Specifically, the platform operates at the system, network, transport, and application layers, with the requisite functions necessary to address proper cloud-native security. This includes process whitelisting at the system level, microsegmentation at the network level, TLS support at the transport level, and hardening support at the application level.

Monitoring support for cloud workloads operates in a granular manner using observability into process execution, files access, and network behavior. Identities and entitlements are addressed via deployment of a Kubernetes Identity and Entitlement Management (KIEM) solution that has been a popular feature for customers.[17] The functionality includes security for sensitive assets, secrets, 5G and edge workloads (see Figure 4).

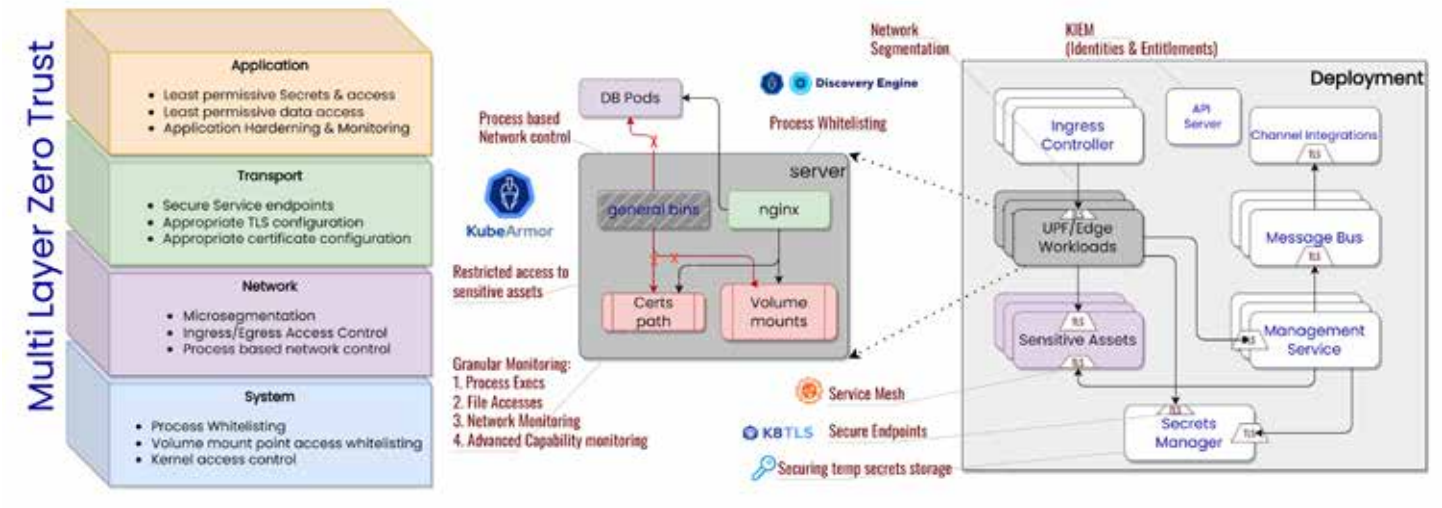

Figure 4. More Detailed View of AccuKnox Multi-Layer Architecture

Technical readers interested in a more detailed view and understanding of the AccuKnox security functionality, including how the specific flows and interactions work between the different software modules that comprise the system (e.g., ingress controller, channel integrations, message bus, API server), should contact AccuKnox directly for a deep dive into the architecture.

ACCUKNOX CONTRIBUTION TO KUBEARMOR

AccuKnox contributes to the KubeArmor project, which uses eBPF for observability of application behavior and Linux Security Modules (LSMs) for enforcement and inline mitigation from unknown zero-day attacks. The CWPP with integrated KubeArmor, includes automated zero trust policy generation, hardening of the workloads, integration with SIEM/SOAR and ITSM, continuous compliance, and unsupervised learning-based anomaly detection.

KubeArmor has been a successful project to date, with AccuKnox reporting over 700,000 downloads of the software. Such extensive use illustrates the value of inline mitigation for cloud workload protection, not to mention comprehensive security for Kubernetes and virtual machine (VM) assets.

ACTION PLAN

Enterprise teams are advised to review their cloud workload protection posture and to consider adding inline mitigation to the protection profile. A source selection process is recommended to identify suitable commercial partners to implement a proof of concept (POC) for inline cloud workload security. TAG strongly advises that AccuKnox be included in this vendor selection process. TAG analysts are always available to provide expert assistance in this regard.

[1] Information on AccuKnox is available on their public website at https://www.accuknox.com/.

