# Prevention, Detection, and Response for Security Operations

REFERENCE ARCHITECTURE GUIDE

APRIL 2020

paloalto® NETWORKS

# Table of Contents

# Preface

## GUIDE TYPES

*Overview guides* provide high-level introductions to technologies or concepts.

*Reference architecture guides* provide an architectural overview for using Palo Alto Networks® technologies to provide visibility, control, and protection to applications built in a specific environment. These guides are required reading prior to using their companion deployment guides.

*Deployment guides* provide decision criteria for deployment scenarios, as well as procedures for combining Palo Alto Networks technologies with third-party technologies in an integrated design.

## DOCUMENT CONVENTIONS

*Notes* provide additional information.

*Cautions* warn about possible data loss, hardware damage, or compromise of security.

Blue text indicates a configuration variable for which you need to substitute the correct value for your environment.

In the **IP** box, enter 10.5.0.4/24, and then click **OK**.

**Bold text** denotes:

- Command-line commands.

  ```
  # show device-group branch-offices
  ```

- User-interface elements.

  In the **Interface Type** list, choose **Layer 3**.

- Navigational paths.

  Navigate to **Network > Virtual Routers**.

- A value to be entered.

  Enter the password **admin**.

*Italic text* denotes the introduction of important terminology.

> An *external dynamic list* is a file hosted on an external web server so that the firewall can import objects.

Highlighted text denotes emphasis.

> Total valid entries: 755

## ABOUT PROCEDURES

These guides sometimes describe other companies' products. Although steps and screen-shots were up-to-date at the time of publication, those companies might have since changed their user interface, processes, or requirements.

## GETTING THE LATEST VERSION OF GUIDES

We continually update reference architecture and deployment guides. You can access the latest version of this and all guides at this location:

https://www.paloaltonetworks.com/referencearchitectures

## WHAT'S NEW IN THIS RELEASE

Palo Alto Networks made the following changes since the last version of this guide:

- This is a new guide.

# Purpose of This Guide

The guide describes how organizations can address the challenges in preventing threats where possible and managing alerts more efficiently. It covers security operations (SecOps), their challenges, and how Palo Alto Networks provides a security architecture for prevention, detection, investigation, orchestration, automation, and response.

## AUDIENCE

This guide is for technical readers—including solution architects, security engineers, and security support staff—who want to understand best practices for preventing, investigating, and responding to threats. It assumes the reader is familiar with the basic concepts of threat prevention, networking, and security operations, as well as possessing a basic understanding of automation, machine learning, and analytics.

## RELATED DOCUMENTATION

The following documents support this guide:

- Artificial Intelligence and Machine Learning in the SOC: Overview—Introduces the concepts of artificial intelligence (AI) and machine learning (ML) and how you can leverage them in the security operations center (SOC)

# Introduction

Organizations continually receive security-related threats. These attacks are on the increase and are top-of-mind for most organizations. Due to the increase of threats and alerts, some organizations struggle with having the right tools or processes in place to properly investigate and respond to these threats, as well as having enough visibility into what is occurring in their environment. To address the security threats, many organizations are investing in *security operations* (SecOps) and *security operations centers* (SOCs). SecOps is a joint effort between IT security and operations teams, working together to identify, investigate, and mitigate threats. These teams defend their organization's infrastructure and data from harm, such as losing data or the ability to operate and function as a business.

Larger organizations might operate a SOC, where dedicated security staff use tools to detect, investigate, report, and respond to threats. SOCs determine the extent of the threat through analysis and threat-hunting techniques. Some organizations might not have a dedicated SOC but still require security staff to be able to investigate and respond to threats. Depending on the size of the SecOps organization, roles can consist of Tier 1-3 analysts, threat intelligence specialists, and hunting specialists.

