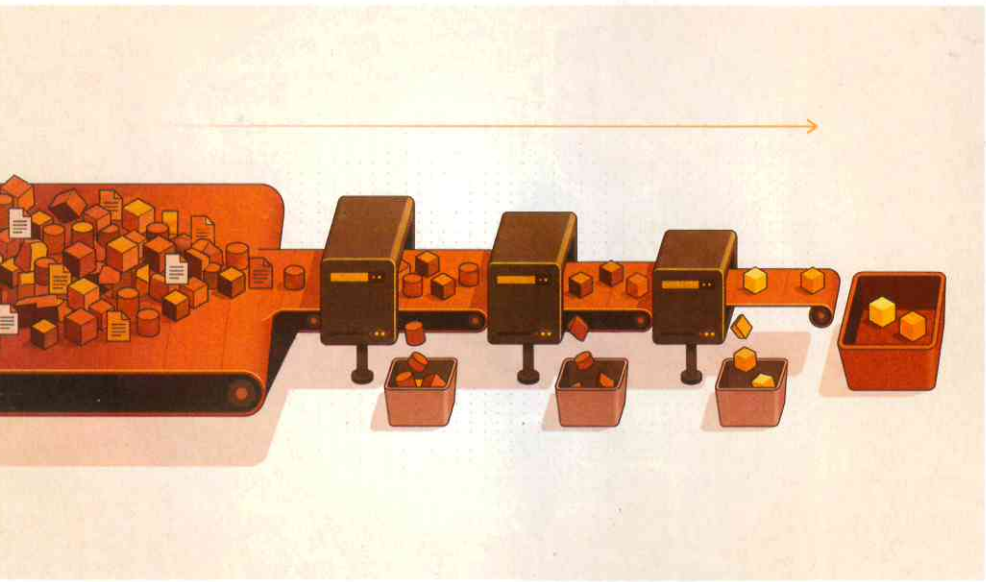


Quelle: Cathrin Kapell / KI / iX

Data Pipelines für optimierte Sicherheitsdaten

Für IT-Sicherheit, aber auch für Netzwerkmanagement, Fehlersuche und weitere Bereiche werden Log- und Eventdaten benötigt – allerdings nicht die gesamte Datenflut, die in Unternehmen üblicherweise anfällt. Plattformen für Security Data Pipelines sind Produkte, die Daten nach Relevanz und für spezielle Zwecke aufbereiten. Das kann beim Weiterverarbeiten bei nachgelagerten Dienstleistern den Geldbeutel deutlich entlasten.

Von Patrick Bäuerle und Stefan Strobel

Security Data Pipeline Platforms (SDPP) bilden eine neue Produktkategorie im IT-Security-Markt, die sich gerade in der gewachsenen Welt der Security Operation Center (SOCs), Security-Information-and-Event-Management-Systeme (SIEMs), XDR-Plattformen (Extended Detection and Response) und moderner KI-Produkte für Security etabliert.

Das tägliche Volumen an Logs, Events, Traces, Netflows oder auch Threat Intelligence steigt stetig, immer mehr IT-Services, Cloud-Komponenten, Microservices und Sicherheitslösungen erzeugen solche Daten. Außer in der IT-Sicherheit benötigt man sie auch im Troubleshooting, im Netzwerkmanagement und weiteren Bereichen. Externe SOC-Dienstleister und SIEMs in der Cloud wie Microsoft Sentinel oder Google Chronicle vereinfachen die Situation nicht, sondern erhöhen die Komplexität weiter. Die frühere Idee, alle Logs und andere relevante Daten in ein einziges zentrales SIEM einzuliefern, hat sich in der Praxis nicht bewährt und kommt allein schon wegen der massiv steigenden Datenmenge an ihre Grenzen.

