



The bridge to possible

AI-Driven Cloud-Native Threat Detection in Realtime

Luca Muscariello, Ph.D. Principal Engineer
BRKNWT-2416

CISCO *Live!*

#CiscoLive

Cisco Webex App

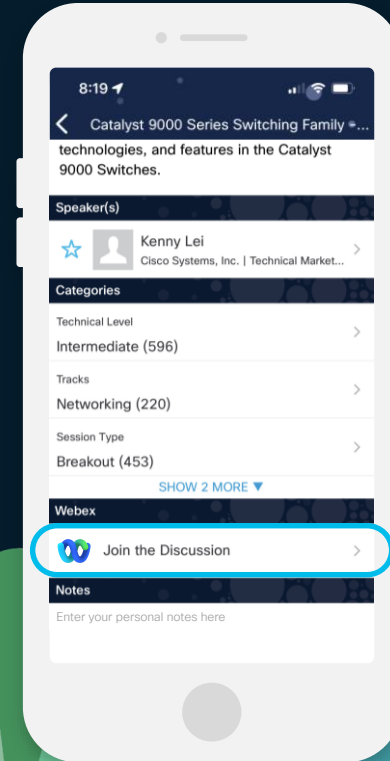
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 7, 2024.





Agenda

- Cloud Detection & Response
- AI-Driven Purple Teaming
 - Formulate Threat Hypothesis
 - Observe Facts and Enrich Data
 - Adversary Emulation
 - Detect Cyber Threats
- Realtime CDR
- Conclusion



*“Cloud-based Cyber-attacks increased by **154%** in 2024 compared to 2023.”*

Checkpoint Research, 2024

Sysdig, AWS Reinvent 2023



*“**10 minutes** – that’s all it takes to execute an attack in the cloud after discovering an exploitable target*



*“The global average cost of a data breach in 2023 was **\$4.45 million**, +15% over three years, highlighting the growing financial burden on organizations*

Mean Time to Detect a Breach today

204
days

is the average time to
identify a security
breach in 2023
(MTTI)

73
days

is the average time to
contain a security
breach in 2023
(MTTC)

Time to identify and contain the breach

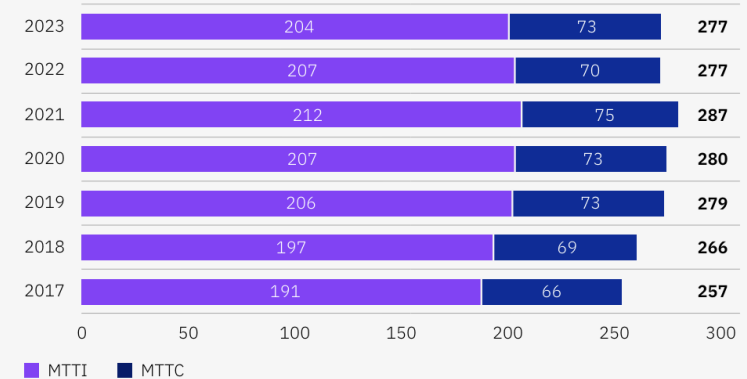
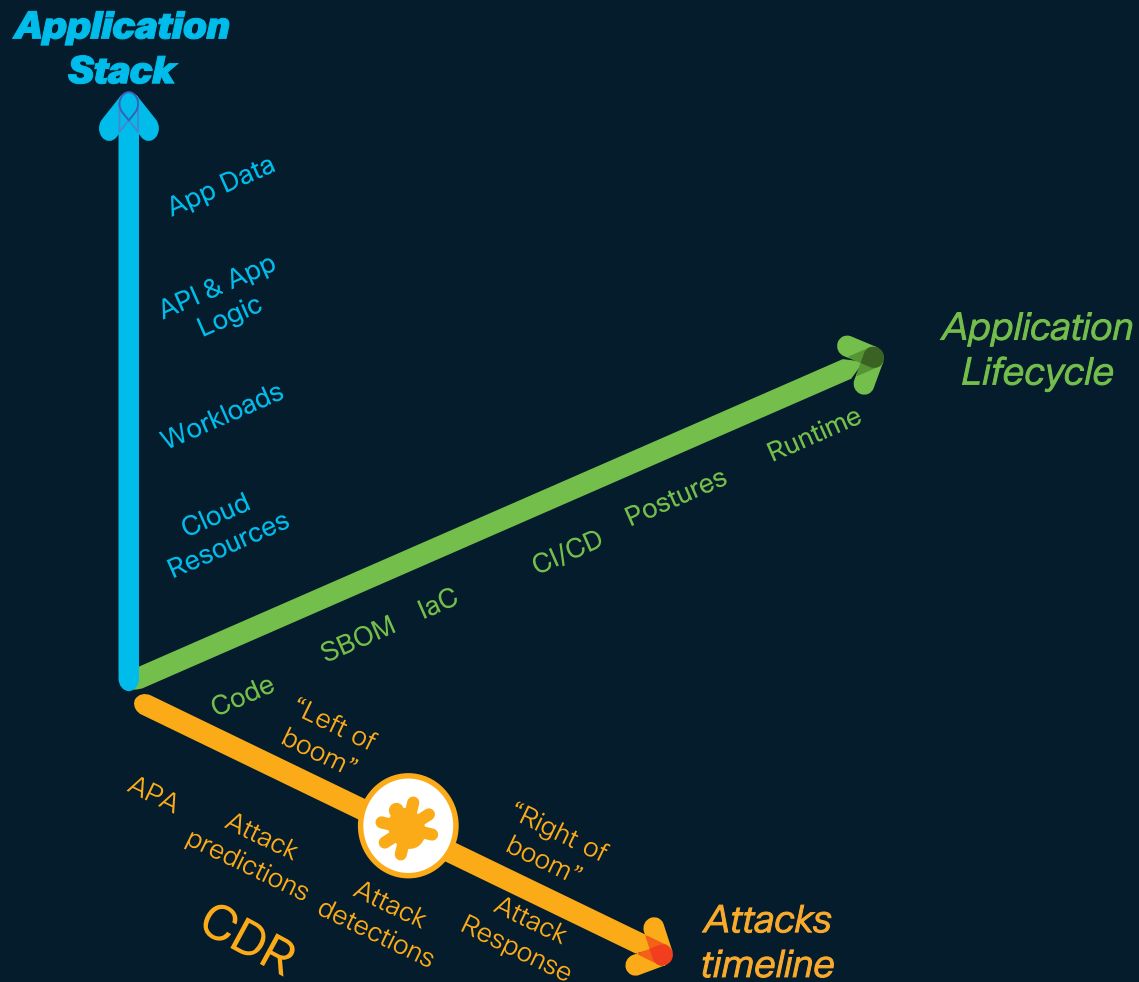


Figure 5. Measured in days



What if we could lower detection
time to ~~minutes~~ seconds?

Application Security Dimensions



Why CDR

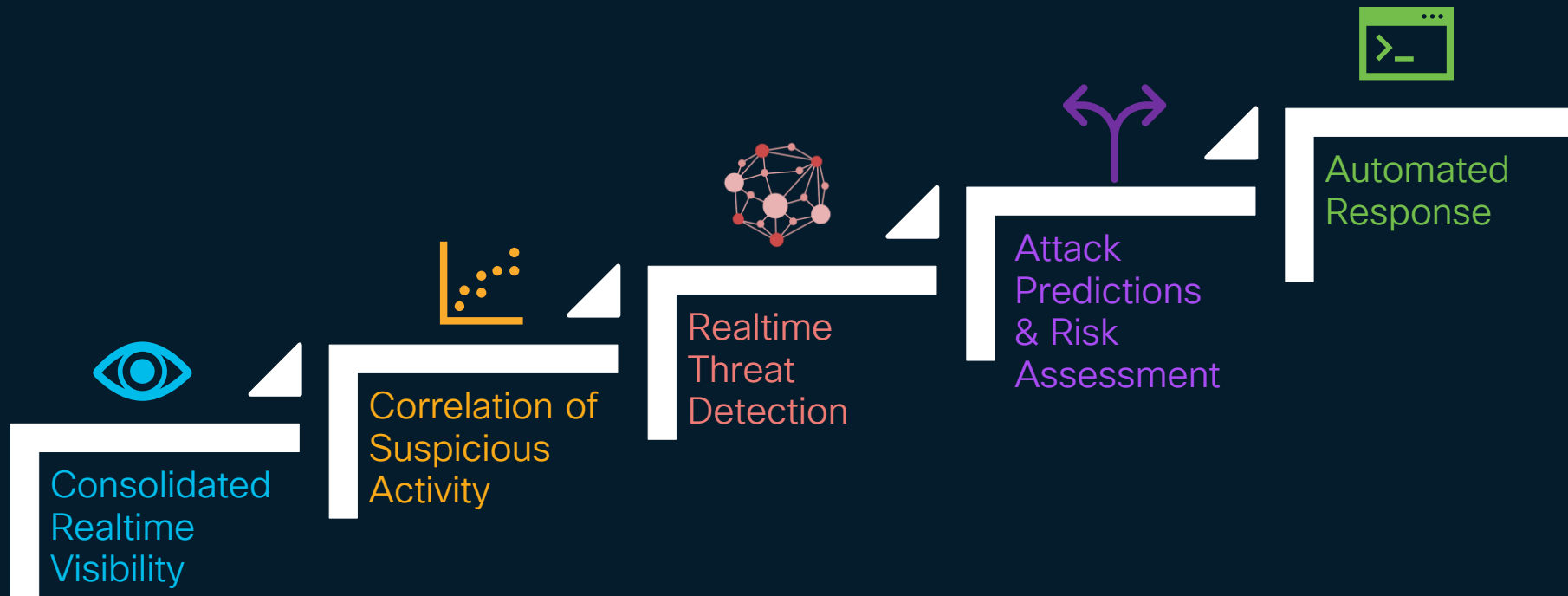
CDR does not replace traditional cloud application security:

CIEM/CSPM are like locks on the doors \leftrightarrow CDR is just like video cameras inside the house providing a live view of what is happening inside, with additional intelligence to:

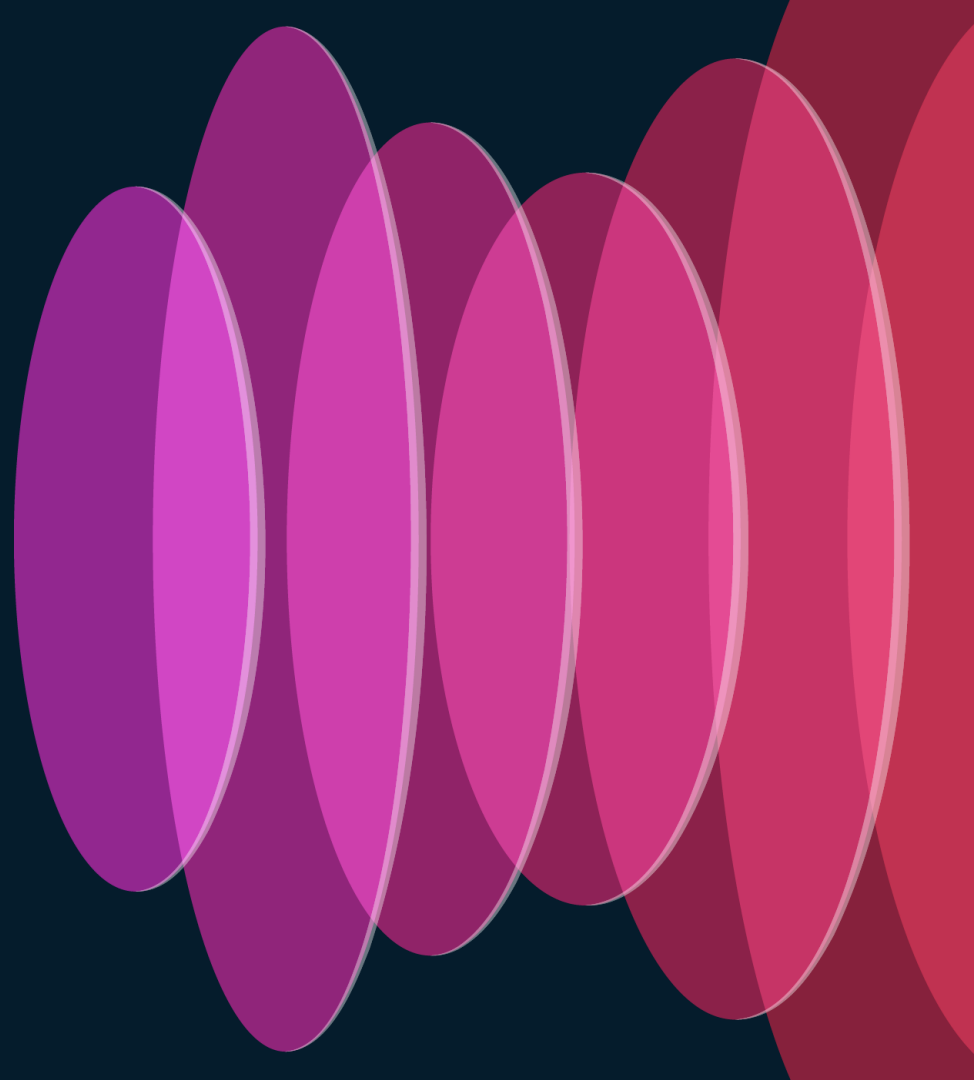


- Identify suspicious activity
- Provide a stream of security events as a forensic proof of threat detection
- Understand the context of what the attacker is doing in the environment
- Predict the next move