Primary among the challenges SecOps faces is *alert fatigue,* or being overwhelmed by alerts. Contributing factors include:

- **Security analyst resources**—Lack of available workforce with cybersecurity expertise makes it hard to recruit the skills and knowledge that organizations need.
- **Non-integrated tools**—As the security industry has evolved, many organizations have collected siloed network security tools lacking integration capabilities, such as endpoint detection and response (EDR), network traffic analysis (NTA), and user and entity behavior analytics (UEBA). As a consequence, analysts must move from console to console, or they must manually query and correlate data in order to understand the user, the host, the application traffic, the endpoint process, and the threat intelligence details related to an attack.
- **Threat response time**—The more analysts must manually respond to alerts, the further behind they get. If analysts have a lack of context when dealing with multiple platforms that are generating alerts, SecOps has to manually integrate multiple data sources and tools to understand the attack. This often leads to delay in assessing and responding to important threats.
- **Manual threat response**—Many organizations respond to threats manually, inevitably wasting analysts' time on false positives. Automation, machine learning, and AI help to reduce the time spent by SecOps teams investigating and correlating events from multiple data sources.

Efficient case management, collaboration, and the ability to assign cases to the right SecOps resource is paramount in providing speedy response to threats. Therefore, common asks from SecOps teams are to reduce the number of alerts, to have tools that easily identify and investigate threats, and to shorten the time it takes to contain a breach.

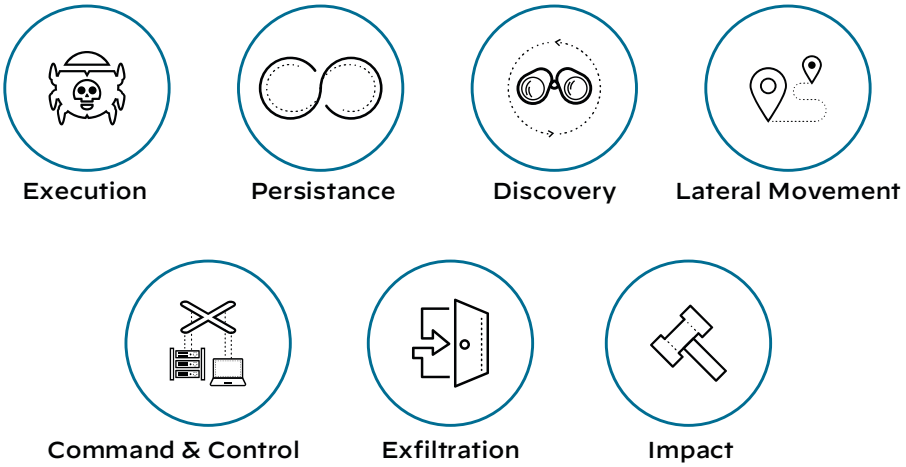
Multiple resources help SecOps categorize and identify threats, behaviors, and techniques. A well-known organization that provides detailed information surrounding attacks and details on cyberattack tactics and techniques is MITRE, which manages the MITRE ATT&CK knowledge base.

## MITRE ATT&CK

Palo Alto Networks leverages the MITRE ATT&CK knowledge base in order to identify tactics related to alerts detected in your environment.

MITRE is a not-for-profit organization that developed MITRE ATT&CK, a global knowledge base of real-world adversary attacks and techniques. ATT&CK stands for *Adversarial Tactics, Techniques, and Common Knowledge*. MITRE ATT&CK provides the cybersecurity community with threat models and methodologies. The knowledge base is open to all organizations and individuals.

*Figure 1 Subset of MITRE tactics*




The MITRE ATT&CK framework describes how adversaries penetrate networks and then move laterally, escalate privileges, and evade defenses. ATT&CK looks at the problem from the perspective of the adversary—what goals they are trying to achieve and what methods they use.

MITRE views persistent threats through tactics, techniques, and mitigations in order to better understand and document their behaviors and methods. Adversary behaviors are organized into *tactics*, which are the objectives the adversary wants to achieve, such as defense evasion, lateral movement, and exfiltration. Tactics represent the "why" of an attack. *Techniques* represent the "how" and the "what" of an attack. MITRE's approach of viewing persistent threats, assists in creating analytics to detect adversary behavior and how they interact with systems such as domains, IP addresses, file hashes, and registry keys. *Mitigations* are the steps you take to detect and deal with attacks.

