クラウドセキュリティにおける AI/MLの役割に関して解説

パロアルトネットワークス株式会社
Chief Technology Officer, Japan & Asia Pacific
Ajay Mishra

Agenda

- AI/MLによる最近攻撃手法の解説
- AI/MLによる防御解説
- AI/ML自体への攻撃/防御解説
- 結論

AI/MLによる最近攻撃手法の解説

paloalto NETWORKS

AI/MLによる最近攻撃手法の解説

LLMによるMalwareの開発1: BlackMamba

- **AI-generated polymorphic malware:**
  - HYAS Labs の研究者は、BlackMamba と呼ばれる概念実証攻撃を実証しました。この攻撃は、大規模言語モデル (LLM) (ChatGPT のベースとなるテクノロジ) を悪用して、ポリモーフィックキーロガー機能をcommand-and-controlを使わずにオンザフライで合成する。 この攻撃は、BlackMamba が実行されるたびに、そのキーロギング機能を再合成するという点で「真にポリモーフィック」である事を報告されてる。.
- **Exfiltration through MS Teams:**
  - キャプチャしたデータを盗み出すために、BlackMamba は、収集したデータを悪意のある Teams チャネルに Webhook 経由で送信することにより、MS Teams を盗み出しチャネルとして使用する。

ML/AIによる攻撃の検知で十分なのか?

最近の攻撃手法は回避型になって来ている:そもそもSandbox防御を回避する複数の手法が報告されてる

- **Sandbox evasions are a major challenge:**
  - マルウェアの作成者はさまざまな手法を使用してサンドボックス環境による検出を回避し、自動マルウェア分析システムが悪意のあるソフトウェアを特定して対抗することを困難になってる
- **Multiple categories of evasions:**
  - 下記の主な手法を使用してサンドボックス環境による検出を回避:
    - マルウェアが仮想マシンで実行されているかどうかを検出しようとするアンチ VM 手法。
    - マルウェアがさまざまな手法を使用して時間を浪費し、意味のある分析を妨げる、タイミングおよびコンピューティング リソースの回避
    - マルウェアがサンドボックスではなく実際の環境で実行されている兆候を探す手法

```
.text:0050AAC7 B8 68 58 4D 56                mov     eax, 'VMXh'
.text:0050AACC BB 00 00 00 00                mov     ebx, 0
.text:0050AAD1 B9 0A 00 00 00                mov     ecx, 0Ah
.text:0050AAD6 BA 58 56 00 00                mov     edx, 'VX'
.text:0050AADB ED                            in      eax, dx         ; read from "VX port"
.text:0050AADC 81 FB 68 58 4D 56             cmp     ebx, 'VMXh'     ; check if reply is from VMware
.text:0050AAE2 0F 94 45 E4                   setz    byte ptr [ebp+fDetected]
```

VMwareの環境かどうかをCheckしている

```
.text:00469691                       lblWaitLoop:                            ; CODE XREF: .text:004696A6↓j
.text:00469691                                                               ; .text:00469706↓p
.text:00469691 F6 C4 FC                      test    ah, 0FCh
.text:00469694 41                            inc     ecx
.text:00469695 81 FF 16 81 D0 5B             cmp     edi, 5BD08116h
.text:0046969B 0F 31                         rdtsc
.text:0046969D 80 FB D3                      cmp     bl, 0D3h
.text:004696A0 81 F9 FF FF FF 00             cmp     ecx, 0FFFFFFh
.text:004696A6 75 E9                         jnz     short lblWaitLoop
.text:004696A8 66 F7 C7 34 93                test    di, 9334h
.text:004696AD 5B                            pop     ebx
```

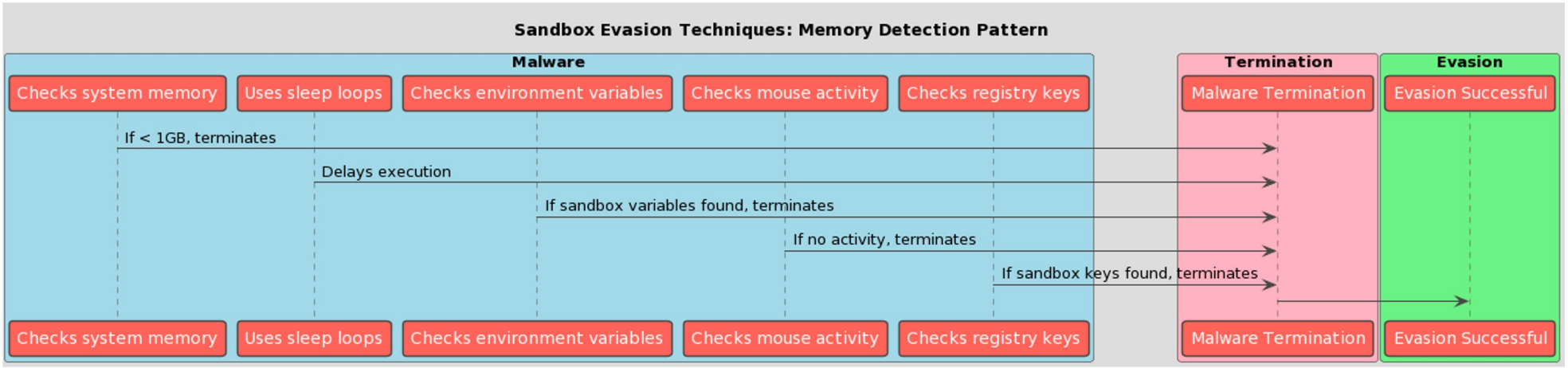
Sleep関数を使って意味のないログの生成して分析を妨げる

```
1  BOOLEAN __stdcall check_memory()
2  {
3    HMODULE LibraryW; // rax
4    void (__stdcall *GetSystemInfo)(LPSYSTEM_INFO); // [rsp+48h] [rbp-B0h]
5    HANDLE hDevice; // [rsp+50h] [rbp-A8h]
6    DWORD BytesReturned; // [rsp+58h] [rbp-A0h] BYREF
7    struct _MEMORYSTATUSEX mem_status; // [rsp+60h] [rbp-98h] BYREF
8    DISK_GEOMETRY disk_geometry; // [rsp+A0h] [rbp-58h] BYREF
9    SYSTEM_INFO sys_info; // [rsp+B8h] [rbp-40h] BYREF
10
11   LibraryW = LoadLibraryW(L"Kernel32.dll");
12   GetSystemInfo = (void (__stdcall *)(LPSYSTEM_INFO))GetProcAddress(LibraryW, "GetSystemInfo");
13   if ( !GetSystemInfo )
14     return 0;
15   ((void (__fastcall *)(SYSTEM_INFO *))GetSystemInfo)(&sys_info);
16   if ( sys_info.dwNumberOfProcessors < 2 )      // check if number of processors is less than 2
17     return 1;
18   hDevice = CreateFileW(L"\\\\.\\PhysicalDrive0", 0, 3u, 0i64, 3u, 0, 0i64);
19   DeviceIoControl(
20     hDevice,
21     IOCTL_DISK_GET_DRIVE_GEOMETRY,
22     0i64,
23     0,
24     &disk_geometry,
25     sizeof(DISK_GEOMETRY),
26     &BytesReturned,
27     0i64);
28   mem_status.dwLength = sizeof(
29   GlobalMemoryStatusEx(&mem_status);
30   return (unsigned int)(mem_status.ullTotalPhys / 1024 / 1024) < 2048;// check if RAM is less than 2GB
31 }
```

実行に必要なプロセッサの最小数と必要なメモリを確認してる

ML/AIによる攻撃の検知で十分なのか?

