Der neue Standard bei CNAPP: Hyper-Priorisierung und autonome Behebung im Cloud-Maßstab

Shrikant Dhanawade, Director, Product Management, Cloud-Sicherheit, Qualys

Die KI-gestützte Erkennung hat eine Schwelle überschritten. Sicherheitsteams können nun Schwachstellen, Fehlkonfigurationen und aktive Angriffspfade mit einer Geschwindigkeit und in einem Umfang aufdecken, der vor einigen Jahren noch unvorstellbar war. Das Problem besteht nicht mehr darin, Risiken zu finden oder zu erkennen, sondern darin, sie schnell genug zu beheben, damit es wirklich etwas bewirkt.

Cloud-Bereitstellungen verstärken diesen Druck auf besondere Weise: Die Infrastruktur, die Sicherheitsteams eilig beheben müssen, verändert sich, skaliert, wird neu bereitgestellt und neu konfiguriert – und das schneller, als es manuell möglich ist. Die Frage, die sich jedes Sicherheitsteam derzeit stellt, lautet daher: „Wie können wir in Cloud-Umgebungen, die niemals stillstehen, Prioritäten setzen und Abhilfemaßnahmen mit der gleichen Geschwindigkeit durchführen, mit der die Erkennung erfolgt?“

Wichtige Erkenntnisse

- Die Cloud ist ein sich ständig veränderndes Ziel. Ihre Sicherheitslage muss damit Schritt halten. Geplante Scans und wöchentliche oder monatliche Überprüfungen können mit einer sich stündlich ändernden Infrastruktur nicht Schritt halten. Ein kontinuierliches Sicherheitsmanagement ist eine Grundvoraussetzung und keine Premium-Funktion.
- CNAPP bietet den größten Nutzen, wenn die Erkennung direkt in die Behebung einfließt. Die Korrelation von Ergebnissen aus CSPM-, CWP- und Schwachstellendaten ist zwar notwendig, aber nicht ausreichend. Der Nutzen liegt darin, Risiken zu beseitigen, nicht darin, sie in einem beeindruckenden Dashboard zusammenzufassen.

- Durch Hyper-Priorisierung werden Cloud-Risiken von Cloud-Rauschen unterschieden. Es gibt Tausende von Fehlkonfigurationen, aber nur ein paar Dutzend ausnutzbare Angriffspfade, die derzeit von Bedeutung sind. Den Unterschied zu kennen, ist entscheidend.
- LLM-gestützte Playbooks ermöglichen eine überschaubare Reaktion auf Zero-Day-Angriffe im Cloud-Maßstab. Wenn neue Bedrohungen auftreten, liegt der Engpass nicht in der Erkennung, sondern darin, schnell genug einen glaubwürdigen, umgebungsspezifischen Reaktionsplan zu erstellen, der tatsächlich etwas bewirkt. KI beseitigt diesen Engpass.
- Eine dauerhafte Behebung muss den Kreislauf vom Code-Commit bis zur laufenden Workload schließen. Das Patchen eines Containers ohne Korrektur des Images oder die Korrektur des Images ohne Aktualisierung der Pipeline-Richtlinie garantiert ein erneutes Auftreten des Problems. Alle drei Ebenen müssen gemeinsam voranschreiten.

Die Cloud entwickelt sich rasant - Angreifer ebenfalls

Fragt man die meisten Sicherheitsteams nach ihrer größten Herausforderung in der Cloud, werden sie etwas über mangelnde Transparenz, zu viele Ressourcen, zu viele Tools und zu viele Daten sagen. Transparenz ist ein echtes Problem. Aber es ist nicht das schwierigste Problem. Das schwierigste Problem ist, dass die Cloud-Infrastruktur von Natur aus dynamisch ist. Dadurch werden die Annahmen untergraben, auf denen traditionelle Sicherheitsprogramme basieren. Das „Weaponization Window“ (die Zeitspanne von der Veröffentlichung eines CVE bis zum aktiven Einsatz eines Exploits in der Praxis) hat sich in den letzten Jahren von Wochen auf Stunden verkürzt. KI unterstützt nicht nur Verteidiger, sondern hilft auch Angreifern dabei, Angriffspfade schneller zu entdecken und abzubilden, als es jede von Menschen durchgeführte Analyse nachhalten kann. Statische Angriffspfadmodelle, die nur wöchentlich aktualisiert werden, sind bereits veraltet.