CDR Objective



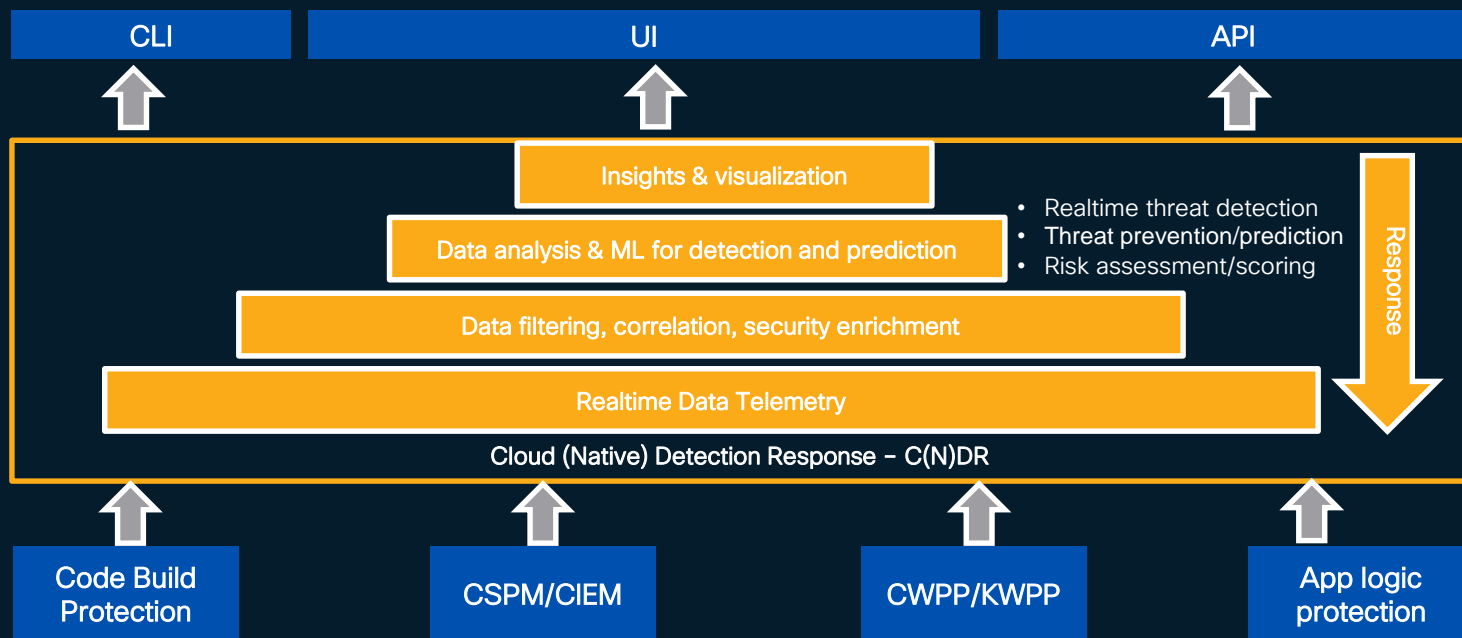
Realtime CDR



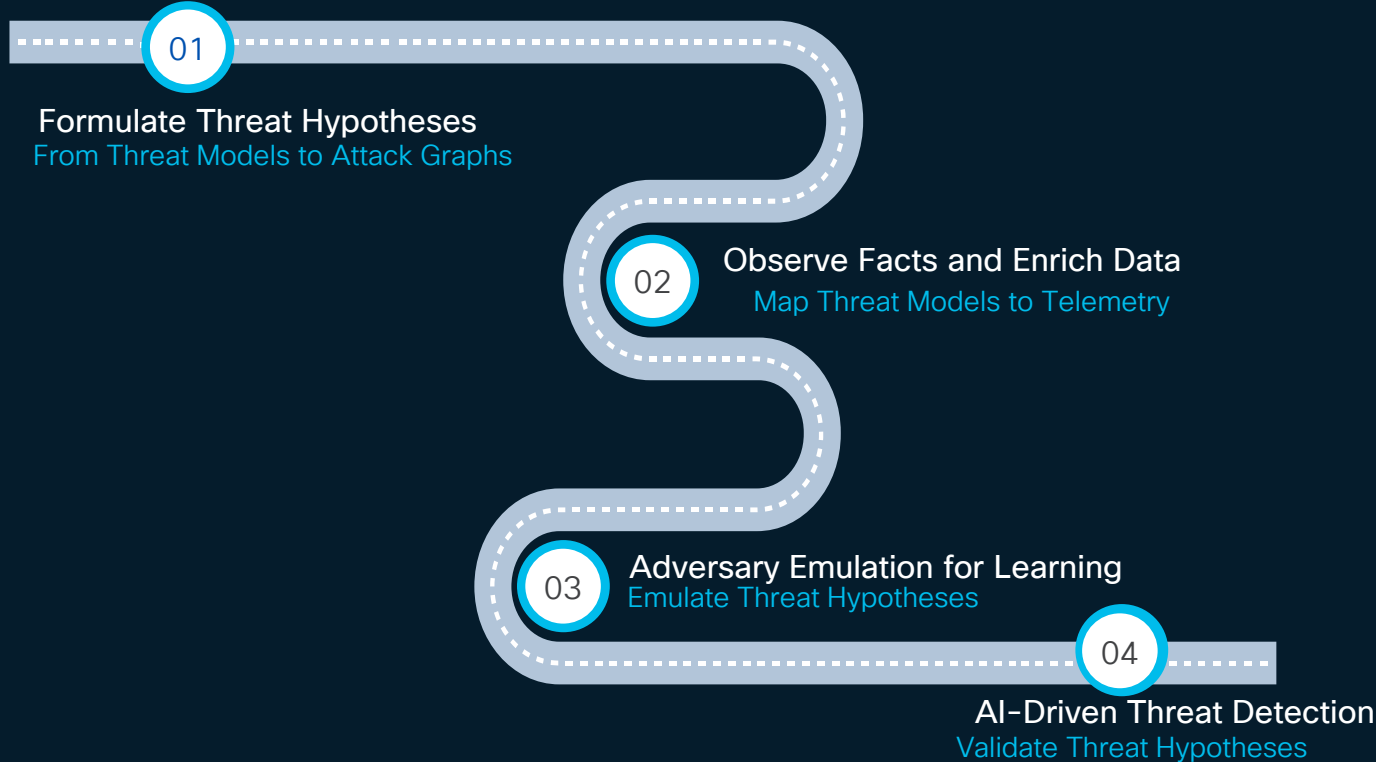
A CNAPP capability

“...many customers are starting to see detection and response as a first-class citizen within CNAPP. Their need is starting to expand beyond just workload runtime security and address the cloud control plane (via analysing cloud logs) to detect suspicious activity across users and services. ”

Sysdig [blog](#), March 17, 2023



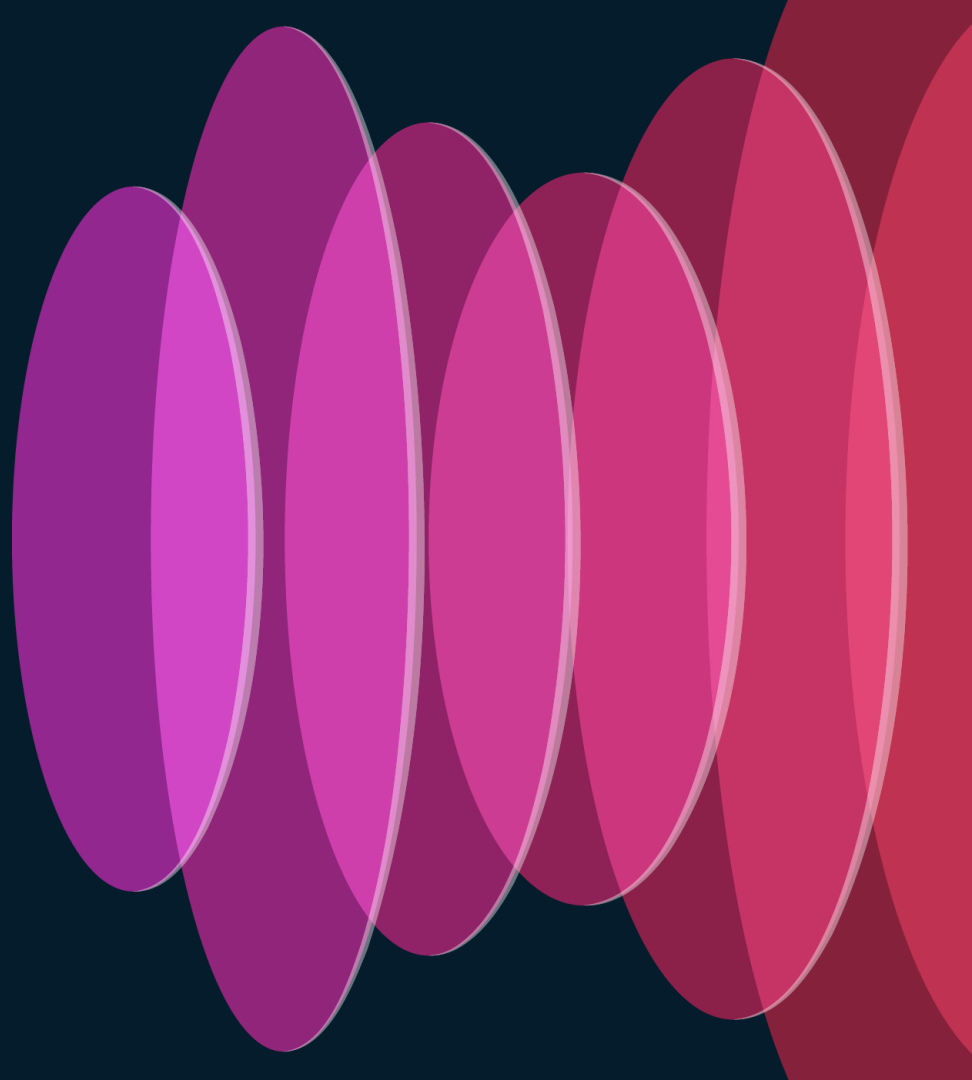
Purple Teaming Approach Overview



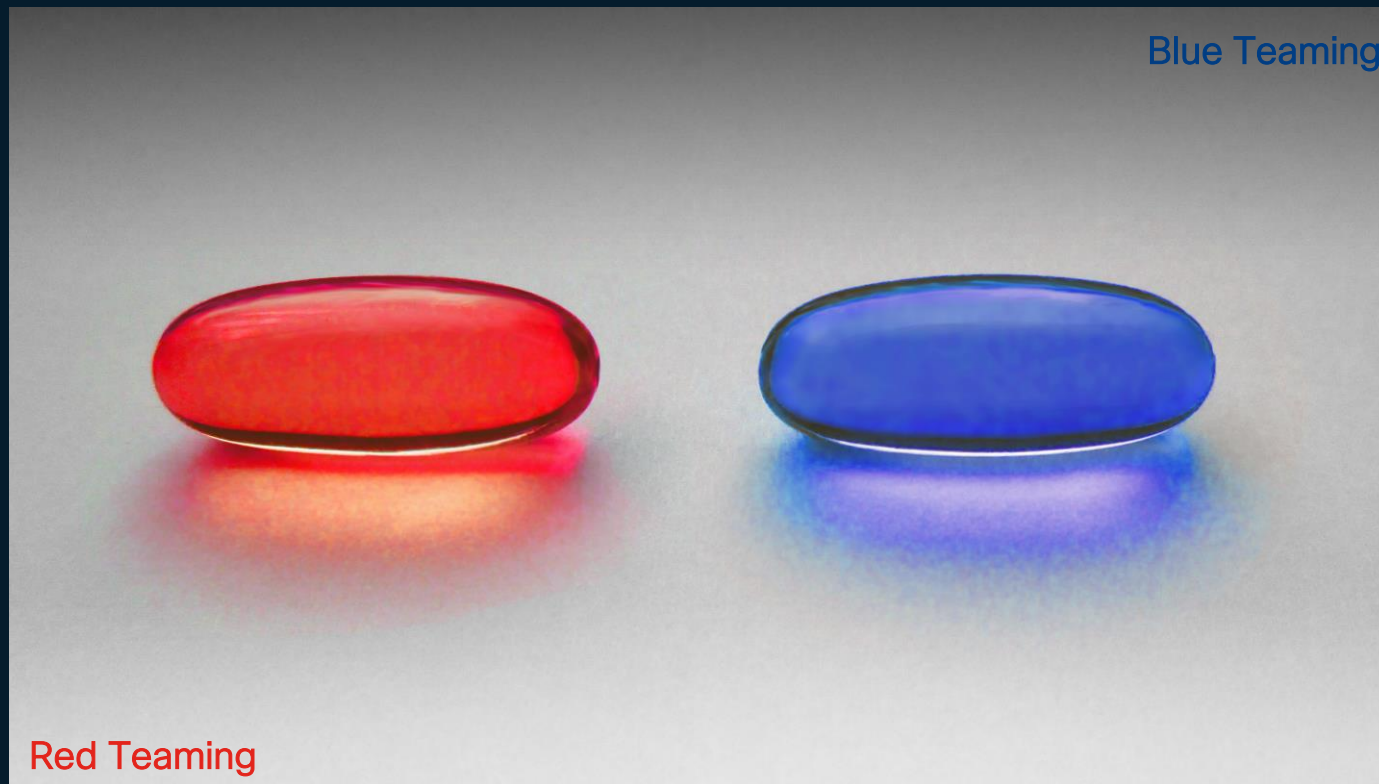
Purple teaming methodology

- Broadly speaking it's the collaboration of red and blue teams
- In this context it's a methodology to train an autonomous agent to formulate threat hypothesis and learn how to detect them based on observations
- Also used to optimize telemetry to include observables required to detect threats and attacks

AI-Driven Purple Teaming



Purple teaming methodology



Training in an emulated environment

Know all the attack techniques first and use that knowledge to defend



We train in an emulated environment:
reproducible, repeatable, ethical

Mapping Emulation to a Real System

- We use a digital twin of a realistic Cloud environment
- Improve accuracy of the emulated world
- Keep the emulation scalable