So findet man heute häufig mehrere Logempfänger und Logspeicher, die zum Teil dieselben, zum Teil unterschiedliche Daten verarbeiten, speichern und unterschiedlich lange aufbewahren. Allein in der IT-Sicherheit unterscheidet man zwischen den Daten, die zur Erkennung von Vorfällen wichtig sind und die meistens für viel Geld in einem SIEM gespeichert werden, und solchen, die nur zur späteren Analyse bereits erkannter Vorfälle relevant sind und in einem deutlich günstigeren Langzeitspeicher abgelegt werden. Moderne KI-Produkte, die im SOC die Analysten unterstützen sollen, benötigen ebenfalls Daten, bevorzugt in einem normalisierten Format.

Moderne Security Data Pipeline Platforms können die Logs mit den jeweils nötigen Inhalten an die relevanten Empfänger leiten. Sie verfügen meist über Features zum Reduzieren der Datenmenge, aber auch zum Transformieren und Anreichern der Inhalte. Die Pipelineprodukte bilden dabei die Schnittstelle zu den Datenquellen, überwachen den Datenfluss, erkennen Störungen und Änderungen im Schema der Daten und stellen einen Überblick dar. Sie entwickeln sich damit zu einer zentralen Komponente in einem SOC.

Cloud-SIEMs: Erkennen von Vorfällen

Die Erkennung von Sicherheitsvorfällen in einem SIEM in der Cloud ist ein typischer Anwendungsfall für die anfallenden Daten. Viele Unternehmen setzen dabei auf Services wie Microsoft Sentinel oder Google Chronicle. Dadurch können sie externen SOC-Dienstleistern Zugriff auf die Cloud-SIEMs im eigenen Tenant geben, ohne dass die Ausgangsdaten im

iX-TRACT

- Die Flut an Daten aus den verschiedenen IT-Systemen, Services, Anwendungen und mehr ist für die benötigten Zwecke wie Sicherheitsvorfallerkennung oder Troubleshooting kaum mehr zu beherrschen. Hinzu kommen hohe Storage-Kosten.
- Security Data Pipeline Platforms, eine neue Kategorie von Produkten, sind in der Lage, diese Daten vorzufiltern, an verschiedene Kanäle weiterzuleiten und bei Bedarf mit zusätzlichen Informationen anzureichern.
- Zahlreiche sich verändernde und flexiblere Features durch neue Möglichkeiten wie KI, aber auch der Kostendruck für immer größere Datenmengen haben dazu geführt, dass sich die Produkte verändern, Produktkategorien überlappen und teils durch andere Produkte und Dienste ersetzbar sind. Hier gilt es, die Bedürfnisse des eigenen Unternehmens zu eruieren und entsprechend zu wählen.

Typische Daten, die über eine SDPP weitergeleitet werden können

- EDR-Telemetriedaten und Logs der Endgerätebetriebssysteme
- Firewall-Logs, IDS/IPS-Events, VPNs beziehungsweise Remote-Access
- Logs von Mailgateways beziehungsweise Mailservern
- Logs von Webgateways beziehungsweise SASE/SSE
- NDR-Telemetrie, Datenflüsse und Anomalien im Netzwerk
- Logs von UEBA-Systemen (User and Entity Behavior Analytics)
- DNS-Logs
- Cloud-Audit-Logs von AWS, Azure, Google et cetera
- Logs der Identity-Systeme beziehungsweise Authentisierungs-Logs von AD, Entra ID, Okta, Ping et cetera
- Applikations- und SaaS-Logs
- DLP- beziehungsweise File-Access-Logs, z. B. von Purview
- externe Threat Intelligence, Geolocation et cetera
- Logs von physischen Zutrittskontrollsystemen
- sonstige Syslog-Quellen

Rechenzentrum des Dienstleisters verarbeitet werden müssen und so eine unerwünschte starke Abhängigkeit vom SOC-Dienstleister entsteht.

Das Einliefern von Daten, die nicht ohnehin schon in der Cloud des Anbieters vorliegen, ist teuer. Daher ist es naheliegend, nur diejenigen Logs, Events et cetera in die Cloud zu senden, die dort auch benötigt werden.