Von Natur aus kurzlebig

Ein für einen Batch-Job gestarteter Container kann maximal vier Minuten bestehen bleiben. Eine serverlose Funktion wird in Millisekunden ausgeführt. Eine Auto-Scaling-Gruppe kann als Reaktion auf einen Traffic-Anstieg Dutzende von Instanzen hinzufügen und entfernen, noch bevor ein wöchentlicher Scan überhaupt ausgeführt wurde. Herkömmliche Schwachstellenscanner wurden jedoch für eine Welt entwickelt, in der Ressourcen lange genug vorhanden waren, um gescannt, bewertet, priorisiert und gepatcht zu werden. In Cloud-Umgebungen kann die Ressource jedoch bereits verschwunden sein, bevor die Priorisierung überhaupt beginnt.

Dies ist kein Problem der Scan-Häufigkeit, das durch schnellere Scanner behoben werden kann. Es handelt sich um eine architektonische Diskrepanz. Sicherheitsprogramme, die auf periodische Bewertungszyklen angewiesen sind, können mit der sich ständig weiterentwickelnden Infrastruktur nicht Schritt halten. Es muss sich nicht nur der Scan-Zeitplan, sondern auch das Reaktionsmodell ändern.

Die Vielfalt der Bereitstellungsarten vergrößert die Angriffsfläche

Moderne Cloud-Umgebungen sind nicht homogen. Ein einzelnes Unternehmen kann beispielsweise Folgendes betreiben:

- Virtuelle Maschinen auf AWS EC2, Azure VMs und GCP Compute Engine mit jeweils unterschiedlichen Patching-Mechanismen und Agent-Unterstützung
- Kubernetes-Cluster, sowohl über Managed Services (EKS, AKS, GKE) als auch selbstverwaltet. Hier gibt es Sicherheitsaspekte auf Node- und Pod-Ebene, die sich nicht auf herkömmliche Host-Modelle übertragen lassen.
- Serverlose Funktionen in Lambda, Azure Functions und Cloud Run, bei denen es kein Betriebssystem gibt, das gepatcht werden muss, und bei denen die Angriffsfläche vollständig in der Laufzeitumgebung und den Abhängigkeiten liegt.
- Container-Images, die aus Basis-Images erstellt wurden. Diese können Schwachstellen enthalten, die bereits Monate vor der Bereitstellung der Workloads entstanden sind.

- Infrastructure-as-Code-Vorlagen, die Fehlkonfigurationen festschreiben, noch bevor eine einzige Workload ausgeführt wird.

Jede dieser Bereitstellungsarten weist ein anderes Sicherheitsmodell, andere Abhilfemaßnahmen und ein anderes Risikoprofil auf. Eine anfällige Anwendung wird verschoben und skaliert, um die Verfügbarkeit zu verbessern. Ein einheitlicher Workflow für das Schwachstellenmanagement, der alle diese Fälle gleichbehandelt, ist für die meisten von ihnen ungeeignet.

Die Erkennungsleistung von CNAPP ist gut, aber reicht die Geschwindigkeit aus?