最近の攻撃手法は回避型になって来ている:そもそもSandbox防御を回避する複数の手法が報告されてる



複数の手法でサンドボックス環境による検出を回避

AI/MLによる防御対策

paloalto NETWORKS

AI/MLによる防御対策

まずマインドセットを変える: Vendor Consolidationの実現で顧客満足

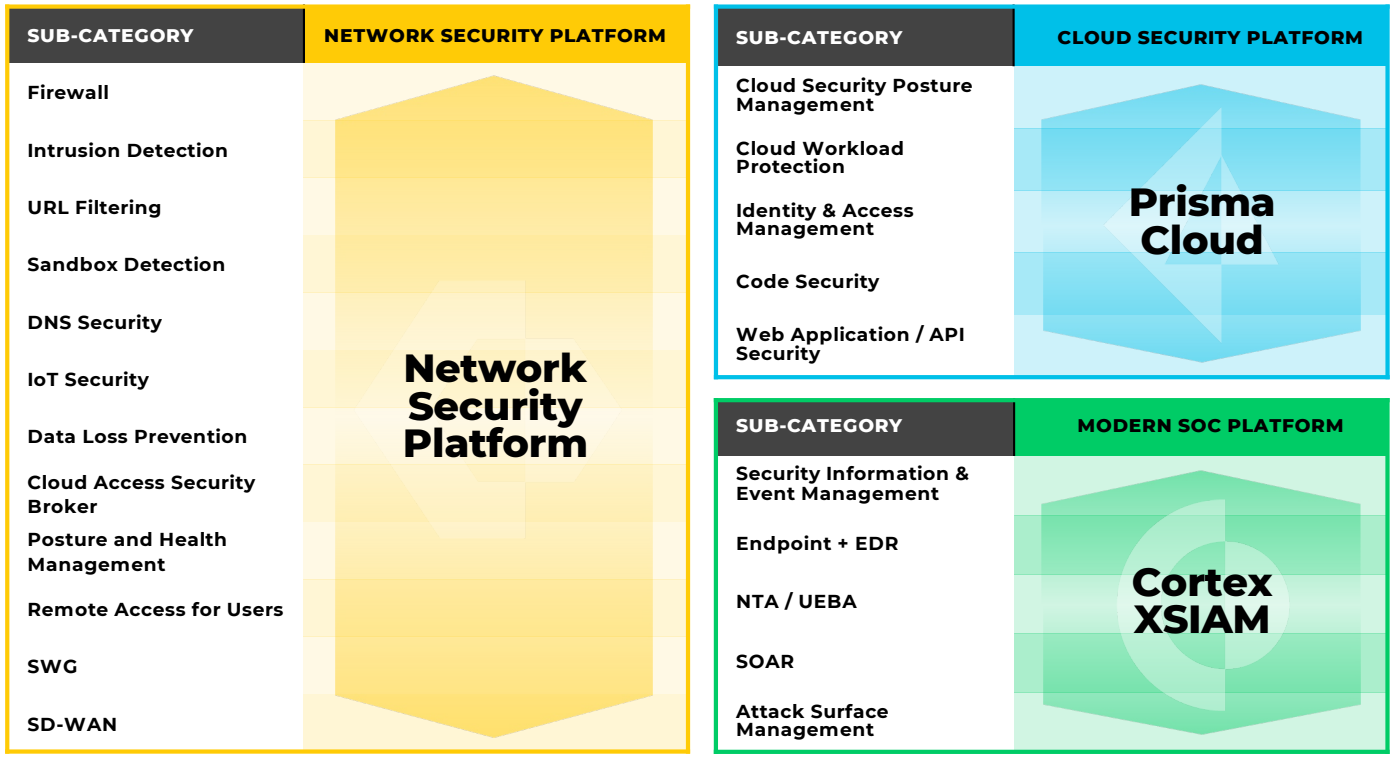
SUB-CATEGORY	POINT PRODUCTS	NETWORK SECURITY PLATFORM
Firewall	Check Point SOFTWARE TECHNOLOGIES LTD	Network Security Platform
Intrusion Detection	CISCO	
URL Filtering	BROADCOM	
Sandbox Detection	FIREEYE	
DNS Security	Infoblox	
IoT Security	ARMIS	
Data Loss Prevention	BROADCOM	
Cloud Access Security Broker	netskope	
Posture and Health Management	tufin dynatrace	
Remote Access for Users	CISCO	
SWG	zscaler	
SD-WAN	velocloud	

SUB-CATEGORY	POINT PRODUCTS	CLOUD SECURITY PLATFORM
CSPM	Dome9	Prisma Cloud
Cloud Workload Protection	aqua	
Identity & Access Management	CLOUDKNOX	
Code Security	snyk	
Web Application / API Security	CLOUDFLARE	

SUB-CATEGORY	POINT PRODUCTS	MODERN SOC PLATFORM
SIEM	splunk>	Cortex XSIAM
Endpoint + EDR	CROWDSTRIKE	
NTA / UEBA	DARKTRACE	
SOAR	Siemplify	
Attack Surface Management	RISKIQ	