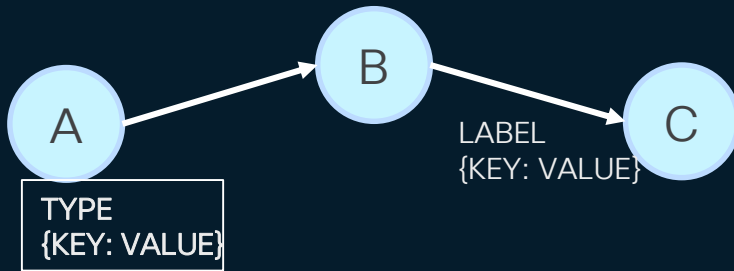
Formulate Threat Hypotheses



Knowledge representation and reasoning

- Best hypotheses can be formulated if all the knowledge of the system is represented in efficient data structures to perform reasoning tasks
- Graphs can be used to model how entities, objects or events are semantically interrelated.
- Graphs are powerful models to store knowledge and compute tasks

A Property Graph of the System under Attack

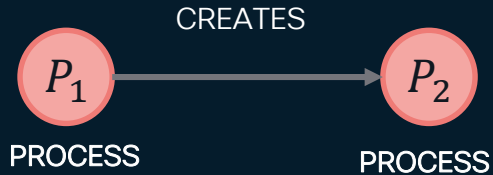


- Nodes
 - Entities with name and type
 - Network flows
 - Processes
 - Files
 - Containers
 - Hosts, Clusters etc.
 - With properties
 - Key-value pairs
- Edges
 - Relationships with a name and type
 - Are directed

Provenance and Context Graphs

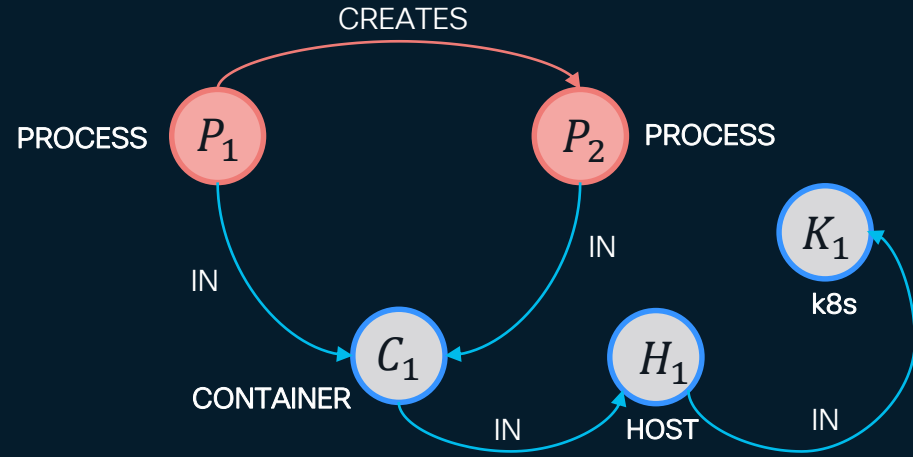
Provenance

- It models activities in the system
- Causality



Context

- System relationship among objects
- Models long term relationship



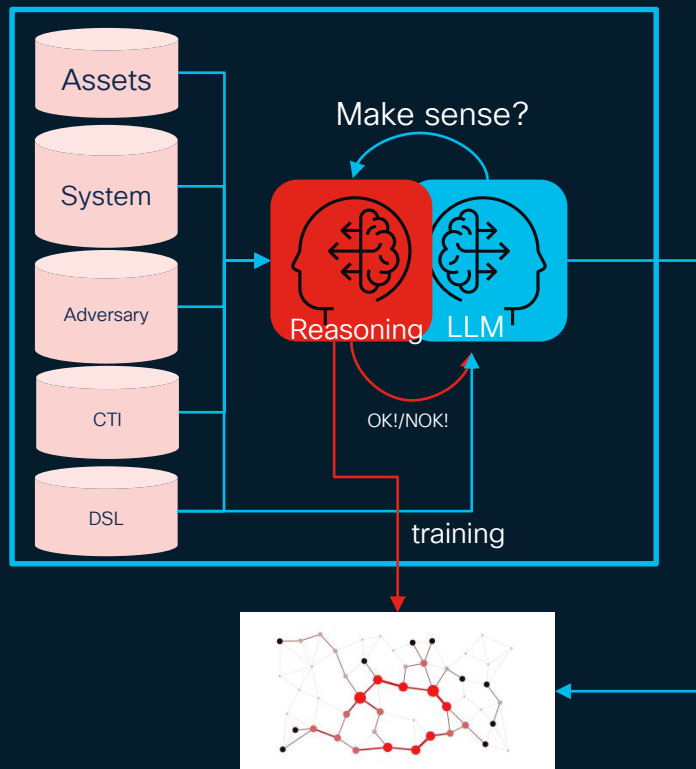
Attack Graphs Generation via DSL Language

- Formulate cyberthreat hypotheses for specific domains (public cloud accounts, private data centres or an enterprise network) and generate attack graphs
- We adopt a meta-attack language (MAL) as a meta domain specific language to specify the threat model of the system S_1
- The DSL language describes assets, events, adversaries of the specific MITRE ATT&CK matrix (tactics/techniques/procedures) and other cyberthreat intelligence.
- Nodes can be assets such as containers, applications etc.
- Edges can be adversaries such as TTPs

```
asset ApiKubectl {  
    // get pods  
    * discovery[1][uniform(20,1)]  
    * execution[1][lognormvariate(2,1)]  
    -> End  
}  
  
asset ContainerClientSide inherits Attacker {  
    * discovery[1][lognormvariate(2,1)]  
    -> End  
}  
  
asset Container {  
    * discovery[1][uniform(20,1)]  
}  
  
asset JuiceShopContainer extends ContainerClientSide {  
    // SSTi  
    fe4445d2-e878-484f-8f85-980b39de0cb5[2][lognormvariate(3,1)]  
}  
  
asset Shell {  
    * discovery[1][uniform(20,1)]  
    * collection[1][uniform(20,1)]  
    * execution[1][uniform(20,1)]  
    -> End  
}
```

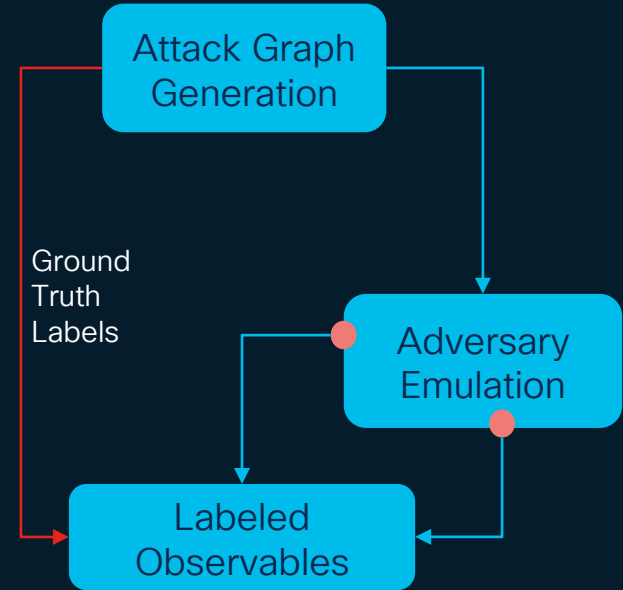
Formulate Threat Hypthesis with Probabilistic Attack Graphs Generation

- Hypotheses can be generated in two ways to produce attack graphs
 - Graph traversal based: it requires defining constraints and an optimization objective to generate focused hypotheses, e.g. “Data Destruction”.
 - LLM based: acts as meta-heuristic that can quickly provide hypothesis to validate.
 - Hypothesis can be verified efficiently.



Adversary Emulation and Observables

- Attack graphs are not observable as they are instances of attack hypotheses
- They can be used to execute an attack on an emulated system
- Observables can be extracted from the emulated systems with different telemetry sensors
- Emulation is powerful and yet expensive and hard to scale.



Current Focus

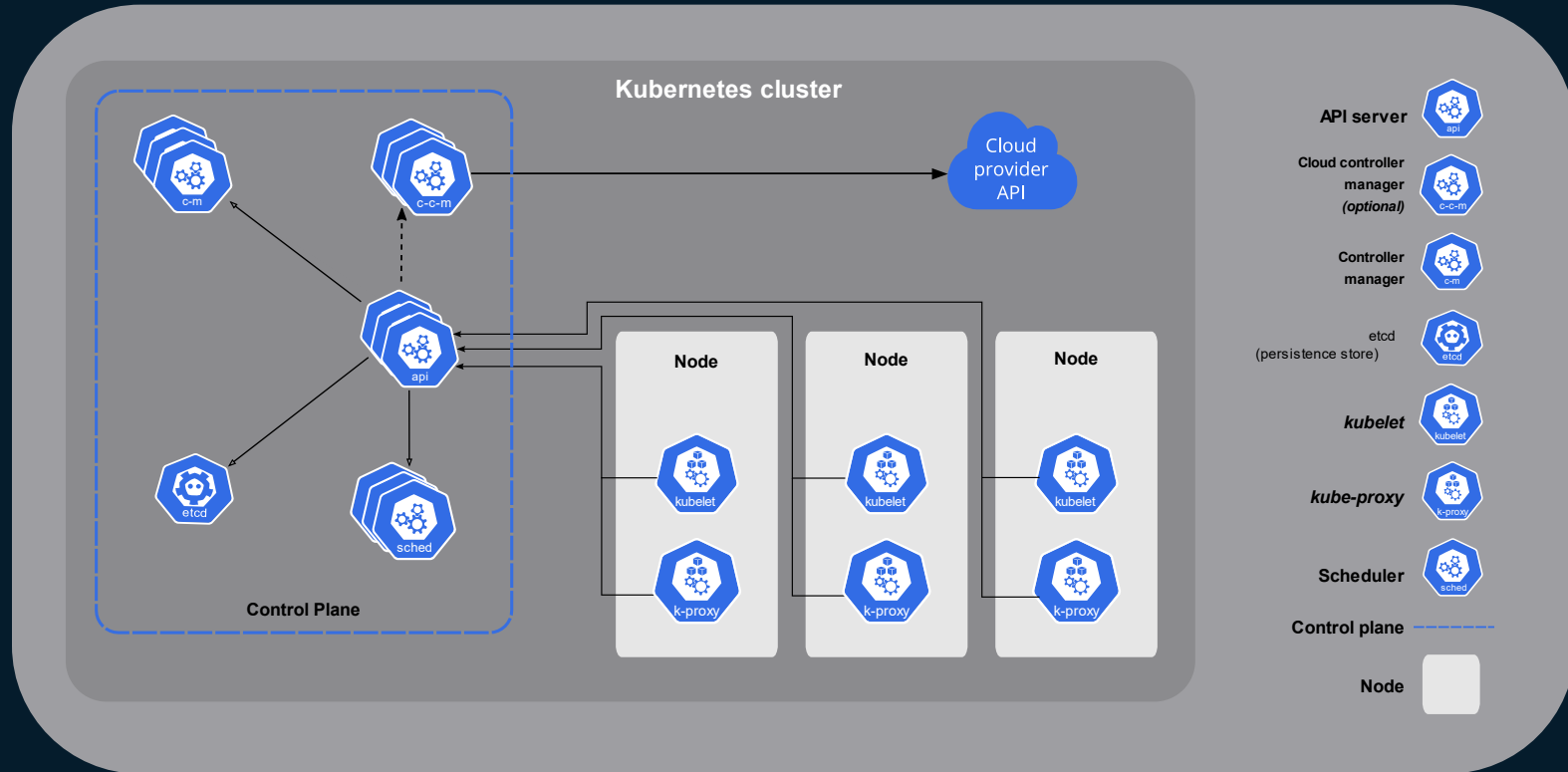
MITRE ATTA&CK Threat Matrix for Kubernetes

Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Impact
Using Cloud credentials	Exec into container	Backdoor container	Privileged container	Clear container logs	List K8S secrets	Access the K8S API server	Access cloud resources	Data Destruction
Compromised images in registry	bash/cmd inside container	Writable hostPath mount	Cluster-admin binding	Delete K8S events	Mount service principal	Access Kubelet API	Container service account	Resource Hijacking
Kubeconfig file	New container	Kubernetes CronJob	hostPath mount	Pod / container name similarity	Access container service account	Network mapping	Cluster internal networking	Denial of service
Application vulnerability	Application exploit (RCE)		Access cloud resources	Connect from Proxy server	Applications credentials in configuration files	Access Kubernetes dashboard	Applications credentials in configuration files	
Exposed Dashboard	SSH server running inside container					Instance Metadata API	Writable volume mounts on the host	
							Access Kubernetes dashboard	
							Access tiller endpoint	