Die meisten modernen CNAPP-Plattformen sind zwar hervorragend darin, Cloud-Sicherheitssignale über diese weitläufige und flüchtige Angriffsfläche hinweg zu aggregieren und zu korrelieren, sie sind jedoch keine Plattformen zur Behebung von Schwachstellen. Sie sind jedoch keine Plattformen zur Behebung von Schwachstellen. Die Lücke zwischen einem einheitlichen Befund und einer behobenen Schwachstelle wird nach wie vor überwiegend durch einen manuellen Arbeitsablauf geschlossen. Es wird ein Ticket eröffnet, ein Team benachrichtigt, eine Änderung genehmigt, eine Bereitstellung durchgeführt und ein Scan zur Bestätigung erneut ausgeführt. In einer dynamischen Cloud-Umgebung dauert dieser Prozess Tage. Bis das Ticket gelöst ist, wurde die Infrastruktur, die den Befund verursacht hat, möglicherweise bereits zweimal ersetzt.

Qualys TotalCloud™ wurde entwickelt, um diese Lücke zu schließen. Die Erkennung erfolgt sofort, in Echtzeit und auf Basis von Cloud-Ereignissen. Als einheitliche CNAPP-Lösung korreliert die Lösung Signale zu Schwachstellen, Fehlkonfigurationen, Identitätsproblemen und Befunden zu sensiblen Daten zu einem einzigen, kontinuierlichen Risikobild und verknüpft dieses Bild direkt mit der Behebung. Das Ergebnis ist eine Sicherheitslage, die nicht nur mehr erkennt, sondern auch schneller reagiert. Von einer falsch konfigurierten IAM-Rolle bis hin zu einem aktiv ausnutzbaren Angriffspfad stellen erstklassige CNAPP-Lösungen sicher, dass jedes von der Erkennungs-Engine aufgedeckte Signal einen direkten, automatisierten Weg zur Behebung hat.

Um diese Lücke zu schließen, muss das korrelierte Signal eines CNAPP mit einer autonomen Behebungsfunktion verknüpft werden, die mit der Geschwindigkeit und dem Umfang der Cloud agieren kann.

Eine CNAPP, die zwar innerhalb von Sekunden einen ausnutzbaren Angriffspfad aufdeckt, aber eine Woche benötigt, um ihn zu schließen, bietet kein angemessenes Sicherheitsergebnis. Sie verursacht lediglich einen größeren Rückstau.

Entscheidend ist nicht, wie viele Schwachstellen erkannt wurden, sondern wie lange sie offenblieben, sodass ein Angreifer sie ausnutzen konnte. Die manuelle Priorisierung ist der Hauptgrund dafür, dass sie so lange offenbleiben.

Hyper-Priorisierung in Cloud-Umgebungen: Den Überblick behalten

Eine ausgereifte CNAPP-Implementierung in einem Unternehmen kann wöchentlich Zehntausende von Befunden zutage fördern. Allein das CSPM generiert in großen Umgebungen üblicherweise Zehntausende Richtlinienverstöße. Ohne konsequente Priorisierung sehen sich Sicherheitsteams einer unüberschaubaren Triage-Last gegenüber und greifen standardmäßig auf die Vorgehensweisen „vom Ältesten zum Neuesten“ oder „mit dem höchsten CVSS-Wert zuerst“ zurück. Dabei spiegelt keine dieser Methoden das tatsächliche Risiko wider. Zudem ignorieren die meisten Unternehmen fast ein Drittel ihrer als geringfügig eingestuften Warnmeldungen. Dies ist problematisch, da das ausschließliche Verlassen auf den CVSS dazu führen könnte, dass Warnmeldungen aufgrund falscher Annahmen übersehen werden, die eigentlich sofort isoliert und behoben werden müssten.

Unternehmen, die Compliance-Vorgaben und -Rahmenwerken wie NIST 800-53 unterliegen, müssen zudem die Anforderungen an eine kontinuierliche Überwachung und schnelle Reaktion berücksichtigen. Die Lösung? Eine fortgeschrittene Priorisierung, die sich auf die wichtigsten Exploits konzentriert – eine Art „Hyper-Priorisierung“ auf der Grundlage mehrerer Kriterien.