Aber auch unabhängig von der Kostenfrage muss es eine zentrale Sammelkomponente geben, die lokale Logs im richtigen Format in die Cloud weiterleitet. Statt der vergleichsweise einfachen Gateways, die die Cloud-SIEM-Plattformen bereitstellen, kann man das mit komfortableren Pipelines eleganter lösen. Sie bieten oft Hunderte vordefinierter Konnektoren zu bekannten Securityprodukten. Für neue Datenquellen, für die die Pipeline noch keinen fertigen Konnektor hat, werden typischerweise Varianten wie Syslog, eigene APIs, Webhooks, Abfragen von Speichern et cetera angeboten und für nicht bekannte Datenformate geht der Trend zum KI-basierten Parsen und Transformieren.

Wenn man mit einem klassischen externen SOC-Dienstleister zusammenarbeitet, der alle relevanten Daten in seinem Rechenzentrum benötigt, ist die Situation ähnlich. Auch in diesem Fall ist es wenig hilfreich, alle Daten unreflektiert an den Dienstleister zu schicken, denn auch dabei entstehen in der Regel volumenabhängige Kosten.

Daten sinnvoll verteilen

Zudem erzeugen die relevanten Quellsysteme auch Logs, die intern für den IT- und Netzwerkbetrieb benötigt werden. Oft ist es nicht möglich, am Quellsystem mehrere Logziele mit unterschiedlich gefilterten Inhalten einzustellen, von der dabei entstehenden Komplexität einmal abgesehen. Schon aus diesem Grund ist eine interne Logverteilungskomponente wünschenswert, die sämtliche Logs von allen Quellen empfängt und dann selektiv die relevanten Daten an verschiedene Ziele weiterleiten kann. Nur der Teil der Logs, den das SIEM in der Cloud für Erkennungs-Use-Cases benötigt, gelangt dorthin, während ein größerer Ausschnitt der Logs für den internen Betrieb an das Logmanagementsystem weitergeleitet wird.

SDPPs können die Daten meist zusätzlich auf viele Arten filtern und verdichten, was das Datenvolumen weiter reduzieren kann. Beispielsweise können sie Spalten oder Felder entfernen, die ein bestimmtes Zielsystem nicht benötigt, oder sich häufig wiederholende Logs zusammenfassen. Gerade bei Netzwerkfirewalls gibt es oft Situationen, bei denen fast identische Logeinträge mehrmals in einer Minute erzeugt werden. Eine Pipeline kann durch Weglassen der Wiederholungen oder Zusammenfassen der Einträge Redundanz vermeiden.

Auch im Kontext Datenschutz spielt Filtern eine Rolle. Je nach Zielsystem oder auch nach Land, in dem ein Zielsystem steht, kann es wünschenswert oder nötig sein, personenbezogene Daten durch pseudonymisierte Daten zu ersetzen oder ganz wegzulassen. Die in der DSGVO geforderte Datenminimierung kann dabei von einer SDPP auf dem Transportweg abhängig vom Ziel umgesetzt werden.

Es geht aber nicht nur um die Reduktion von Datenvolumen. Auch das Gegenteil kann manchmal sinnvoll sein. Zum Beispiel kann man im internen Netz Informationen zu Identitäten, Geolocation oder Daten aus einer Konfigurationsmanagementdatenbank (Configuration Management Database; CMDB) zu Logs hinzufügen, was später bei einem externen Dienstleister oder in einem SIEM in der Cloud vielleicht nicht mehr so einfach möglich ist.

Beim Transport Informationen hinzufügen

Auch Threat Intelligence kann von SDPPs direkt zu Events hinzugefügt werden. Die Enrichment-Features von SDPPs reichern meist Kontext an und verbessern damit die spätere Erkennung. Zu einer Logzeile von der ACL eines Routers, die zunächst nur auf eine blockierte TCP-Verbindung von einer internen Quelladresse zu einer externen Zieladresse hinweist, könnte die SDPP Geolocation, Benutzer und Threat Intelligence hinzufügen. Im Ergebnis enthält das Event dann neben den ursprünglichen Daten die Information, dass vom PC des Benutzers Müller am Standort München eine Verbindung zu einem bekannten C&C-Server in Russland blockiert wurde. Natürlich sind solche Enrichments auch in Form von Use Cases in einem SIEM umsetzbar, sofern alle Daten in das SIEM eingeliefert werden. In Anbetracht der Last und der Kos-