AI/MLによる防御対策

まずマインドセットを変える: Vendor Consolidationの実現で顧客満足

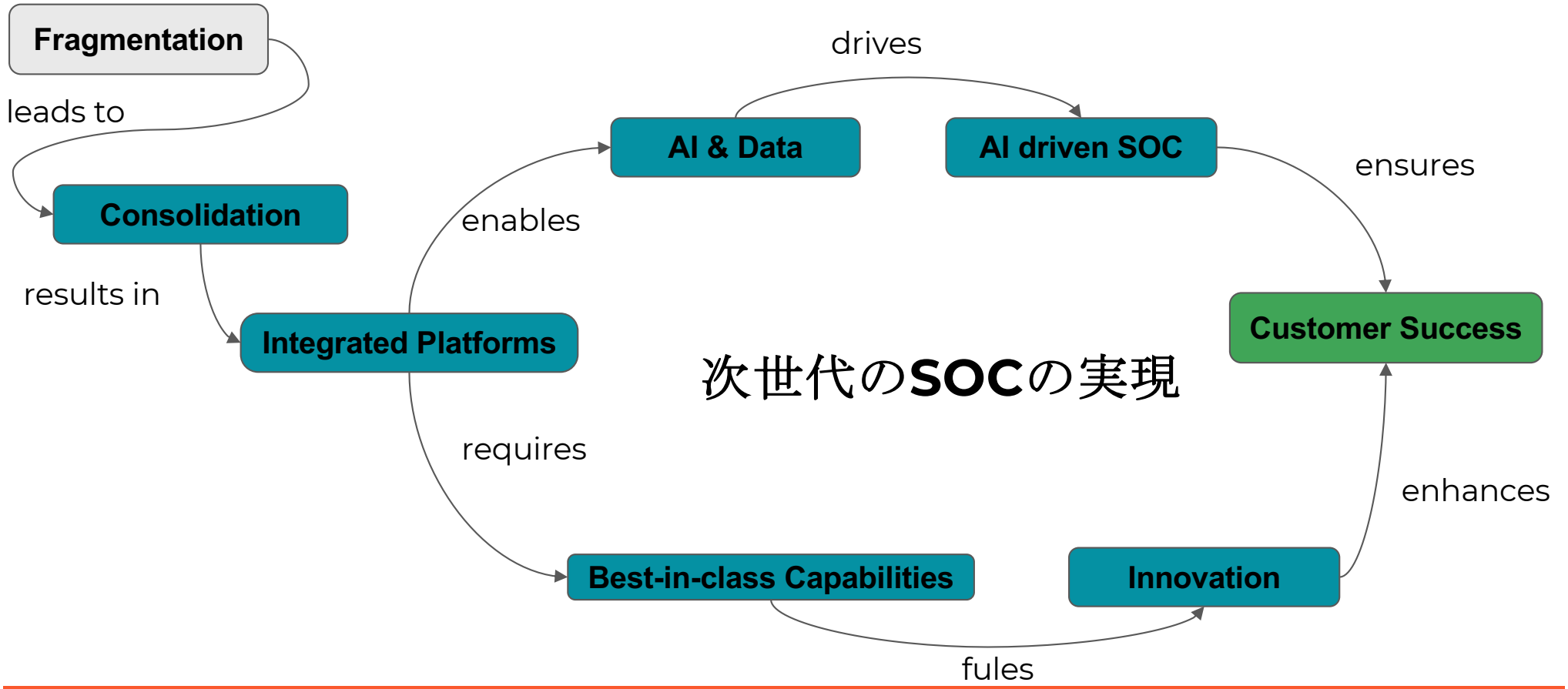


AI/MLによる防御対策

まずマインドセットを変える: Vendor Consolidationの実現で顧客満足

AI/MLによる防御対策

まずマインドセットを変える: Vendor Consolidationの実現で顧客満足



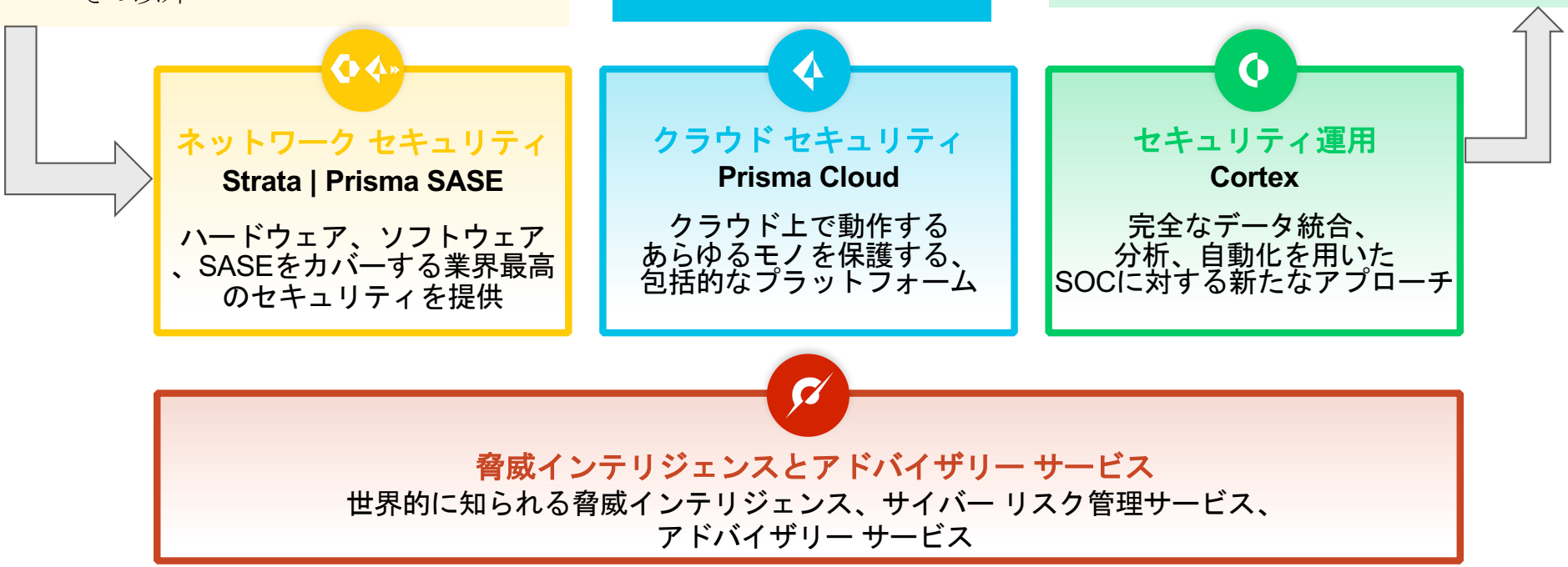
AI/MLによる防御対策

まずマインドセットを変える: AI/MLによる攻撃をPlatformとして防御する,LLM系の攻撃に準じた観点

- App-ID, Phishing Domain
- その以外

- DNS攻撃の防御
- その以外

- Chat GPT based XSOAR
- その以外

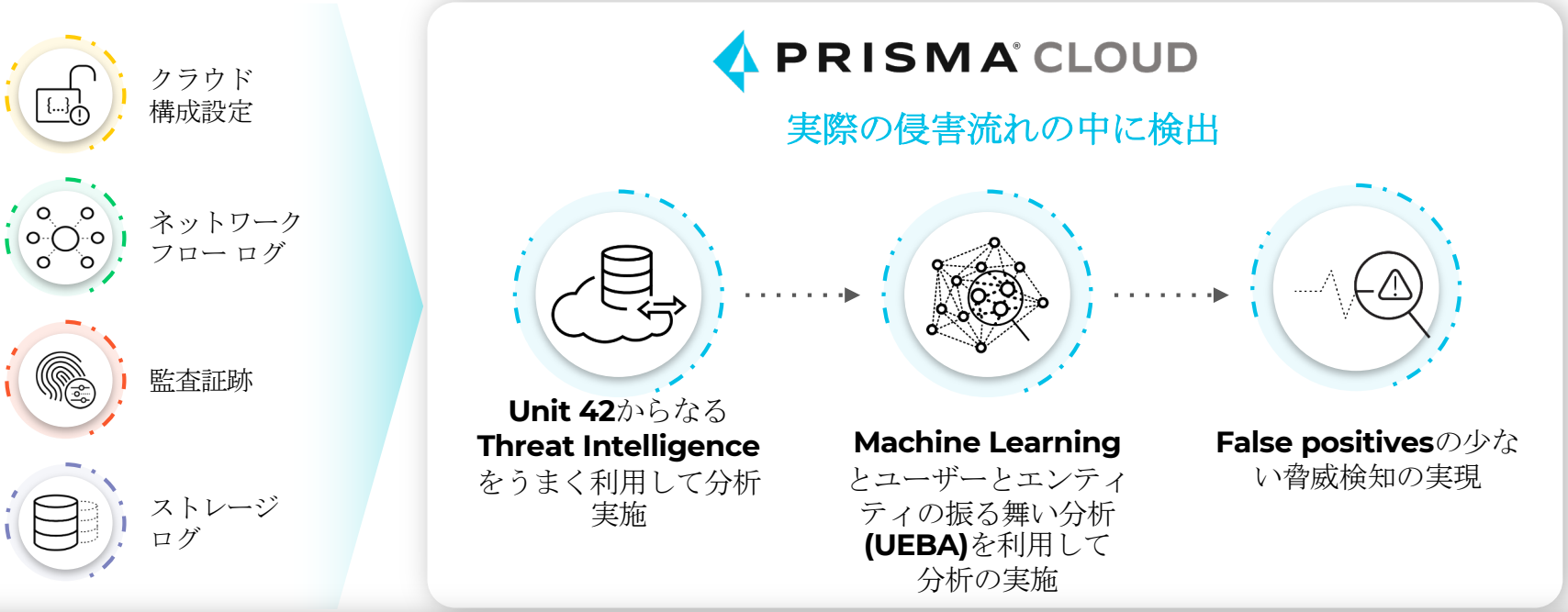


Cloud Securityにおける AI/MLの役割を解説

paloalto NETWORKS

Prisma CloudのCSPMとは?