Ausnutzbarkeit im Kontext, nicht isoliert betrachtet

Jede Ressource wird kontinuierlich anhand eines Live-Bedrohungs-Feeds neu bewertet, während sich die Bedrohungslandschaft weiterentwickelt. Dabei werden auch die folgenden Risikofaktoren berücksichtigt:

- **Exposition:** Ist die betroffene Workload mit dem Internet verbunden? Ist der anfällige Port von außerhalb der VPC erreichbar? Eine kritische Schwachstelle bei einem isolierten internen Dienst unterscheidet sich

grundlegend von derselben Schwachstelle bei einem öffentlich zugänglichen Endpunkt mit Lastenausgleich.

- **Identität und Zugriff:** Verfügt die kompromittierte Workload über eine IAM-Rolle mit weitreichenden Berechtigungen? Kann ein Angreifer, der diese Schwachstelle ausnutzt, auf andere Konten oder Regionen zugreifen oder Daten exfiltrieren? Der Auswirkungsbereich eines Angriffs hängt stark davon ab, wozu die Workload berechtigt ist.
- **Wege der lateralen Bewegung:** Die Analyse von Angriffspfaden über die Cloud-Topologie hinweg zeigt, welche Schwachstellen, wenn sie ausgenutzt werden, einen Weg zu den wertvollsten Assets bieten. So kann ein Befund mit niedrigem CVSS-Wert bei einer Workload mit Netzwerkzugriff auf eine Produktionsdatenbank eine höhere Priorität haben als ein kritischer CVE-Befund auf einer isolierten Entwicklungsinstanz.
- **Geschäftlicher Kontext:** Nicht alle Ressourcen haben das gleiche Gewicht. Eine Schwachstelle im Zahlungsabwicklungsdienst, im Kundendatenspeicher oder in einer für die Compliance kritischen Workload erfordert ein anderes Maß an Reaktionsdringlichkeit als derselbe Befund in einem internen Entwicklungstool.
- **Aktive Bedrohungsfeeds:** Laufzeitsensoren beobachten tatsächliches verdächtiges Verhalten, ungewöhnliche Prozessausführungen, unerwartete Netzwerkverbindungen sowie Zugriffsmuster bei Anmeldedaten und erhöhen die Priorität der damit verbundenen Schwachstellen von „theoretisch“ auf „bestätigt – aktiv“.
- **Ausgleichende Kontrollmaßnahmen:** Eine Schwachstelle, die durch eine WAF-Regel abgedeckt ist, welche den spezifischen Angriffsvektor blockiert, oder die sich in einer Workload befindet, in der der anfällige Codepfad niemals ausgeführt wird, birgt ein geringeres effektives Risiko. Durch die Berücksichtigung von Kontrollmaßnahmen wird verhindert, dass die Behebungswarteschlange von Schwachstellen dominiert wird, die bereits gemindert wurden.