Listing: Entgegennahme, Filtern und Weiterleiten der Daten durch die Pipeline

```
accept_tcp "10.0.0.1:1514" {
  read_delimited "</Event>\n", include_separator=true
}
this = data.parse_winlog()
// Send the raw data to a blob store.
fork {
  to_s3 "s3://my-bucket/events/**/data_{uuid}.ndjson.zstd",
    // Produces objects under:
    //   s3://my-bucket/events/year=<year>/month=<month>/day=<day>↵
                                                    /data_<uuid>.ndjson.zstd
    partition_by=[year, month, day],
    max_size=100M {
    write_ndjson
    compress_zstd
  }
}
// Keep only Login/Logout Events.
where System.EventID in [4624, 4625, 4634]
// Forward security relevant data to a SIEM.
to_splunk "https://10.23.0.10:8088",
  hec_token=secret("splunk-hec-token")
```

ten, die dabei entstehen, ist es jedoch eine interessante Alternative, dies auf dem Transportweg in einer Pipeline zu tun.

Das Orchestrieren der einzelnen Funktionen für eine Pipeline kann je nach Produkt via Code geschrieben oder grafisch via UI zusammengeklickt werden. Der Anbieter Tenzir setzt beispielsweise auf Code, wie im Listingkasten zu sehen ist. Er zeigt, wie eine Pipeline die Windows-Event-Logs im XML-Format im VPN auf Port 1514 annimmt, die Daten in Hive Schema in einem S3-kompatiblen Blob Store ablegt sowie Login- und Logout-Events zu einem SIEM weiterleitet.

Die IT und insbesondere Cloud-Infrastrukturen und -Services entwickeln sich ständig und immer schneller weiter und führen so zu neuen Eventquellen und immer wieder geänderten Formaten. In Anbetracht dieser vielfältigen Datenquellen, die heute Logs, Events, Traces beziehungsweise Telemetrie oder auch Threat Intelligence in ein SOC schicken, ist das Detection Engineering eine Sisyphusarbeit geworden.

Spätestens für KI-Anwendungen im SOC ist eine Bereitstellung der Daten in einer normalisierten Form wichtig. Mit dem Open Cybersecurity Schema Framework (OCSF), dem Elastic Common Schema (ECS), dem Advanced Security Information Model (ASIM) von Microsoft, dem Common Information Model (CIM) von Splunk oder dem Unified Data Model (UDM) von Google stehen zahlreiche offene oder herstellerspezifische Standards zur Verfügung, die von vielen SDPP-Herstellern unterstützt werden.

Setzt ein Unternehmen weniger auf Cloud-Infrastruktur und möchte oder muss on Premises oder gar in einer Air-gapped-Umgebung arbeiten, sollte das SDPP-Produkt auch ohne Internetanbindung funktionieren. Die Anbieter gehen mit verschiedenen Strategien an das Thema heran: Einige bieten On-Premises- oder Hybridvarianten, andere bleiben jedoch bei „Cloud only".

Für alle Produkte interpretierbar

Gerade für Logs, die von der Quelle per Syslog als Text-String gesendet werden, ist das initiale Parsen wichtig. Weitere Normalisierung, die die Bedeutung eines Logeintrags von der Syntax abstrahiert, erleichtert es, später Erkennungsregeln zu formulieren, und macht so unabhängiger vom jeweiligen SIEM-Produkt. Eine fehlgeschlagene Authentisierung wird beispielsweise von Applikationen in unendlichen Formen protokolliert. Varian-