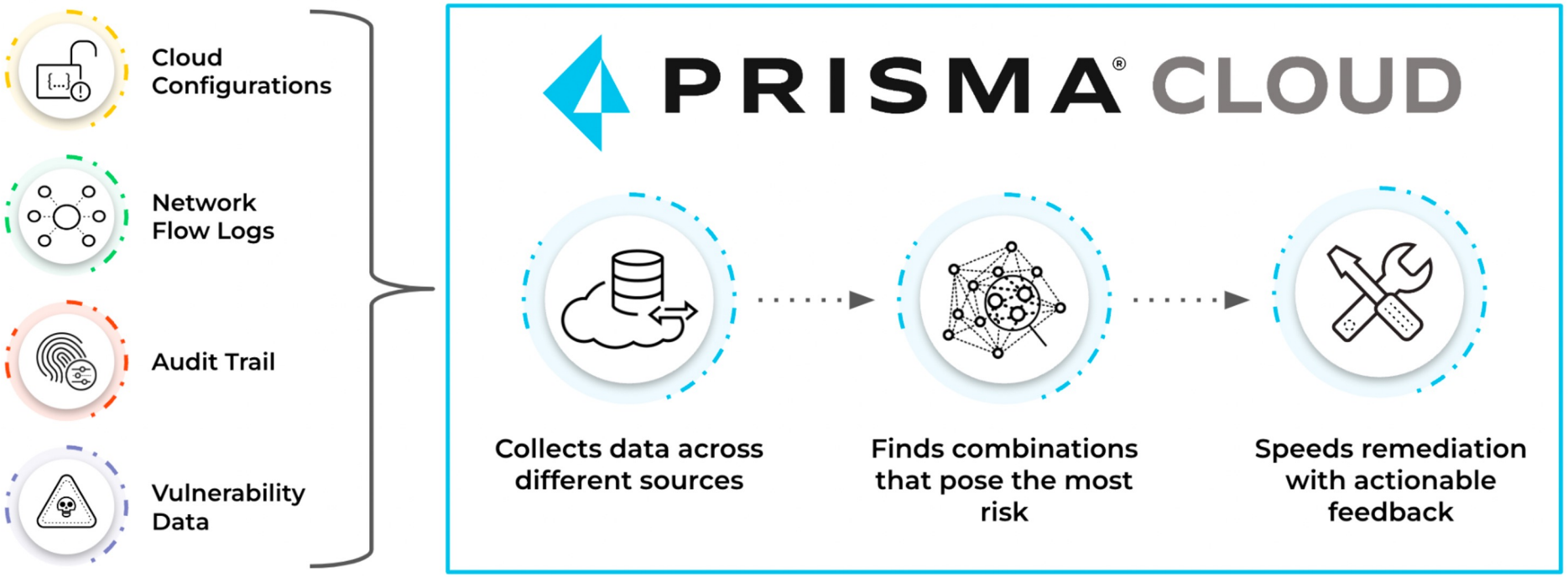
Machine Learningを利用して脅威検知する:通常と異なるユーザーアクティビティがないかクラウド環境を監視



何か危ない事起きるだろうの論理を排除し、今何ができるのかを実践する

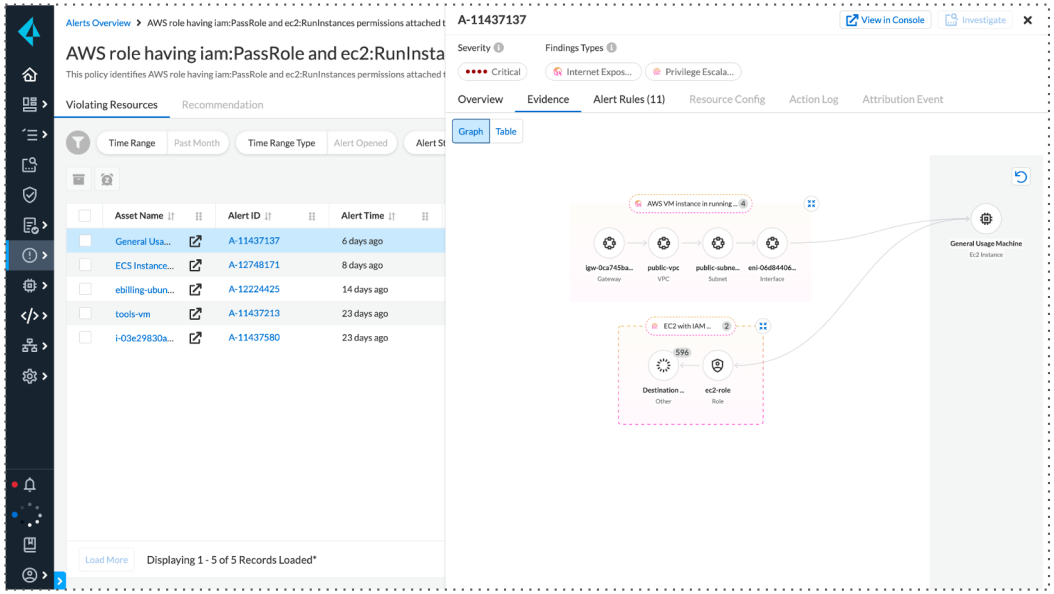
Prisma CloudのCSPMとは?

Machine Learningを利用して脅威検知する:Disrupt Attack Paths: How to Prioritize Your Most Harmful Risk



Prisma CloudのCSPMとは?