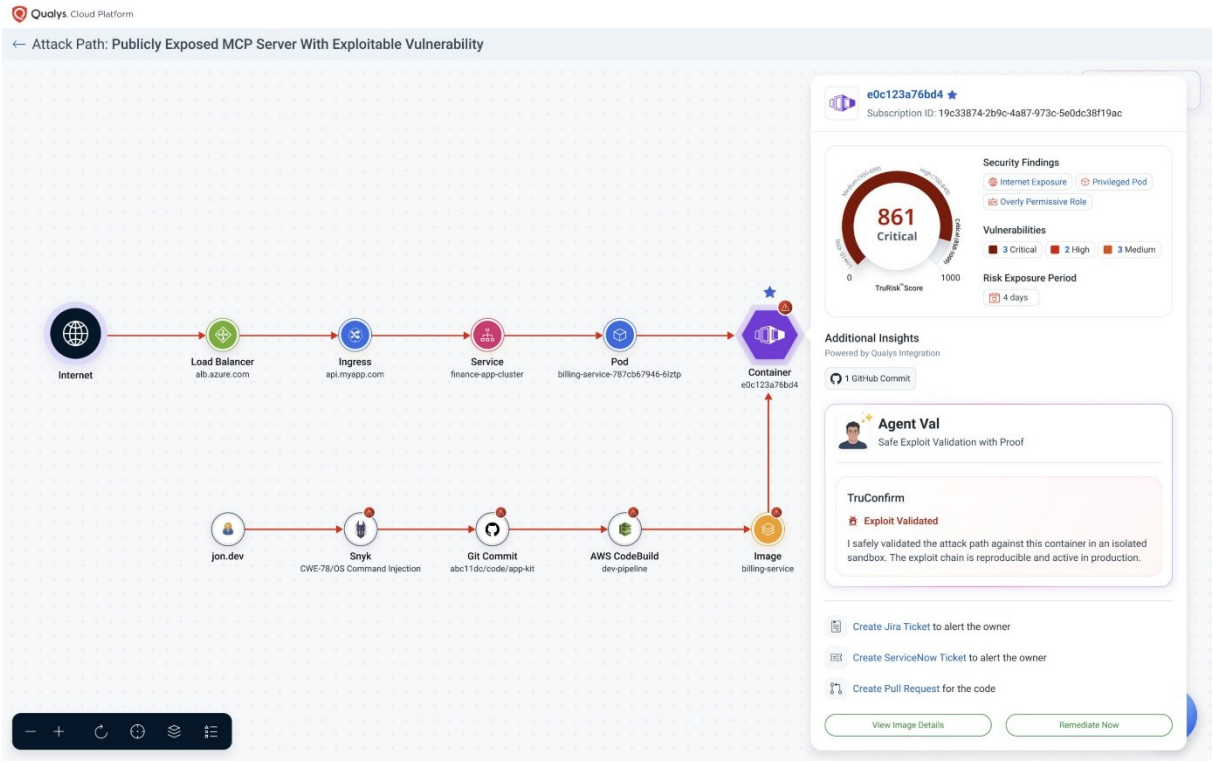
Der Befund mit der höchsten Priorität ist nicht derjenige mit der höchsten Punktzahl in einer einzelnen Dimension, sondern derjenige, bei dem mehrere Risikofaktoren zusammenkommen. Nicht Tausende von Befunden, sondern nur Dutzende. Wenn sie nicht behoben werden, stellen Schwachstellen und

Fehlkonfigurationen einen glaubwürdigen Weg zu einer schwerwiegenden Sicherheitsverletzung dar. Dies ist die Liste, auf die die autonome Behebung zuerst reagieren sollte.

> *Die Hyper-Priorisierung in der Cloud geht über eine genaue Bewertung hinaus. Sie ist unerlässlich, um eine Liste mit Tausenden von Schwachstellen auf die wenigen Dutzend zu reduzieren, die aktuell in Ihrer spezifischen Umgebung ein echtes, unmittelbar bevorstehendes und ausnutzbares Risiko darstellen.*

Analyse von Angriffspfaden und Exploit-Validierung: Priorisieren Sie vorrangig das, was für Sie gerade wichtig ist