To better understand the types of attacks and threats your organization might be facing, cyber-threat intelligence requires knowing what your adversaries do. Threat groups are organized as *advanced persistent threats* (APTs) and are categorized. You can use the MITRE website to search for a specific APT to understand the group, techniques, and software they have used to launch an attack. For example, APT28 is a threat group also known as *Fancy Bear*, which is a known Russian cyber-espionage group.

This Reference Architecture guide covers four cyber-threat attack use cases:

- **Malware**—A malicious program, file, or code aimed at causing harm or disruption to computer systems
- **Credential abuse**—Using those compromised user credentials and passwords to access the network, authenticate applications, or steal data
- **Lateral movement**—An adversary accesses one asset within a network and then spreads their reach from that device to others within the same network
- **Data exfiltration**—The unlawful copying, transfer, or access of network data

Each use case details the Palo Alto Networks capabilities and approach for prevention, detection, investigation, and response. But first, the next section provides a high-level overview of the Palo Alto Networks approach to SecOps.

## PALO ALTO NETWORKS APPROACH TO SECOPS

The Palo Alto Networks approach to SecOps is to enable the SOC to be more successful in preventing attacks, to detect and investigate events rapidly, and to automate response with security orchestration.

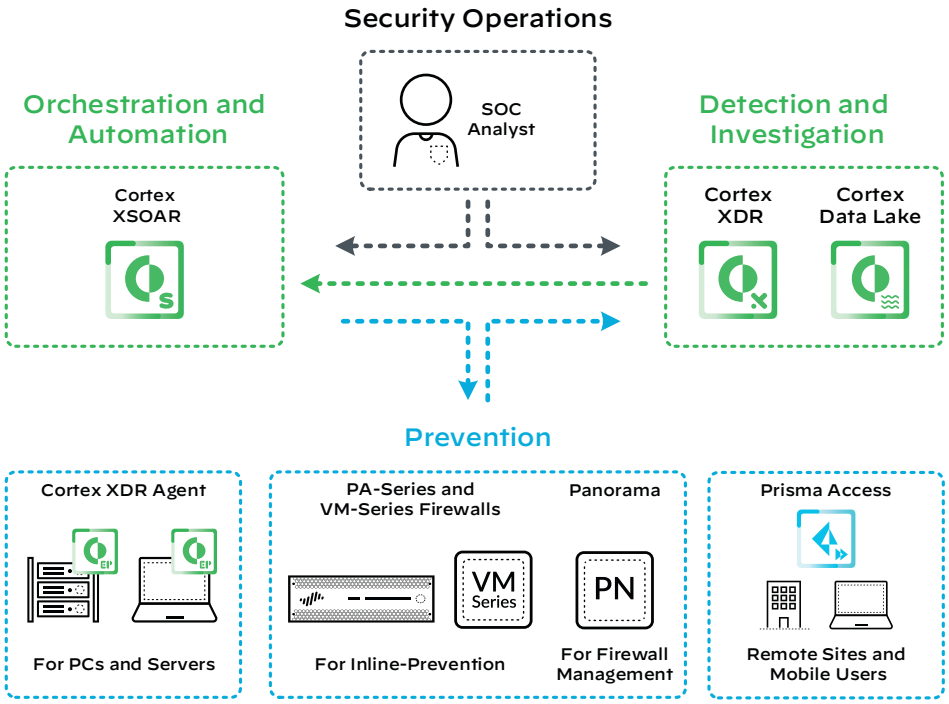
Cortex™ is the platform for standardizing and scaling SecOps. The Cortex main components include Cortex Data Lake, Cortex XDR, Cortex XDR agent, and Cortex XSOAR. Think of Cortex as your one-stop shop for SecOps, solving all key challenges in a more efficient way with improved security outcomes. With Cortex, you can speed up investigations by having the right data, integrated across the network, endpoints, and cloud, with all the context needed for security analysis.