Disrupt Attack Paths: How to Prioritize Your Most Harmful Risk

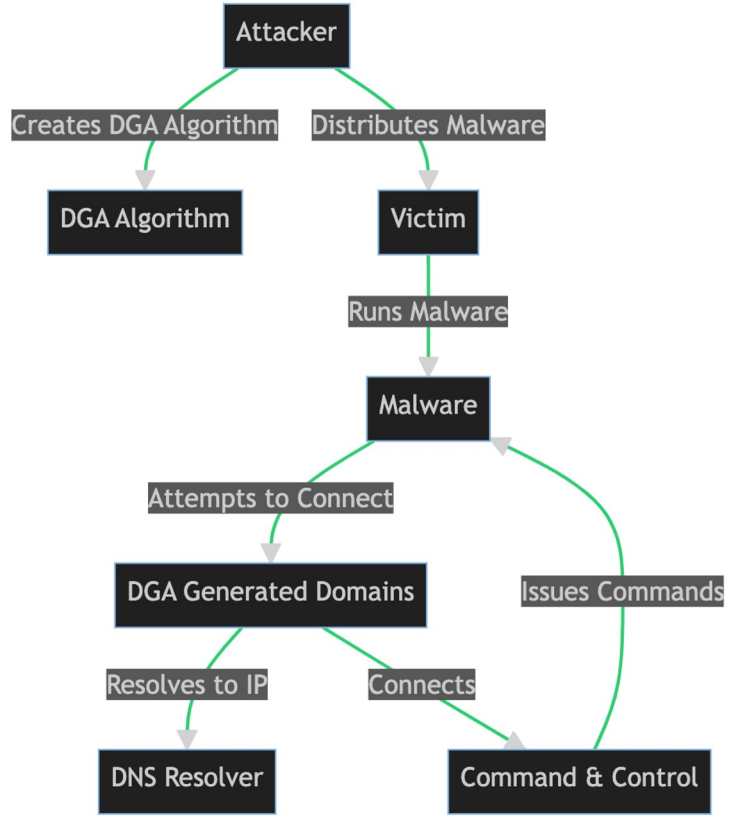


Machine Learning の役割

- In addition to finding what could go wrong in cloud environments, Prisma Cloud applies threat context to identify what is going wrong. With Palo Alto Networks Unit 42 threat intelligence, **coupled with machine learning (ML) and user and entity behavior analytics (UEBA)**, security teams can detect exploited attack paths.Examples combinations potentially indicating attack paths:
  - **A workload with a critical vulnerability that's exposed to the internet and has excessive access permissions.**
  - Azure AD user with Key Vault access performing unusual activity
  - **Detected network data exfiltration activity on a publicly accessible workload with a critical, exploitable vulnerability**

Prisma CloudのCSPMとは?

Detect DNS Threats for AWS Environments with Prisma Cloud: DGA攻撃



Machine Learning の役割

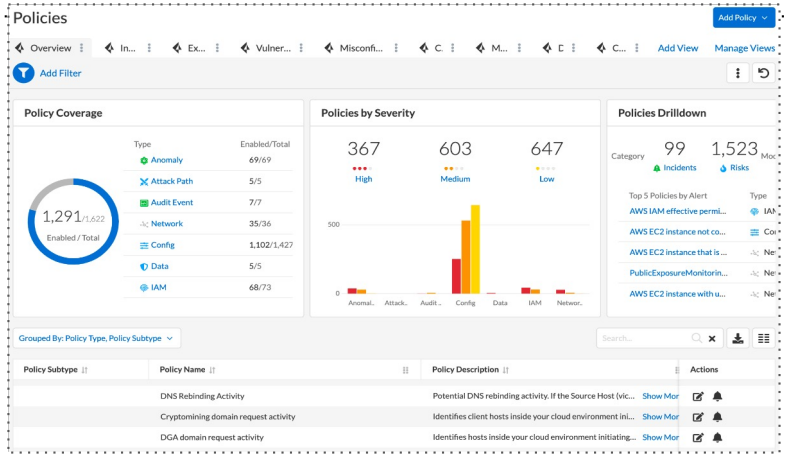
- **Detection of DGA Activity:** **Domain Generation Algorithm (DGA) generates domain names based on a dynamic seed and an algorithm**. Attackers use this technique to register new random-looking domains for their victims to rendezvous with the attacker's network.

Prisma CloudのCSPMとは?

Detect DNS Threats for AWS Environments with Prisma Cloud

Prisma Cloud は、**AWS** 上の **DNS** トラフィックをアクティブに監視し、**DNS** トラフィックに隠れたさまざまな脅威を検出できます。

- 世界的な組織の 88% が DNS 攻撃を受けています。
- 企業は年間平均 7 件の DNS 攻撃を受けています。
- DNS 攻撃あたりのコストは平均 942,000 ドルです。



Machine Learning の役割

- **Detection of DGA Activity:**Domain Generation Algorithm (DGA) generates domain names based on a dynamic seed and an algorithm. Attackers use this technique to register new random-looking domains for their victims to rendezvous with the attacker's network. **Prisma Cloud uses machine learning to monitor the DNS queries and detect suspicious DGA domain request activities**
- **In General:**Using this data, the context-driven platform leverages Palo Alto Networks Unit 42 threat intelligence, third party intelligence streams, **machine learning (ML) and user and entity behavior analytics (UEBA)** to identify threats lurking across cloud environments. With each threat detected, Prisma Cloud provides actionable remediation steps to help you respond.

Prisma CloudのCSPMとは?

Detect DNS Threats for AWS Environments with Prisma Cloud

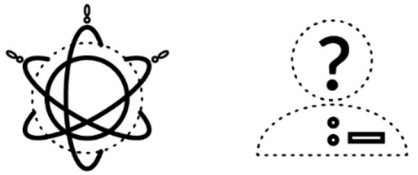
ML-powered & threat intel based detection with contextual insights

Immediately pinpoint highest risk security issue and impact

Data Sources:
IaaS/PaaS logs, user logs, storage logs, network traffic logs

Threat Intelligence:
Malicious / suspicious IP addresses, malware signatures, 3rd party integrations

Advanced Analysis:
Machine learning, statistical modeling, graph analysis, natural language processing



Threat Detection:
Advanced analysis of data sources and threat intelligence to **detect both known and unknown threats**

ML/AIによる攻撃の検知で十分なのか?

最近の攻撃手法は回避型になって来ている:MITRE ATLAS™に関して

- MITRE ATLAS™ は、研究者が機械学習 (ML) システムの脅威の状況を理解するのに役立つように設計されたナレッジ ベース

Reconnaissance &	Resource Development &	Initial Access &	ML Model Access	Execution &	Persistence &	Defense Evasion &	Discovery &	Collection &	ML Attack Staging	Exfiltration &	Impact &
5 techniques	7 techniques	4 techniques	4 techniques	2 techniques	2 techniques	1 technique	3 techniques	3 techniques	4 techniques	2 techniques	7 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution &	Poison Training Data	Evade ML Model	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	ML-Enabled Product or Service	Command and Scripting Interpreter &	Backdoor ML Model		Discover ML Model Family	Data from Information Repositories &	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Adversarial ML Attack Capabilities	Evade ML Model	Physical Environment Access				Discover ML Artifacts	Data from Local System &	Verify Attack		Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access						Craft Adversarial Data		Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets										Cost Harvesting
	Poison Training Data										ML Intellectual Property Theft
	Establish Accounts &										System Misuse for External Effect

MITRE/ATLASはAI/MLに関する脅威を体型的に示す

ML/AIによる攻撃の検知で十分なのか?

一般的にAI/MLに対する防御対策例: Poisoning training data対策(今後のセッションに包括的な説明予定)