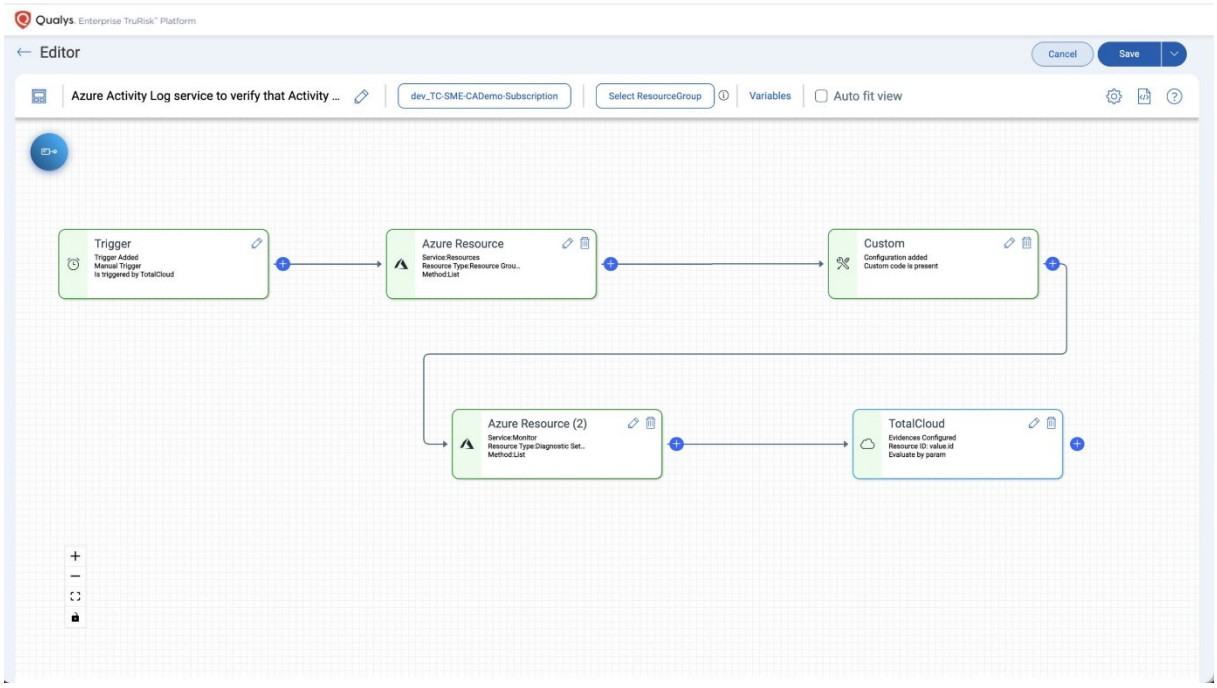
Die Angriffspfadanalyse von TotalCloud übernimmt die Hauptarbeit für die autonome Behebung, indem sie ausnutzbare Pfade in der Cloud-Topologie mit höchster Präzision abbildet und genau ermittelt, welche Schwachstellen einen begehbaren Weg zu Ihren kritischsten Ressourcen darstellen. Qualys geht mit TruConfirm™ noch einen Schritt weiter: Diese Exploit-Validierung bestätigt, ob eine Schwachstelle in Ihrer Live-Umgebung tatsächlich ausnutzbar ist. Das Ergebnis ist keine Rangliste mit Tausenden von Einträgen. Es ist eine eindeutige, validierte Antwort auf eine einzige Frage: Was müssen Sie sofort beheben? Nicht heute. Nicht diese Woche. Sondern genau jetzt.