Anbieter von Security Data Pipeline Platforms

Firmen	Abstract Security	Axoflow	Beacon	Cribl	Databahn	NXLog	Tenzir
Wie viele Mitarbeiter hat Ihr Unternehmen?	>60	35	24	>1000	>75	>100	ca. 10
Gründungsjahr	2023	2022	2025	2018	2024	2014	2017
Firmensitz	USA	USA	Israel	USA	USA	Ungarn	Deutschland
Wofür wird KI eingesetzt?	Datenanalyse, Datenaufbereitung, Weiteres	Datenaufbereitung	Datenaufbereitung, Pipelineerstellung/ -management, Weiteres	Datenaufbereitung, Pipelineerstellung/ -management, Datenanalyse, Weiteres	Datenaufbereitung, Pipelineerstellung/ -management, Datenanalyse, Weiteres	Roadmap	Datenaufbereitung, Pipelineerstellung/ -management, Weiteres
Ist es möglich, ein eigenes LLM/RAG an das Produkt anzubinden (z. B. via MCP-Server)?	✓	–	✓	✓	✓	Roadmap	✓
Deployment-Optionen	SaaS, Self-hosted Cloud, on Premises (Roadmap), air-gapped (Roadmap)	on Premises, hybrid, SaaS, air-gapped	SaaS, hybrid	on Premises, hybrid, SaaS, air-gapped	on Premises, Self-hosted Cloud, hybrid, SaaS, air-gapped	on Premises, SaaS (Roadmap), air-gapped (Roadmap), hybrid (Roadmap)	on Premises, Self-hosted Cloud, hybrid, SaaS, air-gapped
Programmiersprache der leistungsintensiven Komponente	Go, Python	C, Go	k. A.	TypeScript (Node.js)	Go, Rust, Python, Java	C/C++, Rust	C++, Rust, Python
Eigene Storage-Lösung (optional) vorhanden?	✓	✓	–	✓	–	✓	✓
native Konnektoren/Parser	280+	160	500+	70+	500+	50+	50+
Wie wird die Pipeline gebaut? UI? Code?	UI, Code	UI, Code (nur Enterprise User)	UI, Code	UI, Code	UI	UI, Code	UI, Code
Ist es möglich, Daten zu pseudonymisieren?	✓	✓	k. A.	✓	✓	Roadmap, Workaround vorhanden	✓
Ist Data Enrichment möglich?	statisch, dynamisch	statisch, dynamisch	statisch, dynamisch	statisch, dynamisch	statisch, dynamisch	statisch, dynamisch	statisch, dynamisch
Gibt es eine Abfragesprache oder ähnliche Features, um Logdaten zu durchsuchen?	✓	✓	–	✓	✓	–	✓
nativ unterstützte normalisierte Formate	OCSF, UDM, ECS, CEF, Custom	Dynatrace, ECS, OCSF, UDM, ASIM, Custom	ECS, OCSF, ASIM, Custom	OCSF, UDM, ECS, ASIM, CIM, CEF, LEEF, Custom	Plain Text, JSON, XML, Key-Value-Paare, CEF, LEEF, OCSF, ASIM, CIM, ECS, UDM, Custom	UDM, ASIM, ECS, Prometheus, Custom	UDM, OCSF, ECS, Custom
Erkennt die Lösung automatisch Änderungen am Format (Schema Drift)?	Ingoing/Outgoing	Ingoing/Outgoing	Ingoing/Outgoing	Ingoing	Ingoing/Outgoing	Ingoing	Ingoing/Outgoing
Gibt es eingebaute Features zur Erkennung von Sicherheitsvorfällen?	Detection Rules (auch pre-built), Real-time Correlation, UEBA	–	–	händisch möglich	Detection Rules	händisch möglich	händisch möglich

Tabelle beruht auf den Angaben der Anbieter.