Observe Facts and Enrich Data



Observability for Cloud-Native Applications



Realtime observables in Kubernetes

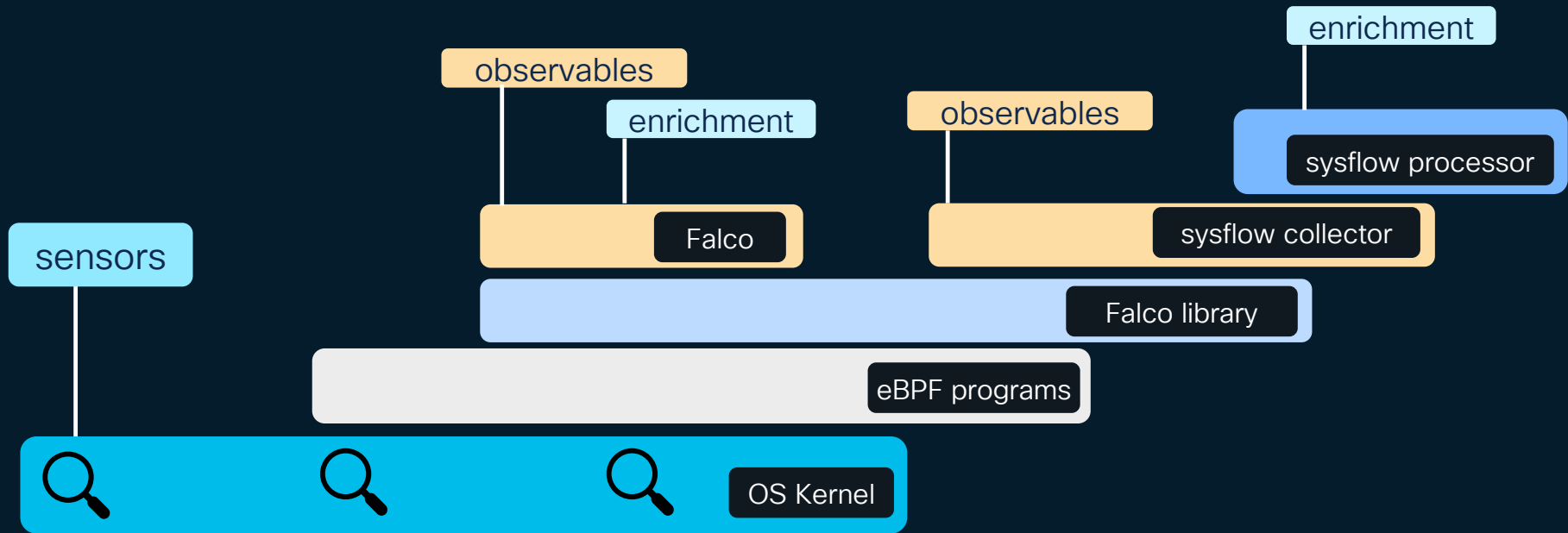
- Kubernetes can belong to a public cloud account (EKS, GKS, AKS etc) or to a private DC.
- We focus on system observables that can be collected from eBPF sensors from the host nodes of the cluster
- We focus on suspicious activities that can be recognized using a state-machine-based rule engine (e.g. falco) or a grammar-based engine (sysflow)
- Falco and Sysflow kernel system-calls based observables
- Kubernetes audits logs and API

Extended OCSF – Open Cybersecurity Schema Framework

- Created by AWS and Splunk to optimize ETL data and cybersecurity events
- A flexible and extensible framework
- We have adopted and extended OCSF to model knowledge graphs for reasoning using ETL data events and system data

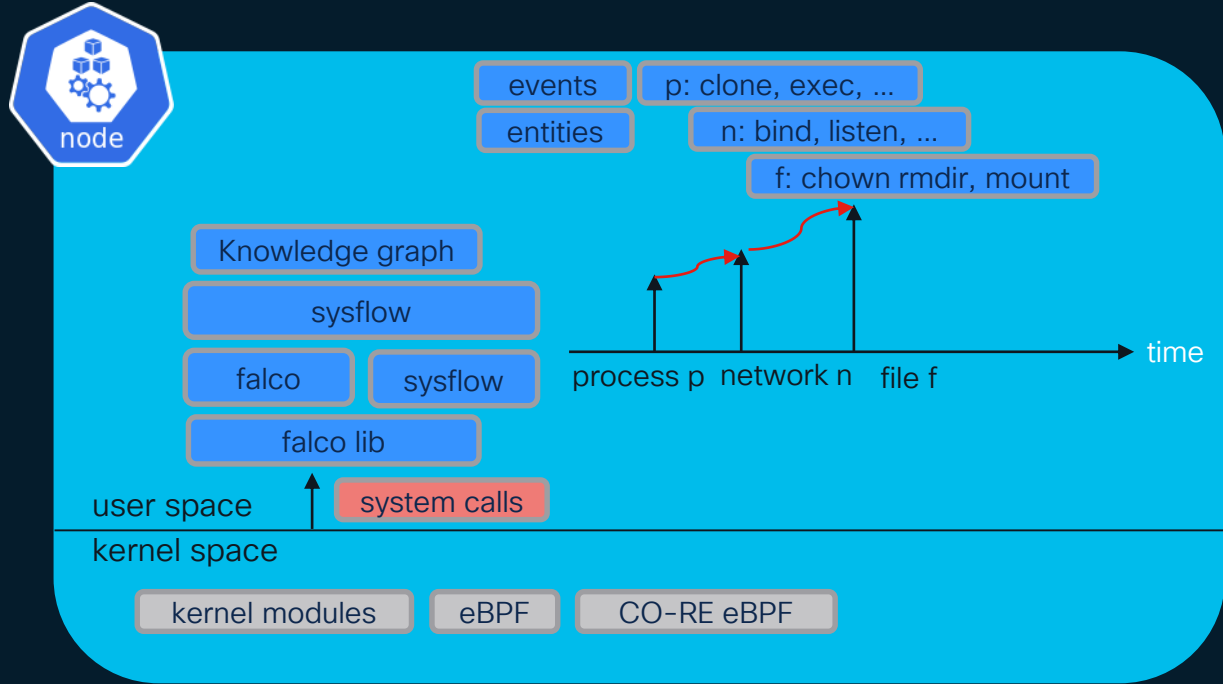


Linux Foundation CNCF/sysdig Falco and IBM sysflow

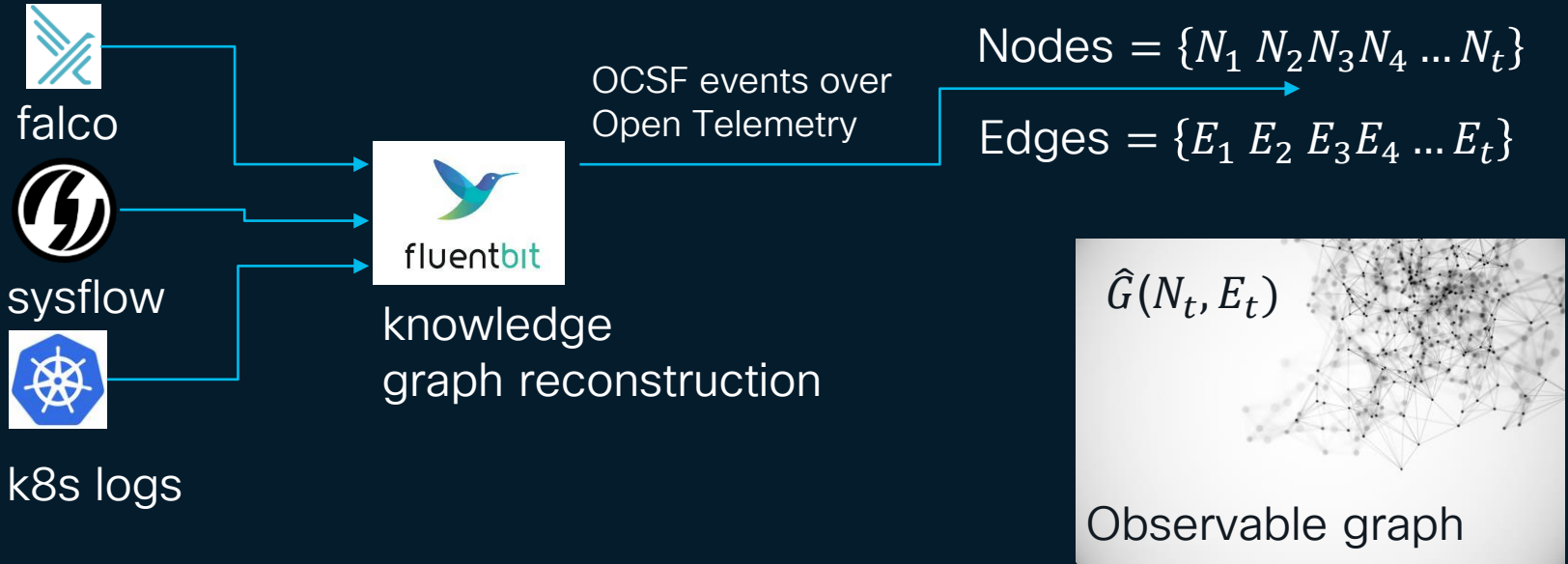


System Based Observability

- Falco extracts efficiently system calls from the node kernel
- Falco models the system activity via a state machine w/ entities and events such as process, network, file.
- sysflow models the system as a DSL language which recognizes events and activities to enrich data to create the provenance graph
- Fluent bit filters and merges information as graph nodes and edges to feed an ETL pipeline.



Knowledge Graph Reconstruction in Real-time



Training and Inference

The sampled graph $\hat{G}(N_t, E_t)$ constitutes the data that can be used for training the detection architecture

The richer the data the more efficient the detection will be

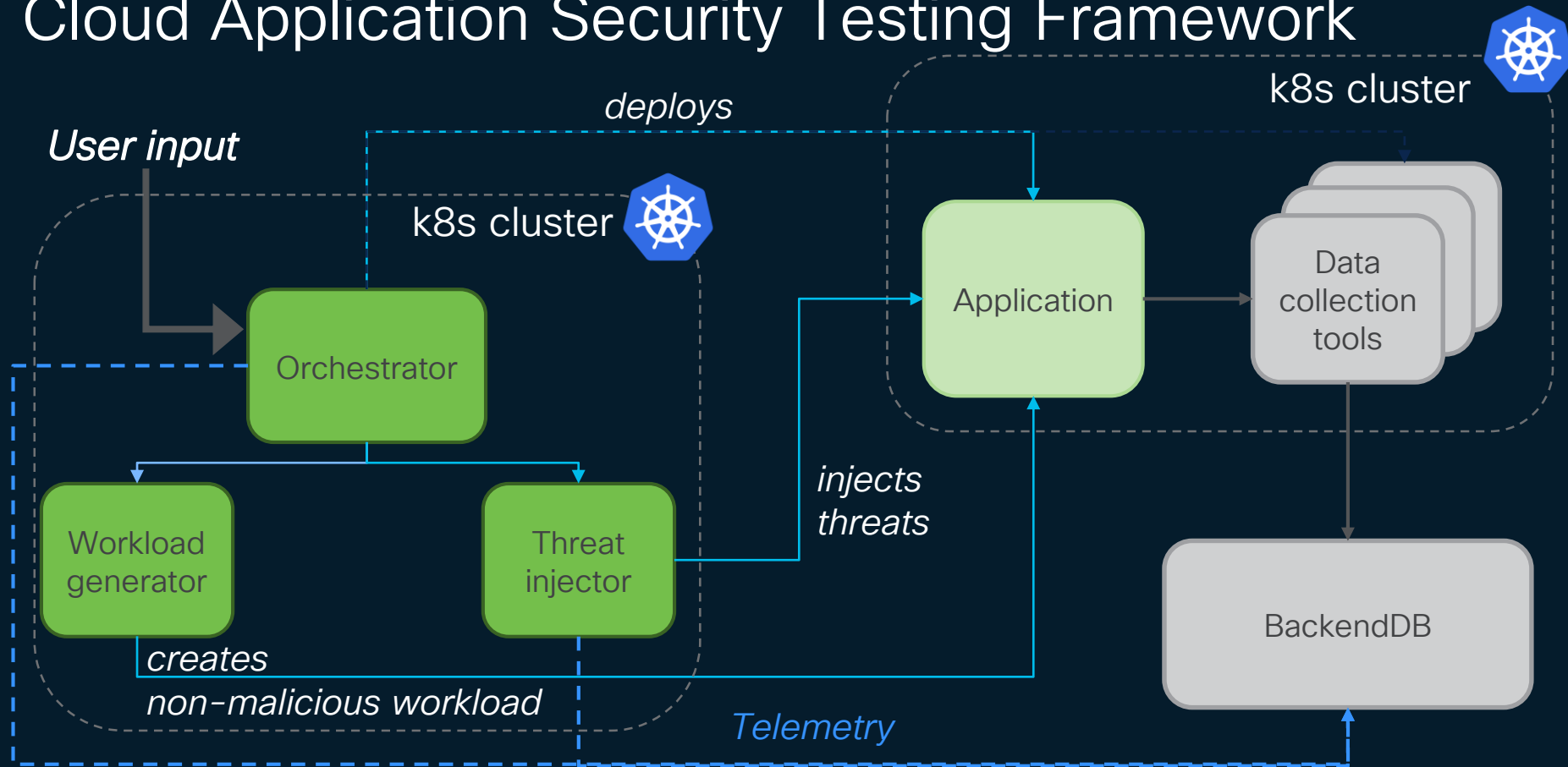
However, there are trade-offs to consider at inference time as the cost of extracting observable data is not zero

Optimal choices about the way $\hat{G}(N_t, E_t)$ is sampled requires extensive training and fine tuning