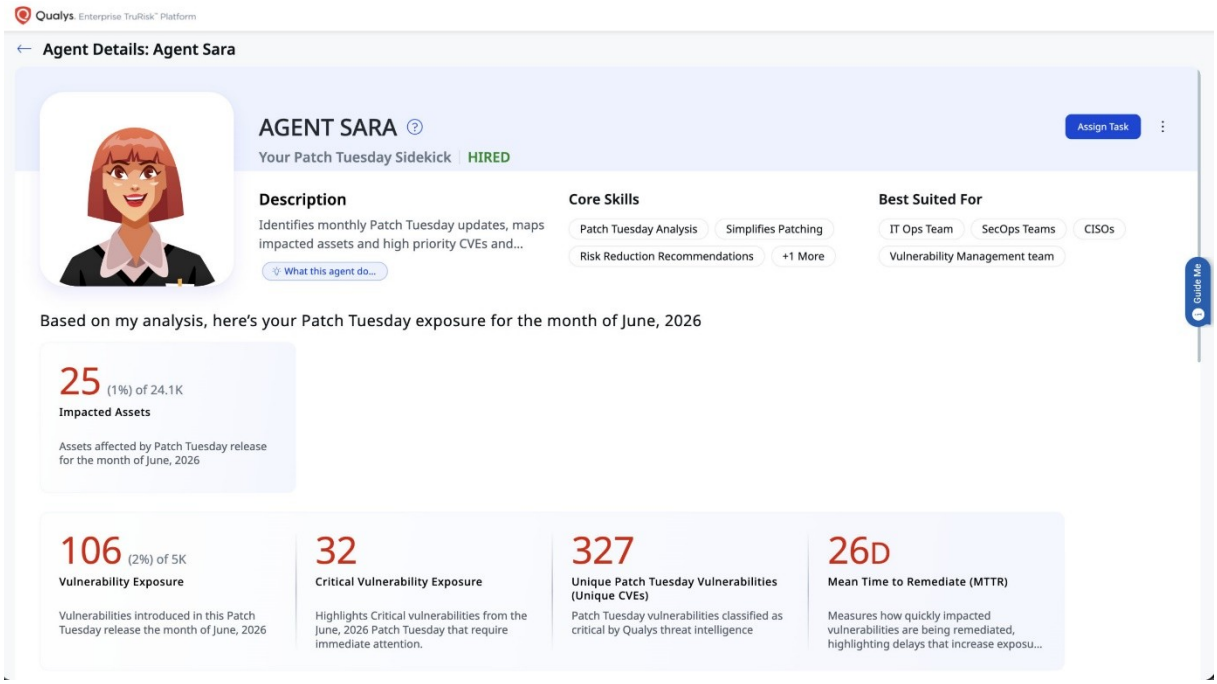
Autonome Fehlerbehebung im Cloud-Maßstab: Was „autonom“ eigentlich bedeutet

Autonome Fehlerbehebung in Cloud-Umgebungen ist keine einzelne Maßnahme. Vielmehr handelt es sich um ein Spektrum an Maßnahmen, die jeweils für unterschiedliche Kategorien von Befunden, Asset-Typen und Konfidenzniveaus geeignet sind. Die richtige Auswahl aus diesem Spektrum ist ebenso wichtig wie die Verfügbarkeit der entsprechenden Funktionen an sich. So sieht das in der Praxis aus:

- **Vollständige Automatisierung bei Konfigurationsabweichungen.** Fehlerhafte Cloud-Konfigurationen sind das Ziel mit dem höchsten Konfidenzgrad. Der Befund ist präzise: Bei diesem S3-Bucket ist der öffentliche Zugriff aktiviert. Die Korrektur ist deterministisch: Deaktivieren Sie den öffentlichen Zugriff und das Risiko der automatisierten Maßnahme ist gut nachvollziehbar. Die CSPM-gesteuerte Behebung wird über die API des Cloud-Anbieters innerhalb von Sekunden ausgelöst und bei erneutem Auftreten der Fehlkonfiguration erneut ausgeführt. Qualys TotalCloud wird mit über 300 „No-Code“-Playbooks über QFlow ausgeliefert. Dabei handelt es sich um die integrierte Cloud-Workflow-Automatisierungs-Engine, die die gängigsten Kategorien bei AWS, Azure und GCP abdeckt, ohne dass Skripte erforderlich sind.