In the SOC, the Palo Alto Networks approach is three-pronged:

1. First, you prevent all of the threats you can. You achieve this with Palo Alto Networks inline and endpoint security.
2. Next, you need to rapidly detect and investigate the threats you can't prevent. You achieve this with Cortex XDR, which identifies, with high fidelity, the security incidents the SOC must investigate and to which they must respond.
3. Then you automate future responses. Using Cortex XSOAR (formerly *Demisto™*), security orchestration allows you to execute automatable playbooks for rapid incident response. *Playbooks* are task-based graphic workflows that allows you to standardize and scale a response process by coordinating tasks across people, processes, and technology.

Cortex is Palo Alto Network's AI-based, continuous-security platform, which continually evolves to help the SOC stop the most sophisticated threats. Cortex XDR stitches all of your managed endpoints, network, and cloud events together. Because these various stacks witness only part of the buildup of an attack, Cortex XDR uses AI to learn which events in one system drive attention to another system. Using AI and ML, Cortex also identifies unknown and evasive threats that are targeting your network.

*Figure 2 The Palo Alto Networks approach*




## Prevention

Leveraging a preventive approach in combination with Cortex provides a much stronger security posture. You can prevent the vast majority of threats first and reduce the number of threats with which the SOC must deal.

You need to have consistent enforcement of your security policy across the network, the cloud, and endpoints for better control of your security posture. Using next-generation firewalls, such as Palo Alto Networks PA-Series firewalls, VM-Series firewalls, and Prisma™ Access (formerly *GlobalProtect™ cloud service*), you can prevent threats through the consistent utilization of the built-in threat-prevention capabilities that provide inline security to the network.

On the endpoint, you can enable Cortex XDR agent endpoint protection to secure Windows, macOS, Android, and Linux endpoints, which provides behavior-based protection that detects and responds to sophisticated attacks. The Cortex XDR agent prevents malware infections by blocking the exploits used by adversaries to compromise endpoints and install malware. By identifying exploit techniques, not just exploit signatures, the Cortex XDR agent protects hosts and stops zero-day threats.

The prevention-based architecture exists outside of the SecOps function itself but affects the efficiency and effectiveness of achieving the SOC mission. Without a prevention-based architecture, analysts can be overwhelmed with false positives and low-fidelity data.

## Detection and Investigation

For every threat you can't prevent, you need to be able to detect and investigate it rapidly. Having the ability to triage alerts from multiple data sources such as the network, cloud, and endpoints helps by providing context to threats, which facilitates the investigation and response.

Cortex XDR provides complete visibility, intelligent alert grouping, ML-based detection, automated root-cause analysis, and a graphical chain of events. The chain of events is created via the Causality Analysis Engine, which corelates activity from all detection sensors and establishes causality chains that identify the alert's root cause. After identifying the root cause of an alert, you can perform immediate response actions.

The right log data is critical for the SOC. The Cortex XDR analytics engine retrieves logs from Cortex Data Lake (formerly *Logging Service*), which is a repository for all logged events from the network, cloud, and endpoints. Cortex Data Lake provides cloud-based storage and allows you to simplify your security operations by collecting, integrating, and normalizing your enterprise's security data.

Cortex XDR provides visibility in order to reduce security breaches, accurate attack detection, simplified investigations, rapid response, increased SecOps efficiency through fewer tools, and better context through root-cause analysis.

## Response with Orchestration and Automation

After you have the right data from multiple sources and have visibility and context, you can then automate responses. Cortex XSOAR allows security teams to ingest alerts across multiple sources and then execute automatable playbooks for rapid incident response. Cortex XSOAR playbook integrations automate tasks across products, while also allowing for human oversight and interaction. Depending on your requirements, playbooks can be automated, manual, or a mixture of both.

Cortex XSOAR has case-management and collaboration capabilities built into the product to ease communication and learning. Cortex XSOAR can ingest alert data across a range of sources, including security information and event management (SIEM) systems, network security tools, email inboxes, vulnerability management tools, and cloud-security solutions. You can intuitively search and query these alerts according to various parameters, allowing you to process the data as needed. You can assign custom service level agreements (SLAs) and link them to playbook execution, resulting in an accurate measurement of your key performance indicators.