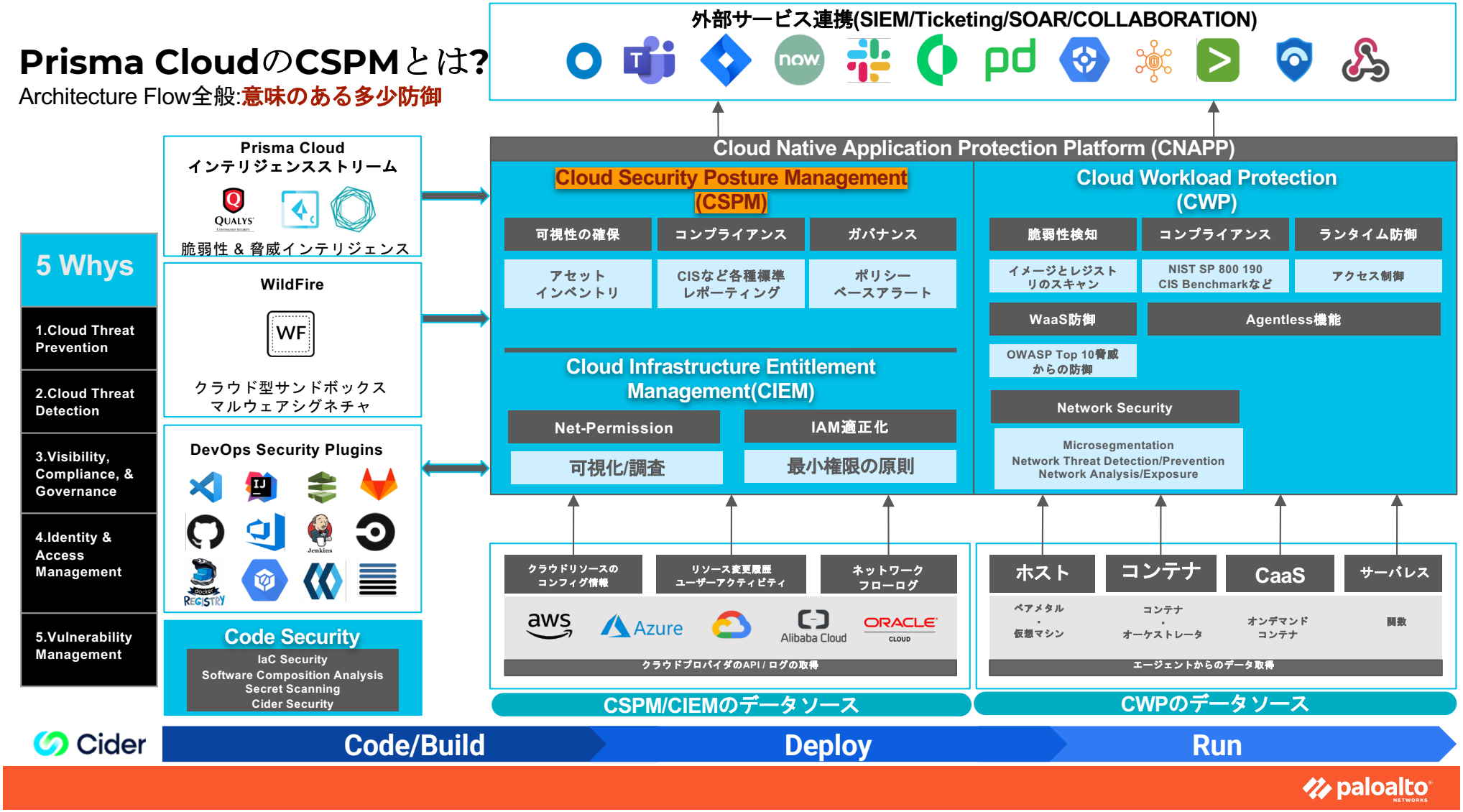
01

Poisoning training data

- 定義
  - By injecting carefully crafted data into the system, attackers may attempt to manipulate the machine learning model's training process, causing it to generate false negatives (failing to detect actual threats) or false positives (detecting non-threatening behavior as malicious).
- 対策
  - Implement robust data validation and sanitization processes to prevent the injection of malicious data.

Prisma CloudのCSPMとは?

Architecture Flow全般:**意味のある多少防御**

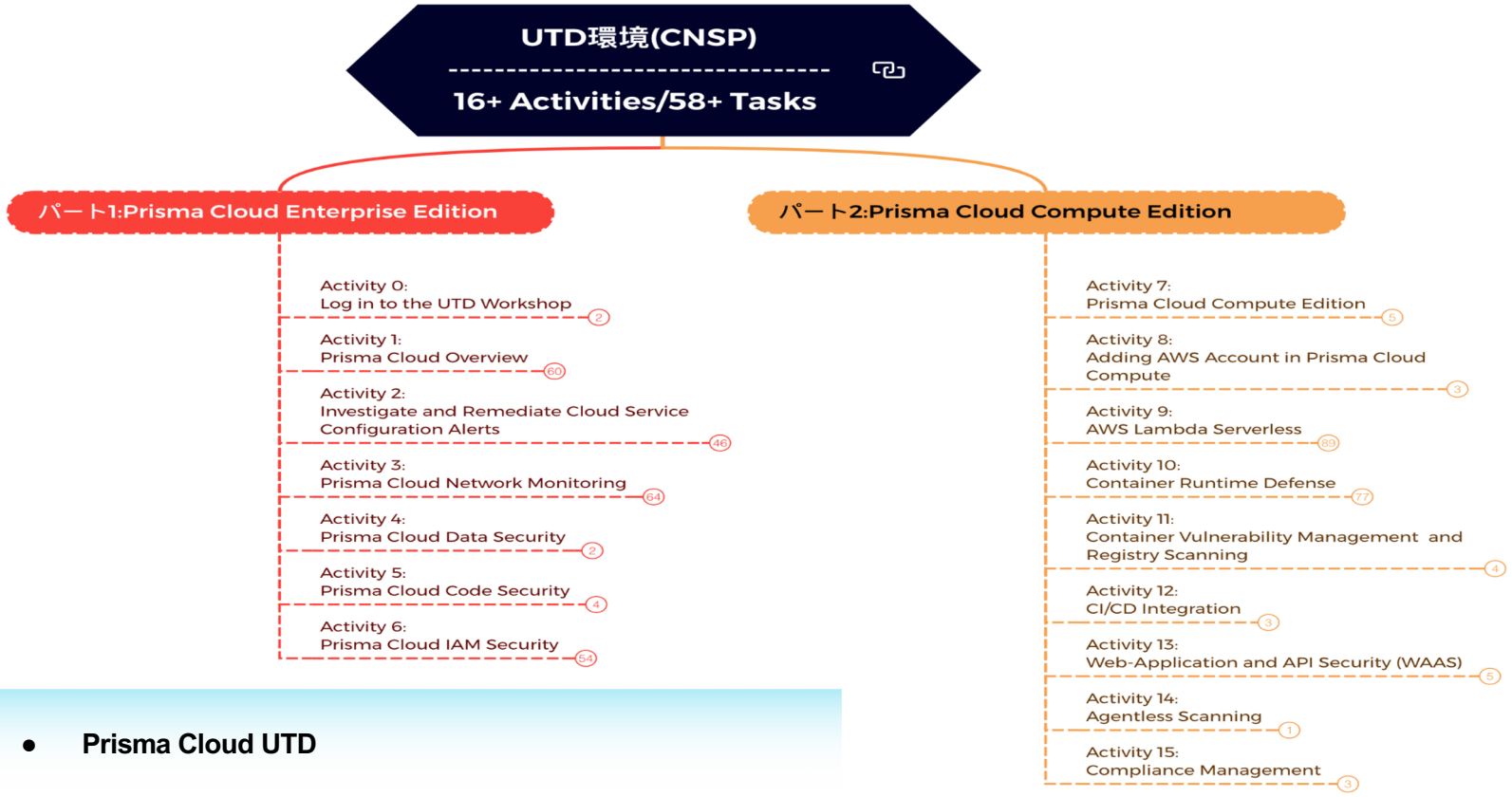


Next Actions

Prisma Cloud製品
Next ActionとしてUltimate Test Drive(UDT)を推薦する

Prisma Cloud体験手法	Pros	Cons
パロアルトネットワークス社員による全般デモ	お客様側に何の準備しなくても済む	Prisma Cloudの機能の理解が深まらない可能性があり
お客様と一緒にハンズオン (UTD環境)	実際に環境を使いハンズオンする事によってPrisma Cloudの機能の理解が深まる	体験系統に応じてお客様に準備する必要があり。基本的に何も準備する必要が無い
自由度の高いPrisma Cloud環境の提供(POC環境)	Prisma Cloudの機能の理解が高度に深まる	お客様側に実践してもらう事が多い