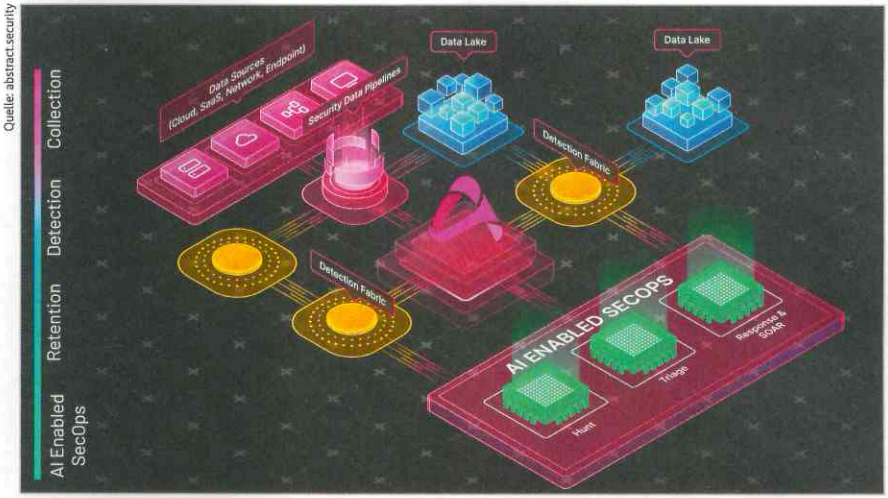
ten von „authentication failed“ bis „invalid login“ könnten nach einer Abstraktion ebenso gemeinsam weiterverarbeitet werden wie verschiedenste Firewall-Log-Varianten, die mit „accepted“, „permit“ oder „allow“ das Gleiche meinen.

Die Transformation beispielsweise in OCSF erhöht jedoch auch das Volumen stark. Je nach Ökosystem, in dem das jeweilige SOC sich bewegt, werden die einzelnen Standards oder auch eigene Formate relevanter sein. Eine moderne Security Data Pipeline Platform wird in diesem Kontext zu einer immer wichtigeren Komponente. Sie initiiert schon am Eingang des SOC das Parsen und die Transformation der Daten, die SIEM, SOAR und KI dann einfacher weiterverarbeiten können.

Oft ist es gewünscht, Logs und ähnliche Daten längerfristig zu speichern. Bei einem erkannten Sicherheitsvorfall beispielsweise können die ersten relevanten Spuren schon viele Monate zurückliegen. Forensiker benötigen dann zum Rekonstruieren des Hergangs umfangreiche Daten aus der Vergangenheit. SIEM-Systeme sind dafür meist zu teuer. Einige Hersteller von Pipelines haben das als Marktchance erkannt und Features eingebaut, um Daten auf einem günstigeren Speicher wie Amazon S3 oder Snowflake abzulegen und bei Bedarf entweder doch noch an das SIEM weiterzuschicken oder direkt über die Pipeline abzufragen. Für diese Abfrage implementieren sie entweder eine eigene Syntax, die der von gängigen SIEMs ähnelt, oder inzwischen auch natürlichsprachliche Abfragen an ein eingebautes LLM, das die Daten bereits bei ihrer Ankunft verarbeitet.

Eine klassische Architektur, die versucht, alle Daten in einem zentralen SIEM zu halten, erfordert ständig neue Kompromisse zwischen Verfügbarkeit wichtiger Daten, der langfristigen Speicherung und den anfallenden Kosten. SDPPs können hier eine Entlastung bringen, wobei ein günstigerer Speicher für die Langzeitarchivierung und für Forensik entsteht und das SIEM mehr auf seine Kernkompetenz in der Erkennung von Vorfällen ausgerichtet wird.

Quelle: abstract.security



Eine Security Data Pipeline Platform, hier vom Hersteller Abstract, besteht aus verschiedenen Bausteinen – vom Sammeln der Daten aus verschiedenen Quellen über das Normalisieren und die individuell benötigte Aufbereitung der Daten bis hin zum Speichern oder Weiterverarbeiten.

SDPP smart dazwischengeschaltet

Sowohl das SIEM als auch andere Zielsysteme werden dabei auch austauschbarer und so ergibt sich aus einer potenziellen SIEM-Migration ein weiterer Einsatz für SDPPs: SOCs, die von einem SIEM-Produkt zu einem anderen Anbieter oder zu einem Cloud-Service wechseln wollen, profitieren von der Einführung einer SDPP, die vorübergehend das alte und das neue SIEM gleichzeitig mit Daten versorgen kann und die Datenflüsse vom SIEM entkoppelt, strukturiert und überwacht.