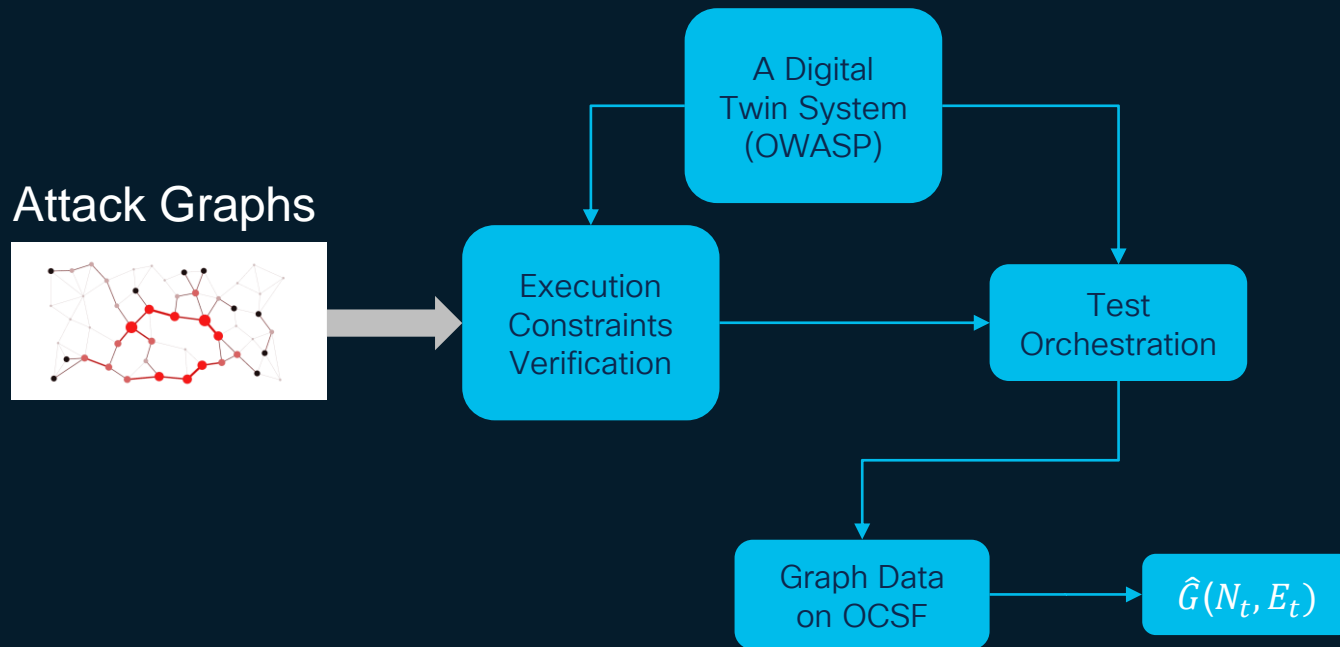
Adversary Emulation



Cloud Application Security Testing Framework



Executing Threat Hypotheses w/ Adversary Emulation



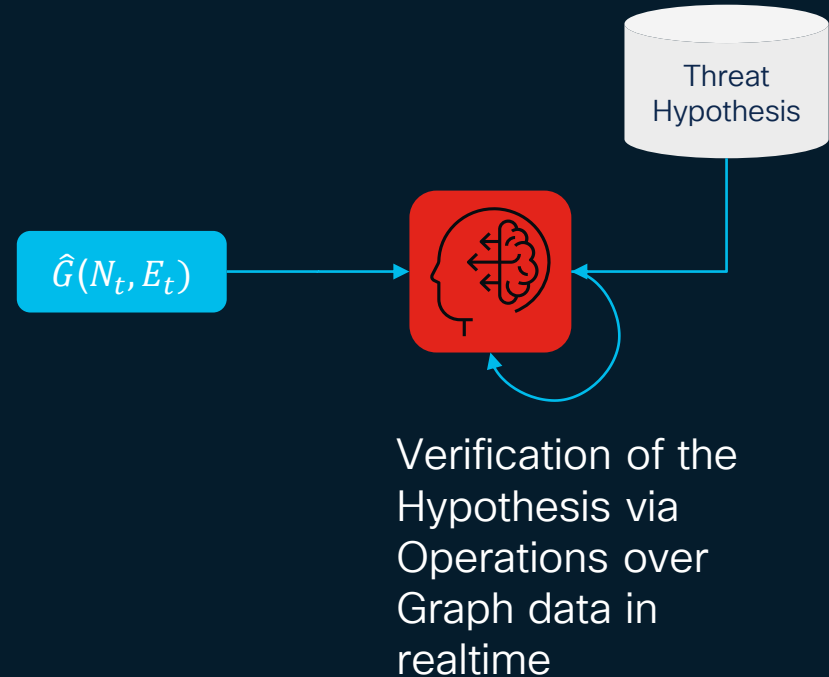
Detect Cyber Threats

CISCO *Live!*



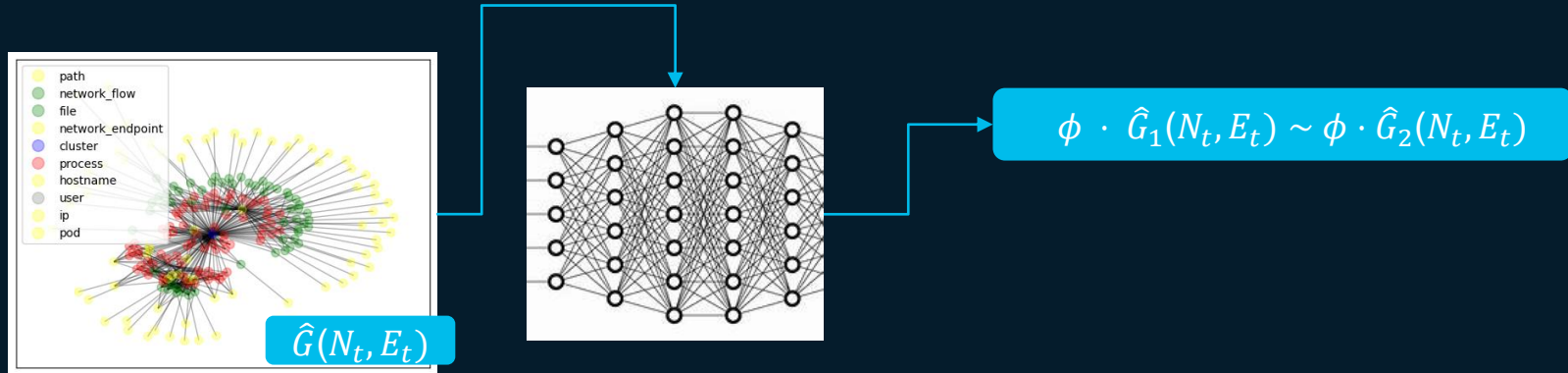
Threat Hunting

- Threat hunting is performed over the graph data $\hat{G}(N_t, E_t)$ in OCSF format that is sampled from the sensors in real-time
- We use ArangoDB as graph DB to perform threat hunting queries to verify threat hypothesis that the system has learned to perform off-line

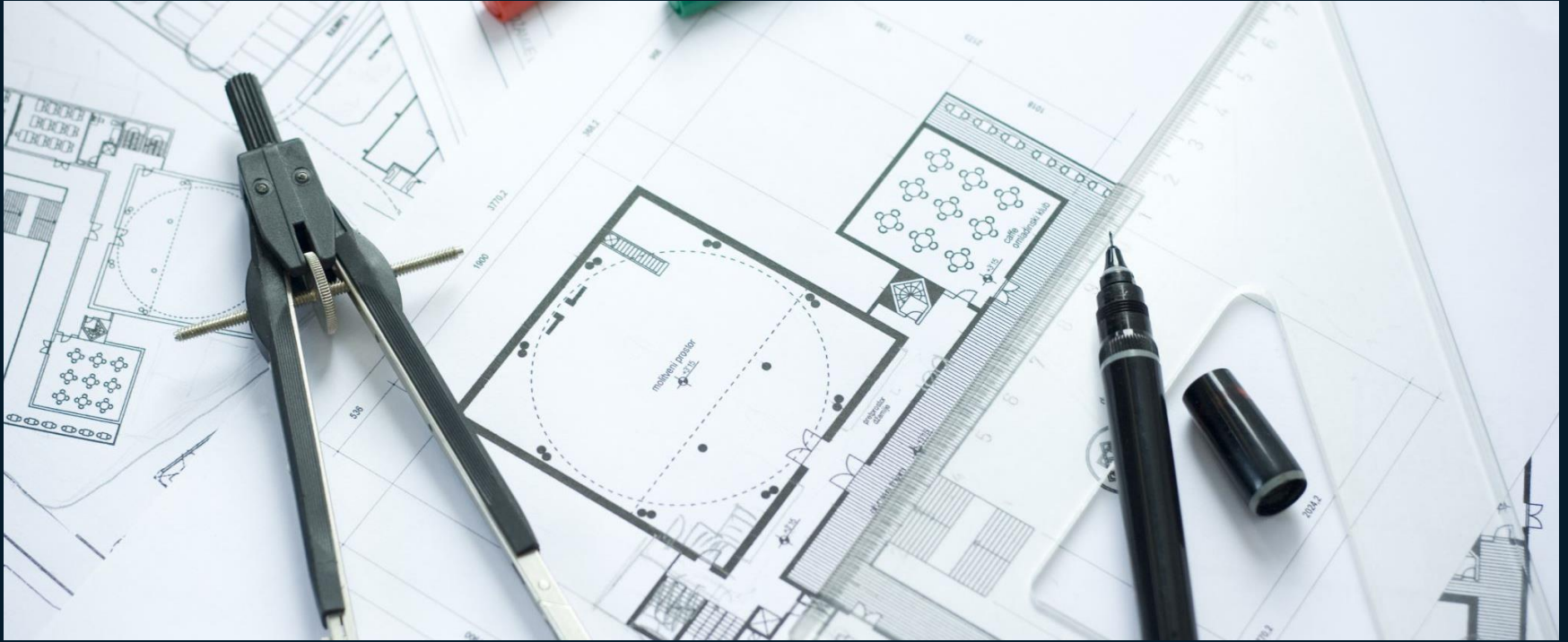


Scalable Detection via Graph Neural Networks

- Automatic Threat hunting remains the formal framework we have developed to verify detections
- Verification of 100% of the sampled graphs in real-time can be costly or generate delay in the response
- We leverage graph embeddings as representations of the sampled graphs to scale up detection computations in real-time
- The core of the detection architecture is based on Graph Matching Network (GMN) embeddings
- Accurate embeddings can be used to perform prediction on detection, risk scoring and more.

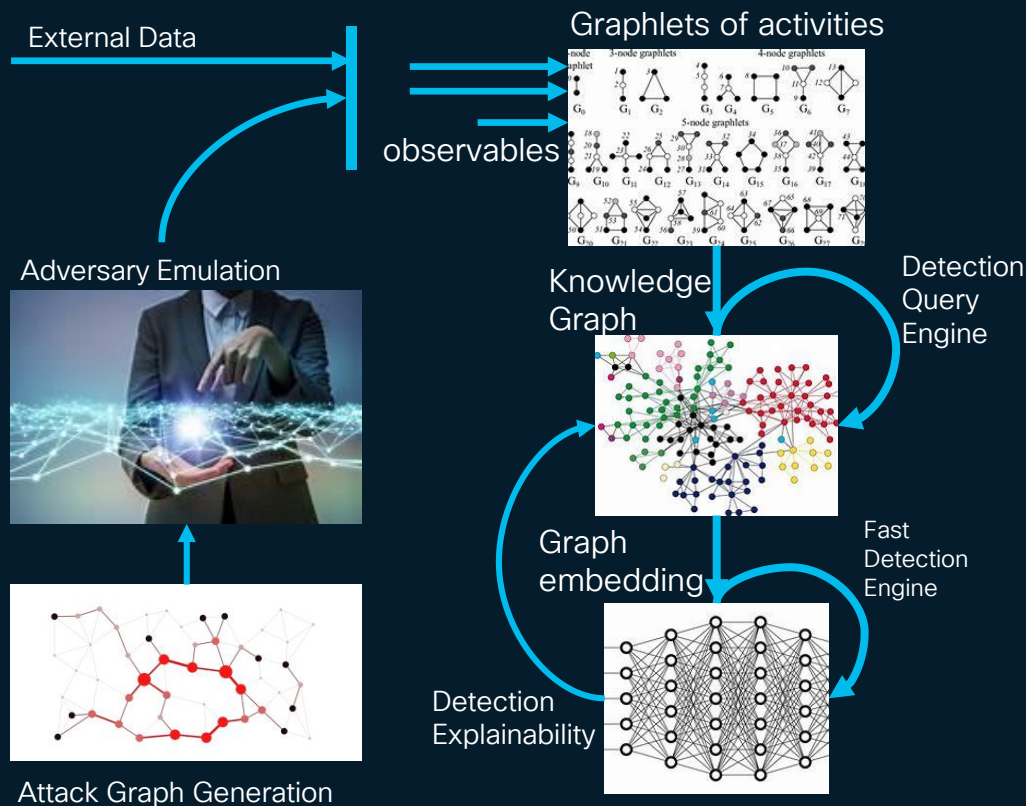


Architecture Summary

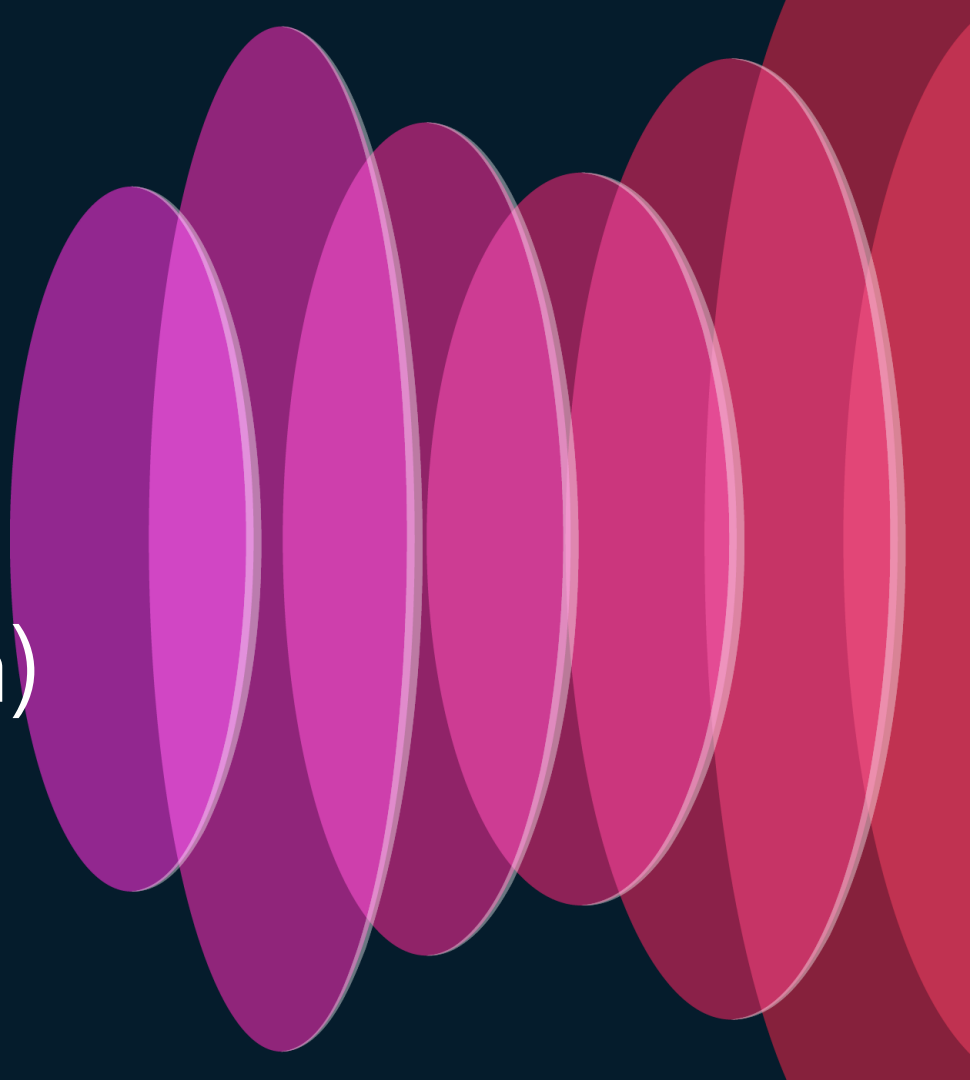


AI-driven purple teaming framework

- **A purple team** simulates malicious attacks and penetration testing to identify vulnerabilities and recommend remediation (combination of red and blue teams)
- **In realtime** We want to empower purple teams with tools that can accelerate testing and extend security test coverage and do it in real-time.
- **Based on graphs:** We have created a framework where complex activities are modeled as graph data as the most suitable representation.
- **AI/ML:** reasoning based on threat intelligence computing scaled up with graph embedding for pattern matching and root cause analysis triggering explainability and real-time forensics.