Next ActionとしてUDTを推薦する: Prisma Cloud UTD



- **Prisma Cloud UTD**

Cloud Risk Assessment

paloalto NETWORKS

Prisma Cloud製品
Cloud Risk Assessmentとは

Prisma Cloud を使用したモニタリング結果を基に、ご利用中のクラウド環境の各種セキュリティリスクについて可視化レポートをご提供致します。

現在運用中のクラウド環境の状況を監査・可視化、潜在的なインシデント要因・影響度を把握頂き、検知されたセキュリティリスクへの対策の検討、ご導入にお役立て頂けます。

アセスメント対象期間：
2022年XX月xx日 00:00～ 2022年XX月xx日 00:00（調整可能）

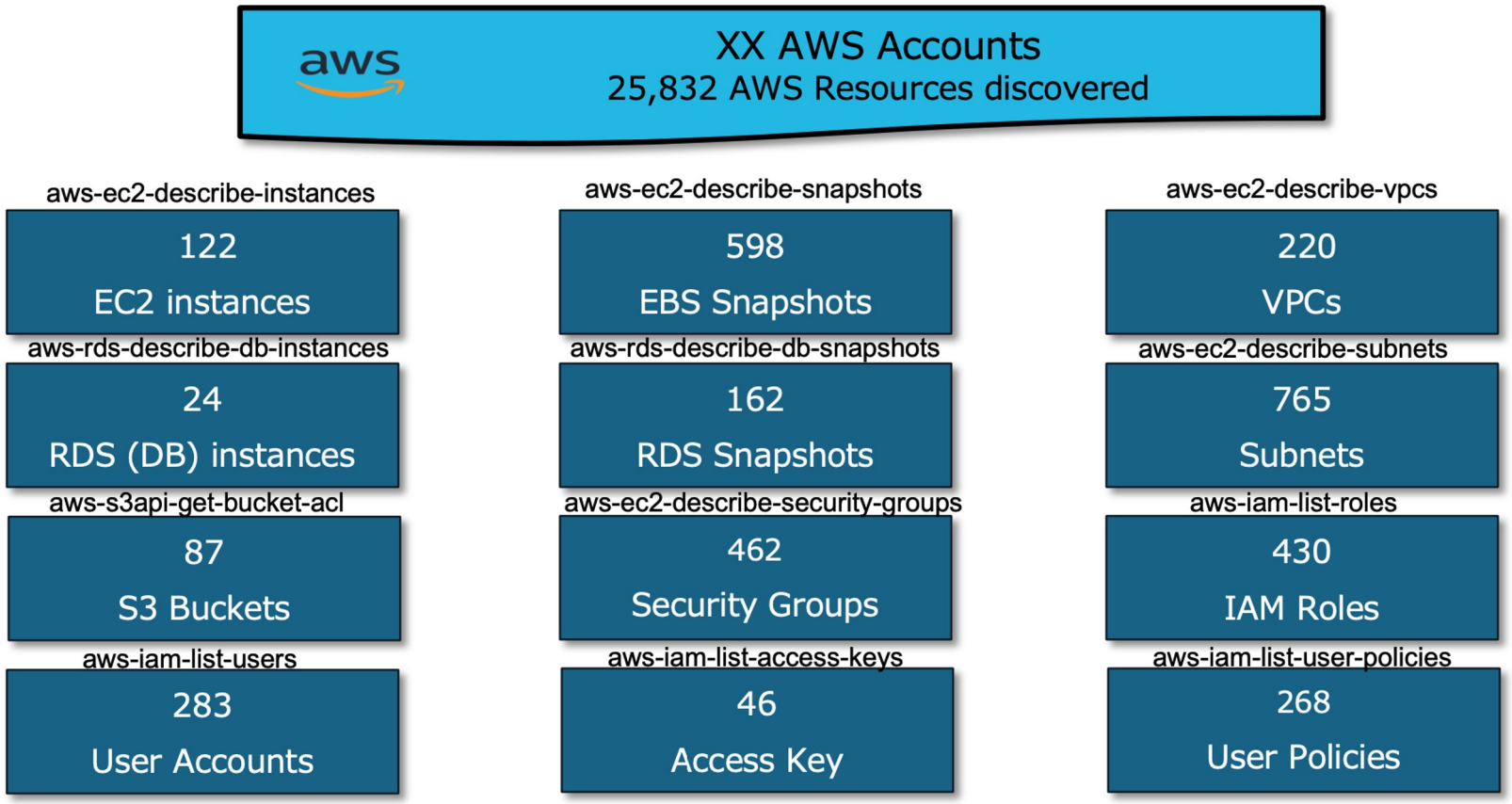
監査対象クラウド環境：
☑AWS　☐Azure (TBD)　☐Google Cloud (TBD)

監査に伴う必要な事前設定：
監査対象のアカウントをPrisma Cloud のAWSアカウントにオンボーディング （Assuming Roleを利用したアカウント間連携によるクラウド内の各種情報参照・取得の許可）して頂きます。