KI spielt auch in der Security eine immer größere Rolle, auch im Kontext von SDPPs. KI, die im SOC eingesetzt werden soll und die Analysten bei der Erkennung und Bewertung von Vorfällen unterstützt, benötigt Daten in einer möglichst hohen Qualität und Konsistenz mit möglichst viel Kontext. SDPPs sind somit ein idealer Lieferant dafür. Aber auch innerhalb der SDPPs selbst kommt vermehrt KI zum Einsatz, zum Beispiel bei der Definition der Datenflüsse in einer Pipeline. Für Unternehmen mit hohen Anforderungen an die Informationssicherheit ist die Frage relevant, ob sie auf die gegebenenfalls in der Cloud betriebene KI des SDPP-Herstellers angewiesen sind oder ein eigenes LLM nutzen können.

Der Markt für SDPP ist seit 2025 stark in Bewegung. Große XDR-Anbieter wie CrowdStrike, SentinelOne und Palo Alto haben Anbieter von Pipelines aufgekauft und integrieren diese Produkte in ihr SIEM oder ihren Data Lake in der Cloud. Die XDR-Anbieter versuchen sich nach der Weiterentwicklung von EDR (Endpoint Detection and Response) zu XDR (Extended Detection and Response) nun auch als Alternative zu SIEM und SOC zu positionieren. Pipelines sind dabei ein wichtiger Baustein, um Zugriff auf möglichst viele Logquellen zu bekommen

iX-Workshop „Cyberabwehr mit Microsoft Defender XDR – Roll-out, Konfiguration, Nutzung“

Extended Detection and Response ist in modernen Sicherheitsstrategien unverzichtbar, um Bedrohungen frühzeitig zu erkennen und effektiv zu bekämpfen. Für einen reibungslosen Betrieb ist eine gelungene Integration in die bestehende Infrastruktur entscheidend.

Administratoren und IT-Sicherheitsexperten trainieren in praktischen Übungen, wie sie Microsoft Defender XDR für Clients und Server effizient bereitstellen und an individuelle Anforderungen in ihrem Unternehmen sinnvoll anpassen. Dazu blicken sie gemeinsam mit den Referenten Lukas Köglsperger und Pascal Schohn auf die unterschiedlichen Komponenten des Tools. Weiterhin lernen die Teilnehmer anhand realistischer Szenarien, wie sie Sicherheitsvorfälle richtig erkennen, sinnvoll bewerten und effektive Gegenmaßnahmen einleiten.

Anmeldung und Termine: heise.de/s/3xX6A

und die Koexistenz mit oder die Transition von eventuell schon vorhandenen Logspeichern zu vereinfachen.

Die Konkurrenz zwischen den XDR-Plattformen, aber auch zwischen ihnen und klassischen SIEM-Produkten und SOC-Services greift damit auch auf den Bereich der Logpipelines und der Data Lakes über. Auch SIEM-Anbieter stellen inzwischen mehrere Speicherebenen bereit. Ein günstigerer Speicher für Langzeitarchivierung vom selben Anbieter soll verhindern, dass die kurzfristig nicht benötigten Logs zu anderen Anbietern gelenkt werden und so Marktanteile verloren gehen.

Pipelineanbieter bauen immer mehr Erkennungsfeatures direkt in ihre Produkte ein, die dann einfache Indicators of Compromise (IoC) schon in der Pipeline erkennen oder mit YARA-Regeln abgleichen können. Das ist performanter als die Use Cases später im SIEM. Damit werden jedoch auch die Pipelines in Teilbereichen zur Konkurrenz der etablierten SIEMs. Manche Organisationen verzichten sogar schon auf ein klassisches SIEM und verlassen sich auf eine XDR-Suite und eine Pipeline mit Erkennungs-Features und günstigem Objektspeicher dahinter.