- **LLM-gestützte Triage für neue Bedrohungen.** Tritt eine Zero-Day-Bedrohung auf, für die es noch kein Playbook gibt, erstellt die LLM-gesteuerte Workflow-Generierung von QFlow innerhalb von Sekunden einen umgebungsspezifischen Reaktionsplan. Dieser umfasst die sofortige Eindämmung, patchlose Abhilfemaßnahmen durch cloudnative Kontrollen und den Weg zu einer dauerhaften Behebung. Dieser Ausgangspunkt wird in Maschinengeschwindigkeit erstellt und ist zur Überprüfung durch den Menschen bereit, anstatt von Grund auf neu verfasst zu werden.
- **Autonomes Patchen, Abwehren oder Isolieren je nach Workload.** TruRisk Eliminate™ leitet jeden Befund automatisch an die richtige Reaktion weiter. Wenn ein Patch verfügbar ist und von der KI als zuverlässig eingestuft wird, wird er autonom bereitgestellt. Wenn die Konfidenzschwellen nicht erreicht werden, wird der Patch zurückgehalten und im Rahmen eines wellenbasierten Rollouts schrittweise bereitgestellt. Wenn kein Patch verfügbar ist, wendet das System patchlose Kontrollen, WAF-Regeln und Richtlinienänderungen an. Und wenn eine Workload zu riskant ist, um sie anzutasten, wird sie vom Netzwerk isoliert, bevor eine laterale Bewegung über Konten oder Regionen hinweg stattfinden kann.
- **Agent Sara koordiniert den gesamten Prozess von Anfang bis Ende.** Als Teil des „Agentic AI“-Frameworks von Qualys sortiert sie die Befunde, wählt den richtigen Reaktionsmodus aus, führt die Maßnahme durch und überprüft den Abschluss – und das alles ohne menschliches Eingreifen. Was früher einen Techniker im Bereitschaftsdienst erforderte, der die verschiedenen Tools koordinieren musste, läuft nun autonom und mit der Geschwindigkeit ab, die die Cloud verlangt.



Die autonome Fehlerbehebung in der Cloud ist kein Schwarz-Weiß-Szenario. Bei Konfigurationsabweichungen ist eine vollständige Automatisierung möglich. Beim „Human-in-the-Loop"-Ansatz werden Änderungen der Arbeitslasten, die mit operativen Risiken verbunden sind, manuell überprüft. Bei neuartigen Bedrohungen, für die es noch kein Playbook gibt, kommt eine LLM-gestützte Triage zum Einsatz.

Fazit für Cloud-Sicherheitsteams

Die Cloud-Sicherheit ist eine besondere Herausforderung, da die zu schützende Umgebung von Natur aus dynamisch ist. Die Infrastruktur verändert sich ständig über verschiedene Bereitstellungsarten hinweg, die jeweils unterschiedliche Risikomodelle mit sich bringen. All dies wird von Teams verwaltet, die schneller agieren als ursprünglich von den Sicherheitsprogrammen vorgesehen.

CNAPP bietet Sicherheitsteams einen einheitlichen Überblick über die gesamte Umgebung. Durch die Hyper-Priorisierung können sie sich auf das Wesentliche konzentrieren. Die autonome Behebung von Sicherheitslücken ermöglicht es ihnen, Risiken schnell zu beseitigen, bevor diese ausgenutzt werden können. Keine dieser Maßnahmen reicht für sich allein aus. Zusammen bilden sie jedoch ein Sicherheitsprogramm, das mit den neuesten KI-gestützten Cloud-Bedrohungen tatsächlich Schritt halten kann.

CNAPP für das Zeitalter der Frontier-KI

Qualys TotalCloud™ ist eine KI-native CNAPP-Lösung (Cloud Native Application Protection Platform), die Unternehmen dabei unterstützt, den Schritt von der Transparenz zur autonomen Risikobeseitigung zu vollziehen. Mithilfe der Funktionen FlexScan, KCS, TruRisk, TruConfirm, CDR, DSPM, QFlow und Eliminate schützt TotalCloud Cloud-, Container-, Kubernetes-, Identitäts-, Daten-, serverlose, Laufzeit- und SaaS-Umgebungen. Gleichzeitig werden die Priorisierung, Compliance und die Geschwindigkeit der Fehlerbehebung verbessert.