Cisco Cloud Application Security (Panoptica) CDR



On boarding CDR

The screenshot displays the Panoptica web application interface. The left sidebar contains a navigation menu with categories: Dashboard, THREATS & VULNERABILITIES (Attack Path Analysis, External Attack Surface, Vulnerability Management, Smart CDR), POSTURE MANAGEMENT (Inventory, Security Posture, Security Graph, Root Cause Analysis, Compliance Frameworks), WORKLOADS & DATA (API Security, Data Security), and BUILDS & APPLICATIONS (Code Security, CI/CD Posture). The main content area is titled 'Kubernetes' and includes a sub-header 'Panoptica integration is minimal and encrypted; the cluster role does not have access to environment secrets.' Below this, a section titled '1. Register your Kubernetes cluster with Panoptica' provides instructions to provide a unique name for the cluster. A text input field labeled 'Cluster Name' contains the value 'my-k8s-cluster-3123'. Under the heading 'Define preferences', there are three toggleable features: 'KSPM' (Kubernetes Security Posture Management) which is currently disabled, 'Smart CDR' (Realtime Cloud Detect and Response) which is enabled, and 'API Security' (Collects and analyzes live API traffic) which is also enabled. A callout box next to the KSPM toggle states 'Image scan is limited to 2GB' and provides a link to documentation. A blue 'Register' button is located at the bottom of the first section. The second section, '2. Install Panoptica's Kubernetes Controller on your registered cluster using HELM', shows a terminal command for installing the controller. The top of the interface includes a header with 'Scope: Global | Account: All Accounts', a date filter set to 'Last 7 Days', and links to 'Help Center', a notification bell, a user profile icon, and a 'DEMO' button.

panoptica
Cisco Cloud Application Security

Dashboard

THREATS & VULNERABILITIES

- Attack Path Analysis
- External Attack Surface
- Vulnerability Management
- Smart CDR

POSTURE MANAGEMENT

- Inventory
- Security Posture
- Security Graph PREVIEW
- Root Cause Analysis
- Compliance Frameworks

WORKLOADS & DATA

- API Security
- Data Security

BUILDS & APPLICATIONS

- Code Security
- CI/CD Posture

Profile

Accounts

Integrations

Repositories

Audit Log

Users & Scopes

API Keys

Alerts & Notifications

External Attack Surface

API Security

Data Security

Frameworks

Kubernetes

Panoptica integration is minimal and encrypted; the cluster role does not have access to environment secrets.

1. Register your Kubernetes cluster with Panoptica

Provide a unique name for your cluster. This name will be used to generate an access key for your cluster.

Cluster Name

my-k8s-cluster-3123

Define preferences

KSPM

☐ Kubernetes Security Posture Management continuously scans and protects Kubernetes environments, identifying risks associated with misconfigurations and vulnerabilities.

Smart CDR

☒ Realtime Cloud Detect and Response monitors your cluster for realtime security events.

API Security

☒ Collects and analyzes live API traffic; automatically builds API catalogs and discover security risks associated with them.

[Image scan is limited to 2GB](#)
For more details about KSPM feature integration, view [our documentation](#)

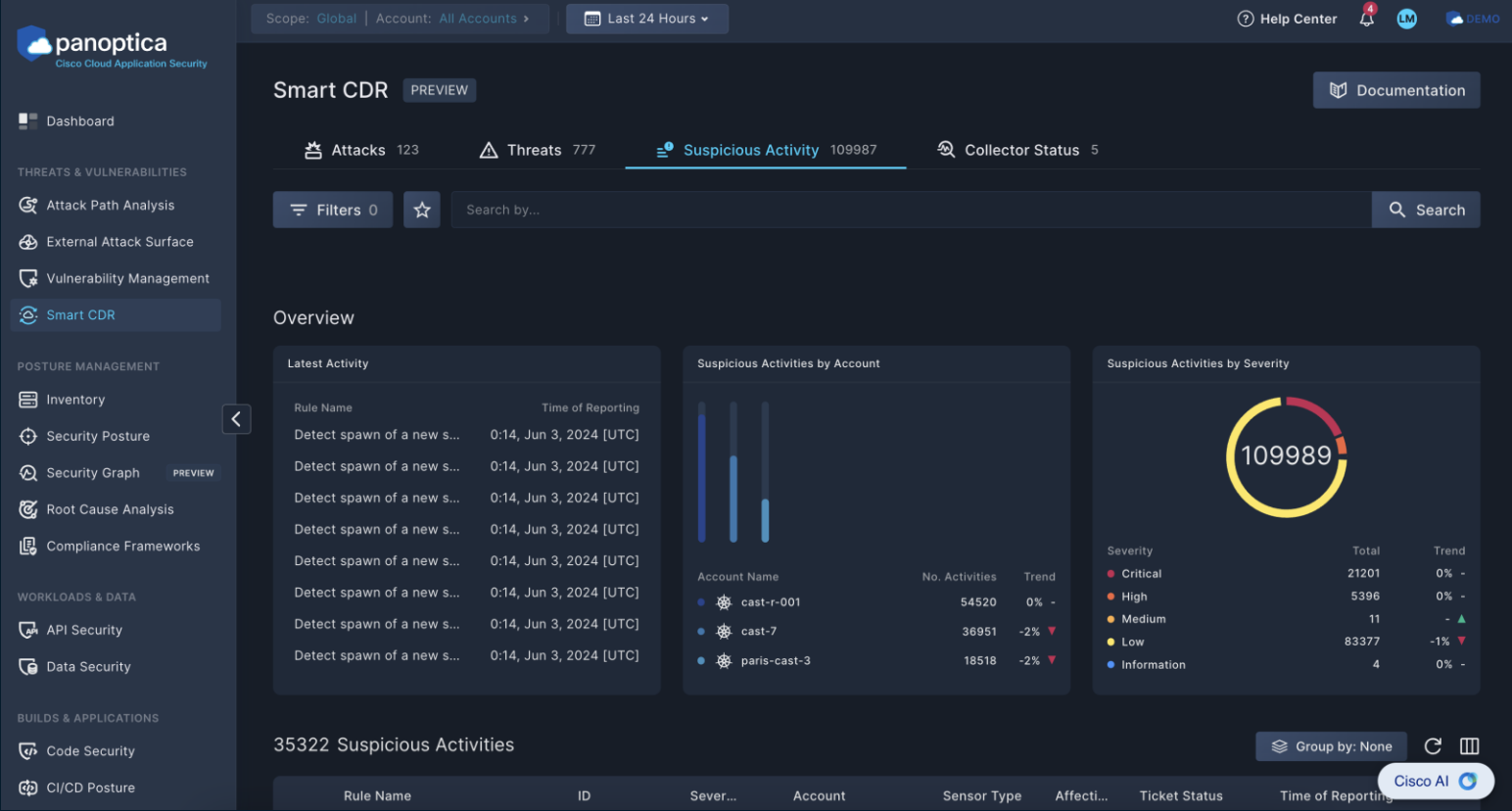
[Register](#)

2. Install Panoptica's Kubernetes Controller on your registered cluster using HELM


```
helm upgrade -i panoptica oci://us-docker.pkg.dev/cloud/panoptica/public-registry/panoptica --create namespace --set controller.secret.sharedSecret=XXXX --set controller.agentID=XXXX
```

Cisco AI

Suspicious Activity Overview



Suspicious Activity Details



panoptica
Cisco Cloud Application Security

Dashboard

THREATS & VULNERABILITIES

Attack Path Analysis

External Attack Surface

Vulnerability Management

Smart CDR

POSTURE MANAGEMENT

Inventory

Security Posture

Security Graph

Root Cause Analysis

Compliance Frameworks

WORKLOADS & DATA

API Security

Data Security

BUILDS & APPLICATIONS

Code Security

CI/CD Posture

Scope: Global | Account: All Accounts >

Last 24 Hours v

Detect spawn of a new s... 0:16, Jun 3, 2024 [UTC]

Detect spawn of a new s... 0:16, Jun 3, 2024 [UTC]

Detect spawn of a new s... 0:16, Jun 3, 2024 [UTC]

Detect spawn of a new s... 0:16, Jun 3, 2024 [UTC]

Write below tmp 0:16, Jun 3, 2024 [UTC]

Account N

ca

ca

pa

35302 Suspicious Activities

	Rule Name	ID	Sever...
☆	NEW Detect spawn of a new sh	531cae57-...	Critical
☆	NEW Detect spawn of a new sh	4e2a1901-...	Critical
☆	NEW Detect spawn of a new sh	30c99db1-...	Critical
☆	NEW Detect spawn of a new sh	70e086b5-...	Critical
☆	NEW Detect spawn of a new sh	a28185f3-...	Critical
☆	NEW Detect spawn of a new sh	f0a42b31-...	Critical
☆	NEW Detect spawn of a new sh	60845d64-...	Critical
☆	NEW Detect spawn of a new sh	55f62359-...	Critical
☆	NEW Detect spawn of a new sh	105f2f12-b...	Critical
☆	NEW Detect spawn of a new sh	3dcf6b7c-...	Critical

Showing 31 - 40 of 35302 results Rows per page 10 v

SA-2e204549-7066-41eb-9874-2a1fb6fa9b5d

Create Ticket Dismiss ☆ ↻

ID	SA-2e204549-7066-41eb-9874-2a1fb6fa9b5d	Affecting	4 Assets & Objects
Severity	High	Labels	T1059 Container
Sensor	Falco Syscall	Time of Reporting	0:13, Jun 3, 2024 [UTC]
Account	Cast-r-001		

Description

07:19:30.748311610: Error Drift detected (chmod), new executable created in a container.

(user=<NA> user_loginuid=-1 command=busybox sh -c mv xmrigr /tmp/xmrigr

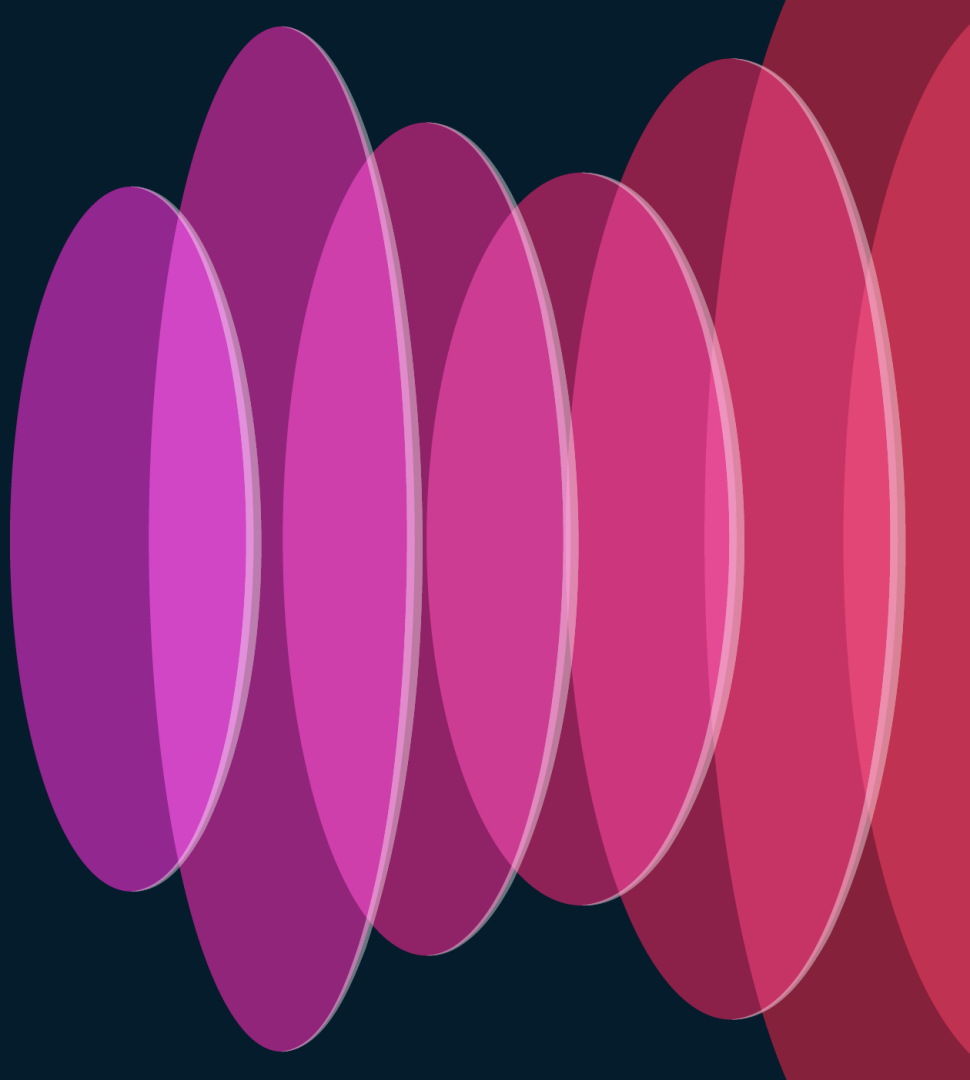
chmod +x /tmp/xmrigr pid=2757556 filename=/tmp/xmrigr name=<NA> mode=S_IXOTH|S_IROTH|S_IXGRP|S_IRGRP|S_IXUSR|S_IUSR|S_IRUSR event=chmod container_id=7e4325ecb58c container_image=dockerhub.cisco.com/espresso-docker/dragonfly-juice-shop-devel container_image_tag=latest container_name=dragonfly-ast-juice-shop k8s_ns=dragonfly-ast-juice-shop k8s_pod_name=dragonfly-ast-juice-shop-ccc477869-bldkp