Weitere Anbieter von Datenpipelines

Im Folgenden finden sich weitere Anbieter, die für die Marktübersicht kontaktiert wurden. Zum Redaktionsschluss standen über ihre Produkte keine Detailinformationen zur Verfügung. Nicht alle der genannten Anbieter sind auf Security spezialisiert, hier gilt es, im Zweifelsfall nachzufragen.

- CETU
- CrowdStrike (Onum)
- Databricks
- Datadog (Vector)
- Edge Delta
- Fivetran
- Fleak Tech
- Google
- Hevo Data
- Honeycomb
- IBM
- Integrate.io
- LimaCharlie
- LogSeam
- Elastic
- Monad (Tarsal)
- Panther (Datable)
- REALM. Security
- Sawmills
- SentinelOne (Observo.ai)
- Snare
- Ziggiz

Wird das SIEM in Zukunft überflüssig?

Auch wenn solche Architekturen für größere Unternehmen heute noch abwegig erscheinen, kann dies in der Zukunft, wenn autonome KI-SOC-Agenten einen größeren Anteil der Aufgaben übernehmen, dazu führen, dass KI zusammen mit Pipelines und Data Lakes ein klassisches SIEM überflüssig macht. Schon heute sprechen manche Hersteller von ihren Produkten als „Security Data Fabrics“.

Neben den SDPPs existiert mit ETL-Datenpipelines (Extrahieren, Transformieren, Laden) eine ähnliche und teilweise überlappende Kategorie an Produkten, die jedoch auf Daten für Business Intelligence ausgerichtet sind. Typischerweise fehlen ihnen die Features für Erkennung und Normalisierung im Securitykontext. Neben den Kernaufgaben ist eine ihrer wesentlichen Funktionen, fehlerhafte Daten zu erkennen – sowohl bei eingelieferten Daten als auch beim Schema, das das Ausgabeformat definiert.

Hersteller gehen das Problem korrupter Daten mit verschiedensten Überwachungen an, beispielsweise von Netzwerkparametern, Logvolumen oder Schema Drifts. Natürlich können diese Metadaten, ganz im Sinne der Logpipelines, verarbeitet und an verschiedene Ziele transportiert werden.

Produkte am Markt

Für diesen Marktüberblick (siehe Tabelle) hat die Redaktion zahlreiche Anbieter mit einem umfangreichen Fragenkatalog angeschrieben. In der Tabelle berücksichtigt sind diejenigen, deren Produkte auf die Bedürfnisse in einem SOC beziehungsweise der IT-Sicherheit ausgerichtet sind und die den Fragebogen ausgefüllt haben. Weitere Anbieter aus dem Themenumfeld sind im Kasten aufgeführt.

Die Tabelle enthält zunächst Basisdaten zum Hersteller selbst, beispielsweise Größe oder Herkunft der Firma. Nicht dargestellt sind Kernfeatures, die in allen Produkten vorhanden sind (zum Beispiel die Möglichkeit, Daten zu transformieren).

KI-Features sind ständig in Entwicklung und sehr dynamisch. Sie wurden in Kategorien zusammengefasst. „Datenaufbereitung“ umfasst etwa Features wie Parsergenerierung, Datenbanknormalisierung und Datenklassifizierung. Die Kategorie „Pipelineerstellung/-management“ bedeutet, dass Pipelines teilweise oder komplett per KI erstellt und/oder angepasst werden können.

Unterstützt ein Produkt statisches Enrichment, können statische Daten, beispielsweise Geodaten, mit in den Datensatz aufgenommen werden. Ist dynamisches Enrichment unterstützt, können Daten auch durch Informationen aus anderen Tabellen angereichert werden. Auf Basis dieser Übersichtstabelle kann möglicherweise noch keine Entscheidung für ein Produkt getroffen werden, sie kann jedoch bei der ersten Orientierung hilfreich sein. (ur@ix.de)

PATRICK BÄUERLE

ist Berater bei der Firma cirosec. Er führt Penetrationstests durch, ist im DFIR-Team und berät Kunden bei der Auswahl von Dienstleistern und Produkten.

STEFAN STROBEL

ist Buchautor sowie Gründer und Geschäftsführer des IT-Sicherheitshauses cirosec.