[2] Many different public sources exist that offer insights into the growing number and intensity of cyber threats that are occurring around the world targeted enterprise and government teams. The Federal Bureau of Investigation (FBI), for example, regularly publishes data – here's a typical report: https://www.fbi.gov/contact-us/field-offices/springfield/news/internet-crime-complaint-center-releases-2022-statistics.

[3] The research and advisory team at TAG Infosphere tracks global commercial vendors offering products and services in cybersecurity (as well as in artificial intelligence and climate/sustainability). The exact number of cybersecurity vendors in the TAG database varies from month to month, but it averages in the 4600 range. Readers interested in more information should contact TAG at https://www.tag-infosphere.com/.

[4] The benefits of AccuKnox, and any runtime security solution, are perhaps best explained in the context of low-level concepts related to operating systems, cloud containers, workload interfaces, and the like. That said, our assumption is that the readers of this article might have more general backgrounds, so we take every effort here to keep things as readable as possible, with minimal assumptions about the reader's understanding of computer science, Linux, software development, low-level system design, and cloud technology.

[5] The NIST Cybersecurity Framework (CSF) is one of the most important contributions to cybersecurity in history and it can be review in detail here: https://www.nist.gov/cyberframework.

[6] Unlike NIST CSF, the MITRE ATT&CK Framework models exploits into a structure that helps defenders (and yes, also attackers) better understand the components of the offensive lifecycle of threats. It can be reviewed here: https://attack.mitre.org/.

[7] MITRE maintains the well-known CVE database here: https://cve.mitre.org/. In our estimation, the MITRE ATT&CK model has tended to be more useful for vendors mapping their features than for practitioners trying to develop an operating framework for defense.

[8] Many excellent articles are available that provide insights into zero-day vulnerabilities. A good sample article is available here that explains how zero-day attacks work: https://www.pcworld.com/article/2113913/zero-day-exploits-how-to-protect-yourself.html.

[9] This author first became aware of the use of artificial intelligence (simple machine learning at first) from Stuart McClure in roughly 2012, when the Cylance product was first being developed. Since then, most cybersecurity products and services include some measure of machine or deep learning in their engines. Cylance has since been absorbed into Blackberry.

[10] Throughout this report, we will use the terms workload and application roughly synonymously. Some practitioners draw distinctions, often viewing workloads as covering a business function such as human resources or payroll, where applications cover a specific function that could be part of a deployed workload. Here, we do not need to rely on any distinction between workloads and applications to explain how in-line mitigation at runtime will work.

[11] Technical and business-related information on these companies is available at their public websites at https://www.wiz.io/ and https://www.paloaltonetworks.com/.

[12] Detailed technical as well as tutorial information on the extended Berkeley Packet Filter (eBPF) and how it operates with hooks into the Linux kernel is available in articles across the Internet. A typical explanation is available at https://ebpf.io/what-is-ebpf/.

[13] Several excellent case studies are available on the eBPF website that show how companies such as Google. Netflix. S&P Global, Shopify, and Cloudflare use the technology for various types of inline security tasks.

[14] This on-line presentation does a good job explaining how Linux Security Modules (LSMs) work, including their architecture and how they integrate into the Linux kernel: https://elinux.org/images/0/0a/ELC_Inside_LSM.pdf.

[15] This author participated in one of the earliest efforts at Bell Labs to insert inline MAC controls into UNIX-style operating systems in the 1980's and 1990's consistent with the then-relevant Trusted Computer System Evaluation Criteria (Orange Book) that guided security designs at the time. LSMs include many of the same design principles identified in those early kernel-level protection projects.

[16] The company was founded in 2020 and is based in Cupertino, California. The executive and technical teams at AccuKnox were instrumental in helping to support the writing of this report and were helpful to explain how their platform works.

[17] See https://www.accuknox.com/press-release/ixel-sas-partnership for details of a recent partnership forged by AccuKnox that is centered on KIEM functionality provision for customers.

ABOUT TAG

TAG is a trusted research and advisory company that provides insights and recommendations in cybersecurity, artificial intelligence, and climate science to thousands of commercial solution providers and Fortune 500 enterprises. Founded in 2016 and headquartered in New York City, TAG bucks the trend of pay-for-play research by offering unbiased and in-depth guidance, market analysis, project consulting, and personalized content—all from a practitioner perspective.

IMPORTANT INFORMATION ABOUT THIS DOCUMENT

Contributor: Dr. Edward Amoroso

Publisher: TAG Infosphere Inc., 45 Broadway, Suite 1250, New York, NY 10006.

Inquiries: Please contact Lester Goodman at lgoodman@tag-cyber.com to discuss this report. You will receive a prompt response.