Affected Assets & Objects

Name	Provider/Account	Asset type	Location
Xmrigr	/Tmp/Xmrigrcluster-Bc7d8dae-5	File	/tmp/xmrigr
2757556		Process	busybox sh -c mv xmrigr /t
Dragonfly Ast Juice Shop C		Pod	dragonfly-ast-juice-shop

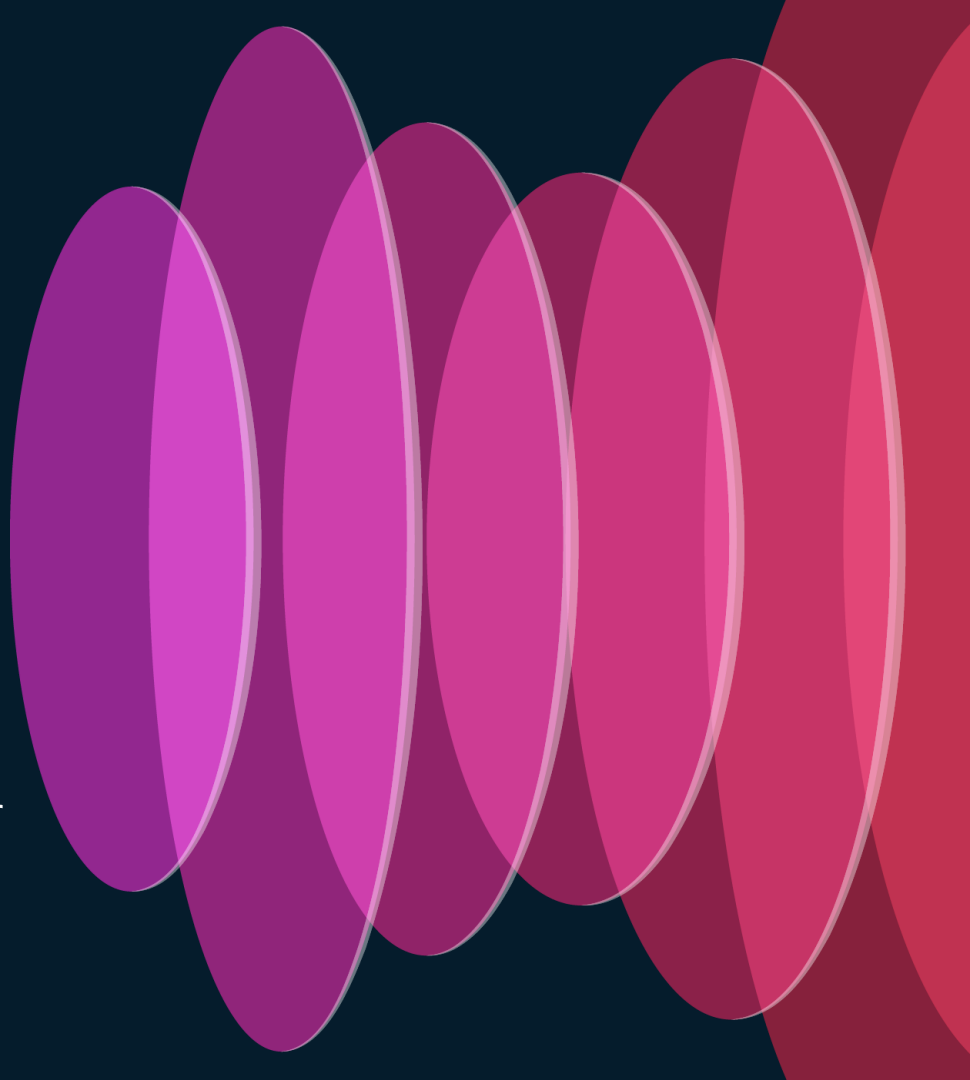
What is suspicious behavior?

Suspicious behaviors are events that have been identified based on rules defined within CDR (see next slide).



Examples: Rules that are used to identify suspicious activity

- Attacker may attempt to obtain account login and credential material
- Detect dumping of process memory
- Scan ports to check for listening ports
- Detect read & writes to file system
- Detect delete system and audit logs
- Connection attempts to Kubernetes cluster local domain (internal services in the cluster)



Notifications (WebEx) Example



Dragonfly Notification

RuleID: 432a61cd-056c-4d02-822a-0e87b33e3833

Cluster uuid: cluster-test-demo

Severity: CRITICAL

Namespace: host

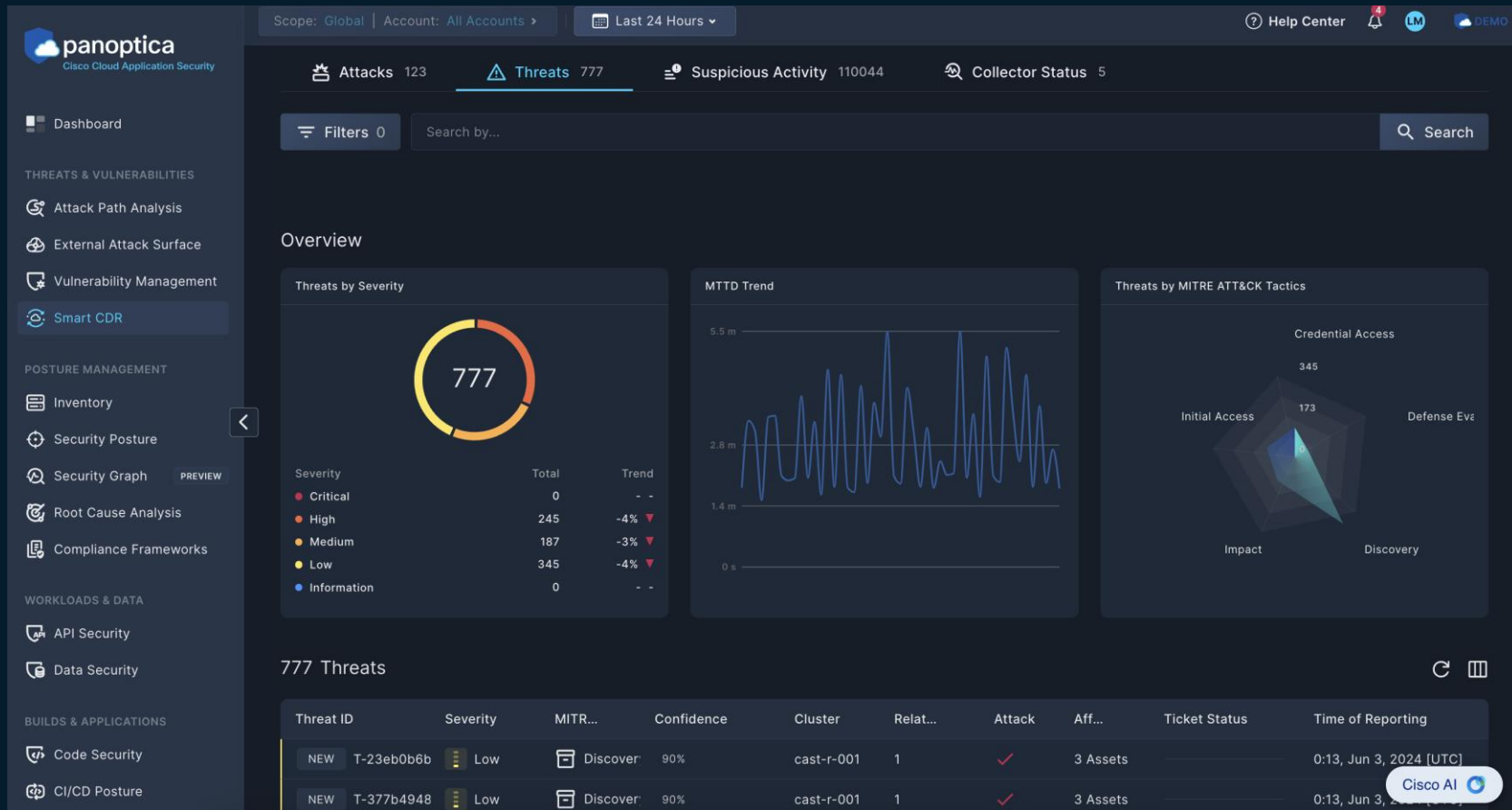
Pod name: undefined

Container name: host


Description: 09:13:17.764680555: Alert Attempts to obtain the mapping between cluster local domain and Cluster Ip (command=nslookup [kubernetes.default.svc.cluster.local](#) pid=603424 container_id=host name_space= pod_name=) k8s.ns= k8s.pod= container=host

Investigate in Panoptica

Threats Overview



Threat Details



Cisco Cloud Application Security

Dashboard

THREATS & VULNERABILITIES

Attack Path Analysis

External Attack Surface

Vulnerability Management

Smart CDR

POSTURE MANAGEMENT

Inventory

Security Posture

Security Graph

Root Cause Analysis

Compliance Frameworks

WORKLOADS & DATA

API Security

Data Security

BUILDS & APPLICATIONS

Code Security

CI/CD Posture

CI/CD Scans

Scope: Global | Account: All Accounts >

Last 24 Hours ▾

Severity	Total	Trend
Critical	0	- -
High	244	-5% ▾
Medium	188	-3% ▾
Low	343	-5% ▾
Information	0	- -

1.4 m

0 s

775 Threats

Threat ID	Severity	MITR...	Confidence
T-29500c78-b296-41	Medium	Credenti	80%
T-b92cbfed-c13c-42	High	Initial Ac	90%
T-f2caf279-7730-4f5	Medium	Credenti	90%
T-06fab0e1-151a-481	High	Impact	80%
T-6fa7f2b9-5596-4b	Low	Discover	80%
T-3bbb95fd-ed16-4a	Low	Discover	90%
T-7b9cf530-7976-48	High	Initial Ac	90%
T-f5e91238-2dce-44	Medium	Defense	60%
T-86e778c6-e6c9-4e	High	Impact	90%
T-98d3c569-136a-4e	Low	Discover	90%

Showing 31 - 40 of 775 results

Rows per page 10 ▾

T-b92cbfed-c13c-4273-a199-eaec2ba96e4d

Create Ticket

Overview

Severity

High

Part of Attacks

✓

MITRE Tactic

Initial Access

Affecting

3 Assets & Objects

Confidence

90%

Labels

T1190

TA0001

Cluster

cast-r-001

Time of Reporting

23:31, Jun 2, 2024 [UTC]

Threat Story

Graph

Timeline

Table

Process

Process

File

cisco Live!