注意：
貴社の全ての環境・リソースに対して確認を行うことは困難であり、クラウドアカウント監査のサンプルとしてご使用下さい。

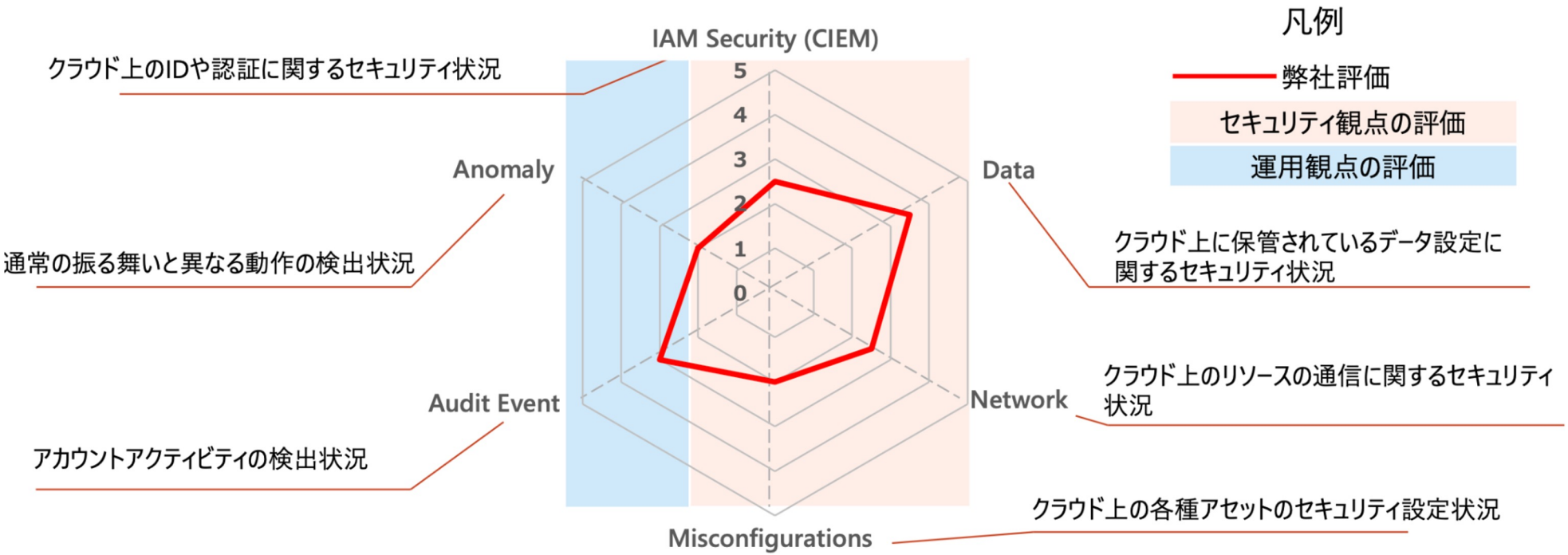
Prisma Cloud製品

Cloud Risk Assessmentのレポートサンプル:検出されたクラウドリソース



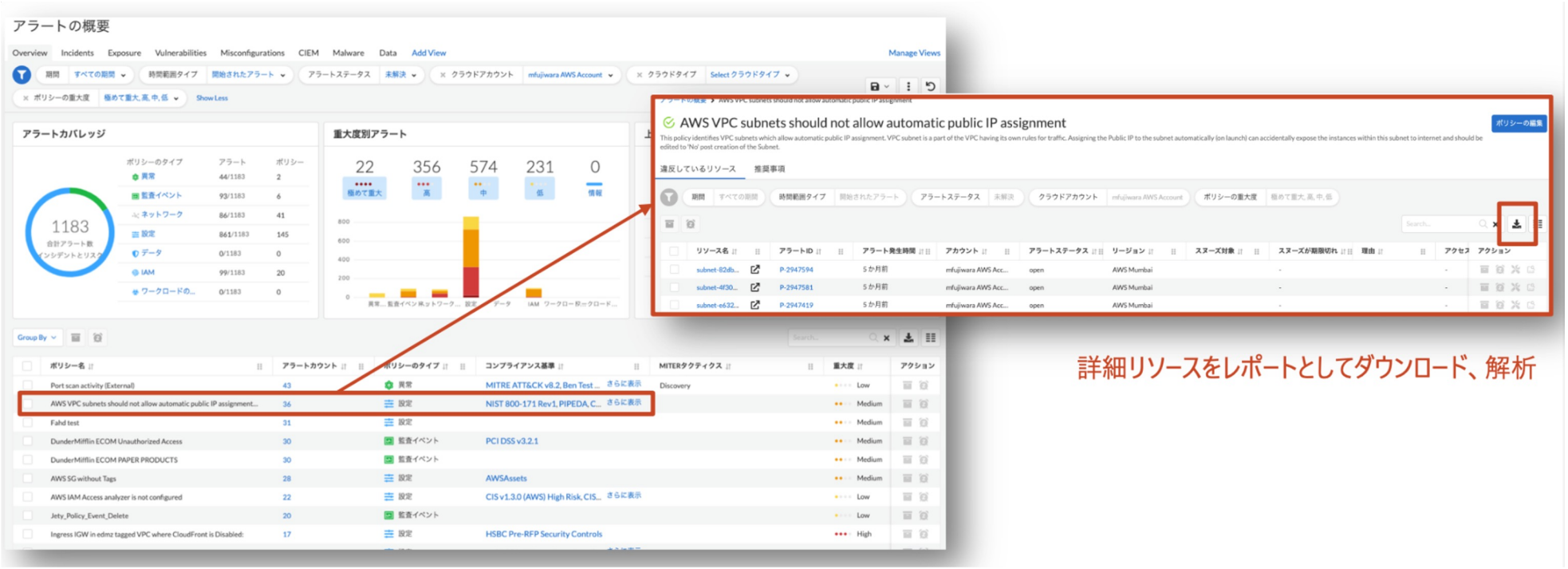
Prisma Cloud製品

Cloud Risk Assessmentのレポートサンプル:エグゼクティブサマリー



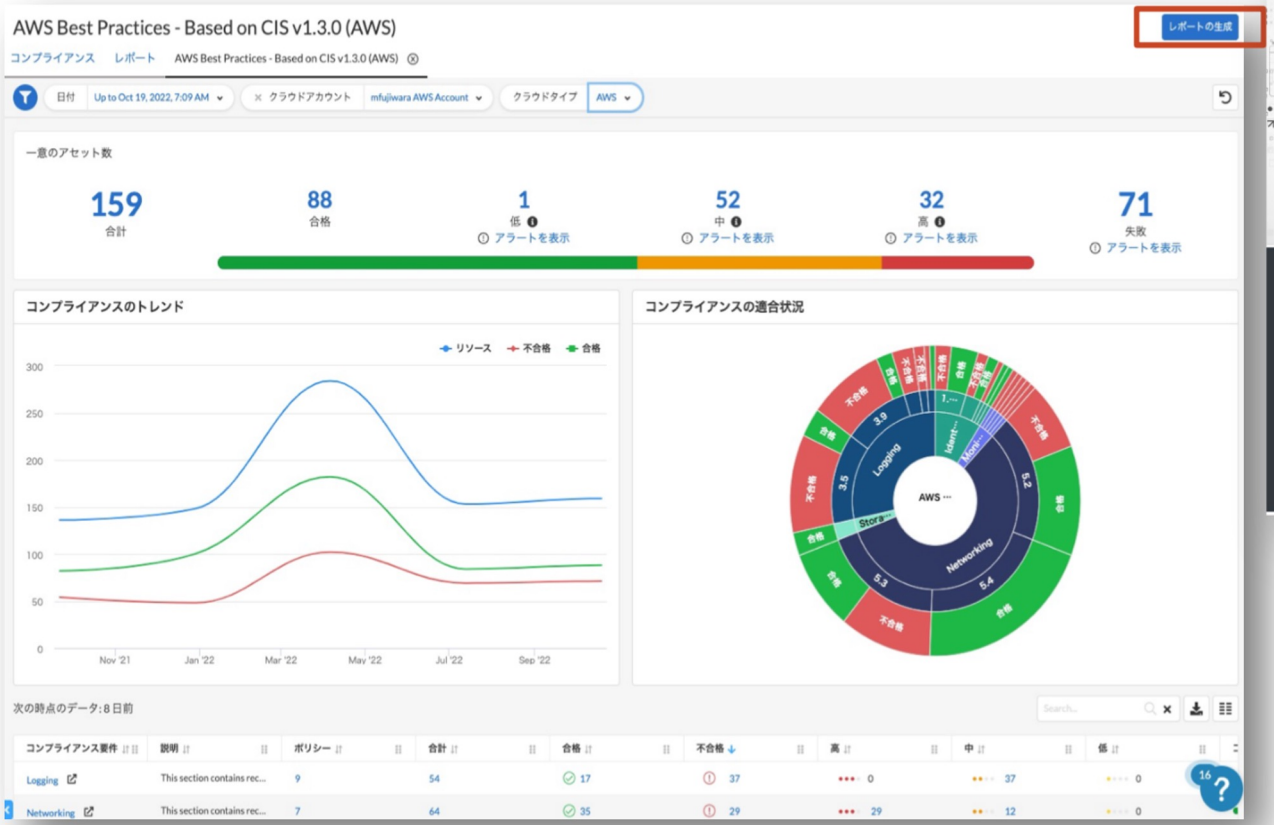
Prisma Cloud製品

Prisma Cloudによって検知された各種アラート、対象リソースを解析



Prisma Cloud コンソール上のアラート報告イメージ

指定したコンプライアンス標準での監査・レポート



paloalto NETWORKS

Thank you

paloaltonetworks.com