#CiscoLive

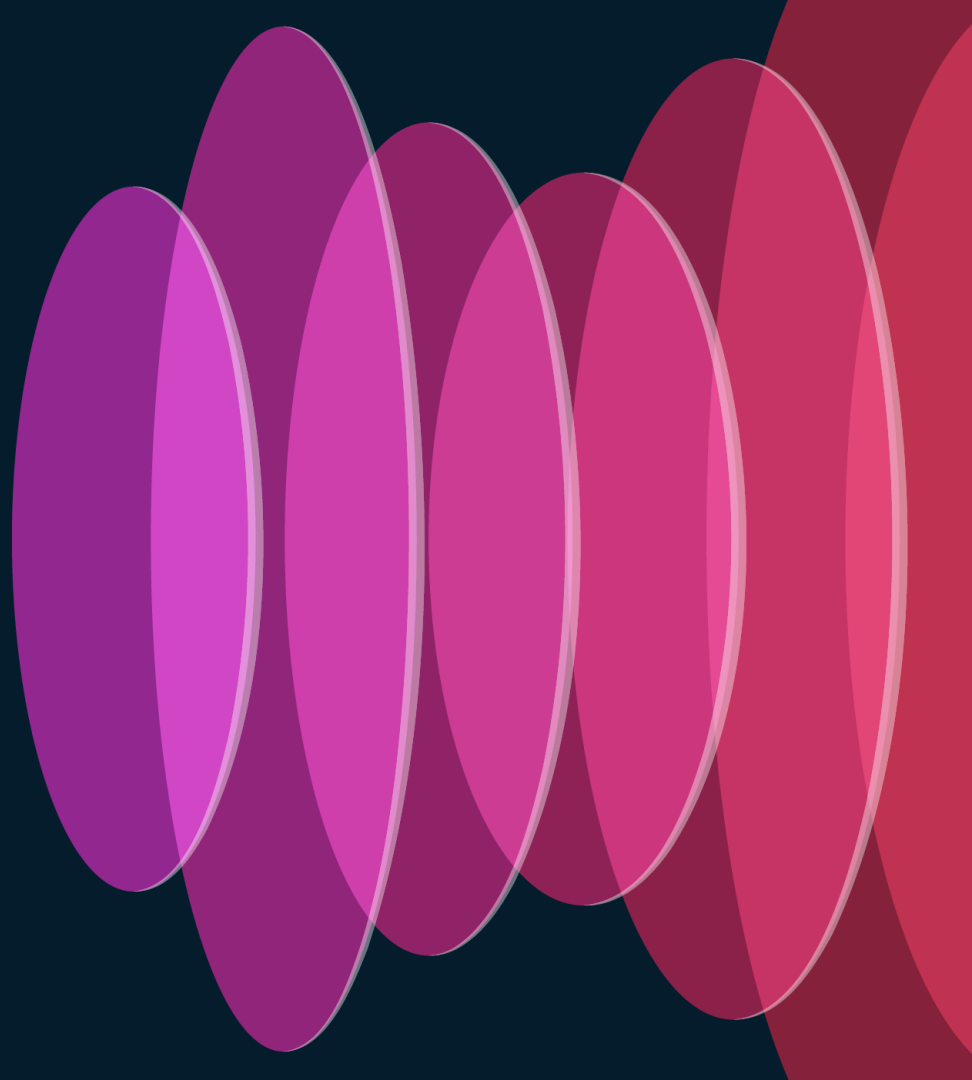
BRKNWT-2416

© 2024 Cisco and/or its affiliates. All rights reserved. Cisco Public

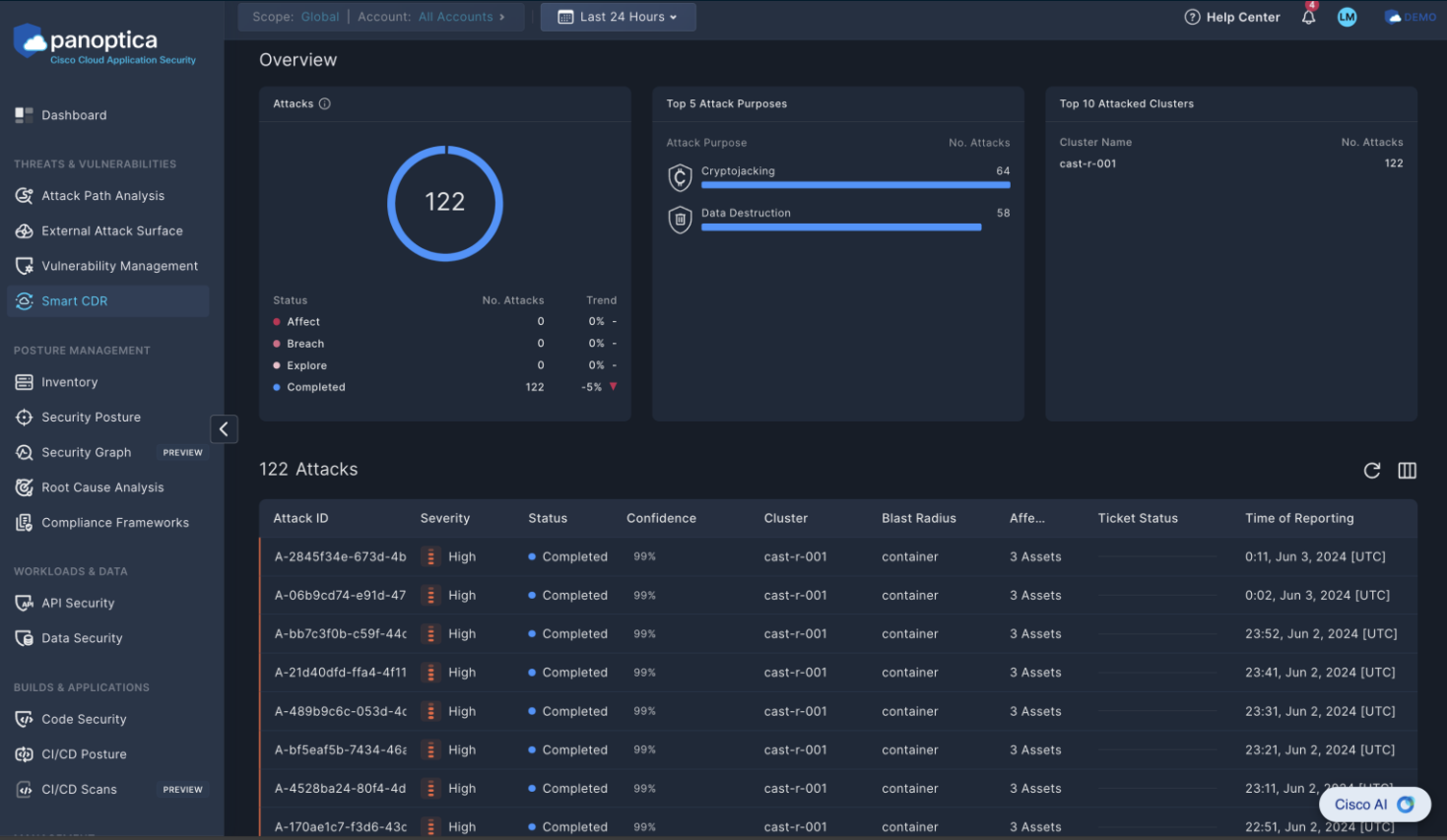
50

What is a threat?


A threat is a series of suspicious activities that are related to each other and are intended to cause harm to the system. Each threat is then mapped to a MITRE TTP (Technique, Tactic, or Procedure).



Attack Overview



Attack Details



panoptica
Cisco Cloud Application Security

Dashboard

THREATS & VULNERABILITIES

Attack Path Analysis

External Attack Surface

Vulnerability Management

Smart CDR

POSTURE MANAGEMENT

Inventory

Security Posture

Security Graph PREVIEW

Root Cause Analysis

Compliance Frameworks

WORKLOADS & DATA

API Security

Data Security

BUILDS & APPLICATIONS

Code Security

CI/CD Posture

CI/CD Scans PREVIEW

Scope: Global | Account: All Accounts > | Last 24 Hours v

Overview

Attacks 0

122

Status

Affect

0

0%

-

Breach

0

0%

-

Explore

0

0%

-

Completed

122

-5%

▼

Top 5 Attacks

Attack Path

Crypto

Data

122 Attacks

Attack ID	Severity	Status	Confidence
A-2845f34e-673d-4b	High	Completed	99%
A-06b9cd74-e91d-47	High	Completed	99%
A-bb7c3f0b-c59f-44c	High	Completed	99%
A-21d40dfd-ffa4-4f11	High	Completed	99%
A-489b9c6c-053d-4c	High	Completed	99%
A-bf5eaf5b-7434-46e	High	Completed	99%
A-4528ba24-80f4-4d	High	Completed	99%
A-170ae1c7-f3d6-43c	High	Completed	99%

A-2845f34e-673d-4bcd-ad78-ce822103935f

×

Create Ticket

↻

Severity

High

Blast Radius

container

Status

Completed

Affecting

3 Assets & Objects

Confidence

99%

Labels

T1003 T1003.008

Cluster

cast-r-001

Time of Reporting

0:11, Jun 3, 2024 [UTC]

Show all ▼

Attack Story

Sequence of realtime security threat events. [Learn more](#)

Attacker

Impact

Defense Evasion

Discovery

Discovery

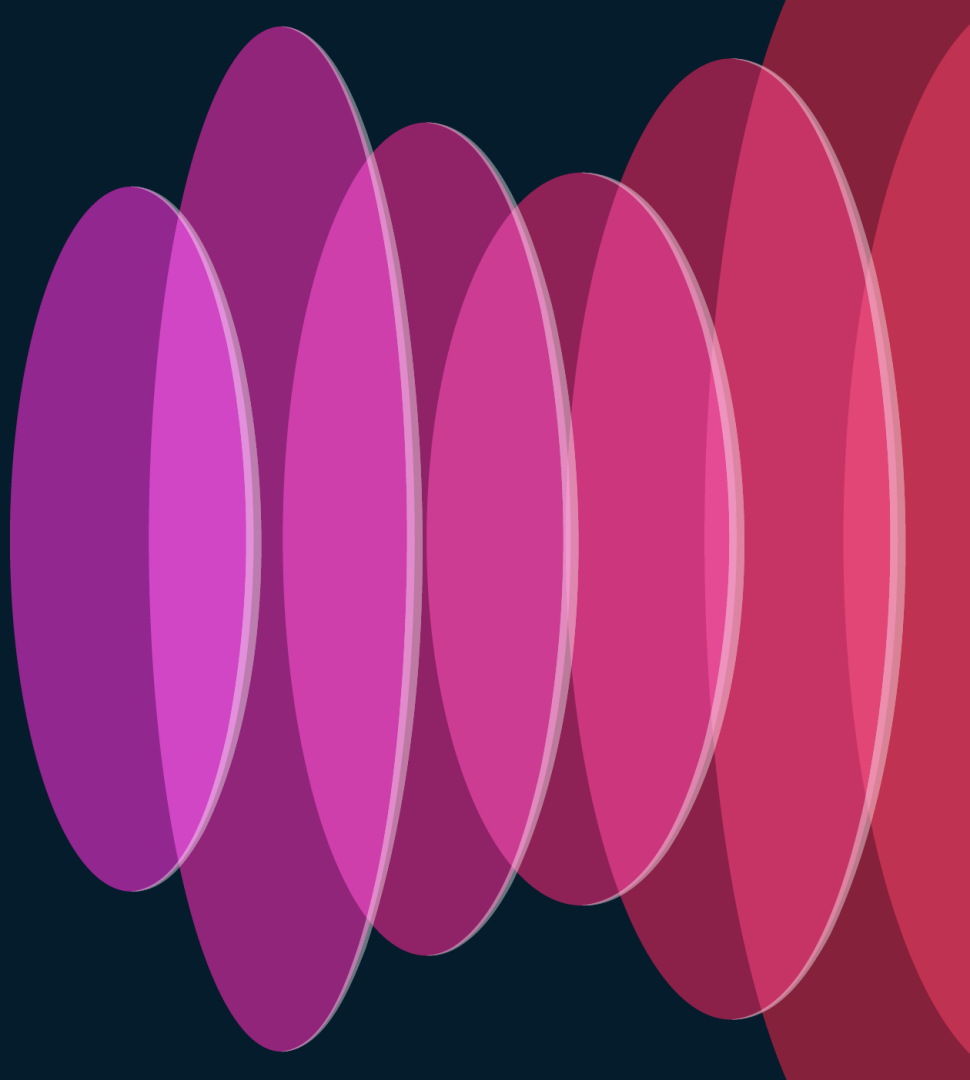
Discovery

Discovery

Credential Access

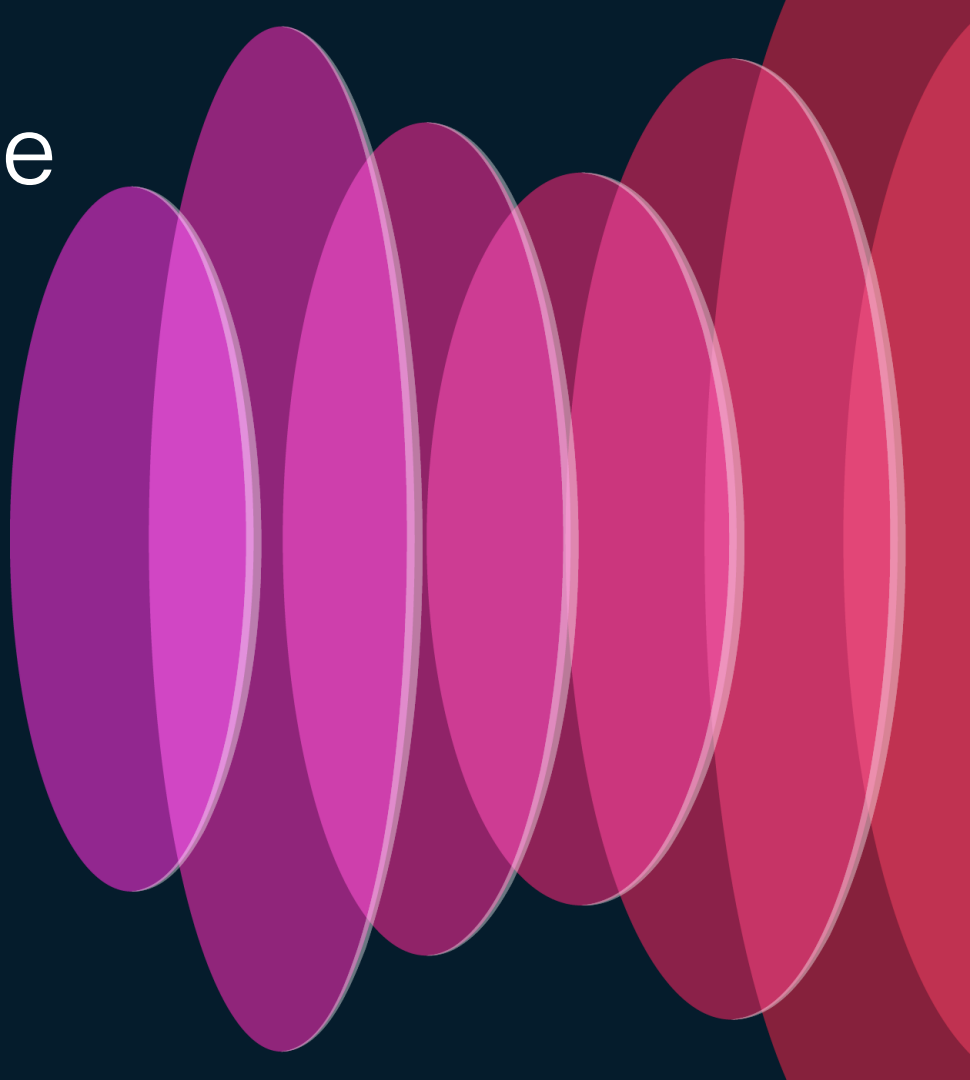
What is an attack?

An attack is an attempt to collect, disrupt, deny, degrade, or destroy the system. CDR focuses on identifying 5 categories of attacks (see next slide) based on a series of threats.



CDR Attack Coverage

- Ransomware
- Data Exfiltration
- Container Escaping
- Crypto-mining
- Data Destruction
- Unknown



Complete Your Session Evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to **win 1 of 5 full conference passes** to Cisco Live 2025.



Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.



Level up and earn **exclusive prizes!**



Complete your surveys in the **Cisco Live mobile app**.

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at: **Insert preferred comms method**



The bridge to possible

Thank you

CISCO *Live!*

#CiscoLive