Cortex XSOAR facilitates real-time investigation through the *War Room*, a shared space where SOC team members can collaborate and remotely execute actions. The War Room records all user actions at one source and, from a common console, allows team members to quickly run specific commands relevant to incidents in the network. All team members have full task-level visibility of the process and are able to run and document commands from the same console, eliminating the need to gather information from multiple sources for documentation purposes.

# Prevention

Organizations face attacks by adversaries around the globe who are looking to profit or to cause malicious damage. Through use of techniques such as packet obfuscation, polymorphic malware, encryption, multiphase payloads, and DNS, adversaries launch high-volume, stealthy, and sophisticated attacks while remaining invisible to traditional defenses within the organization. Therefore, it is important that organizations adopt a preventive approach that provides multiple layers of prevention in order to confront threats at each stage of their attack.

Prevention might exist outside normal SecOps function, but it affects the ability of the SOC to effectively and efficiently protect the organization. Without a prevention-based architecture, the SOC is overwhelmed with false positives and low-fidelity data. This section covers the key Palo Alto Networks components for inline security and endpoint protection that provides a preventative approach.

The ideal security outcome is to prevent all attacks. Making the change to a modern, prevention-based architecture is of strategic importance for organizations to ensure that they can prevent most attacks before they occur. Application and user visibility, optimized and hardened security policies, threat prevention services, segmentation, and decryption policies are all key building blocks for providing a secure posture.

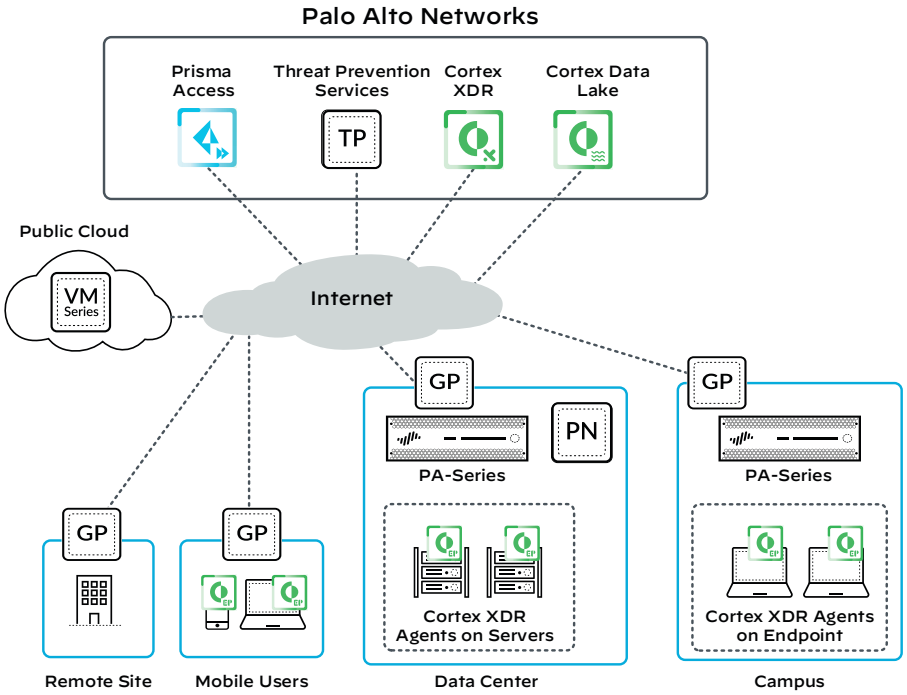
For a preventative approach, an organization needs to reduce the attack surface and provide consistent protection across the network, cloud, and endpoints. Centralized management of security controls and devices is critical to providing consistency and reducing administration time.

As this guide discusses the threat use cases, it highlights how Palo Alto Networks provides a feature-rich and comprehensive offering for both inline security and endpoint security in order to prevent threats. To solve each use case, the full suite of Palo Alto Networks solutions is leveraged, including PA-Series physical firewall appliances, VM-Series virtual firewalls for private and public cloud deployment, and *Prisma Access*, which is a cloud-delivered firewall service. Panorama™ provides centralized network security management, including device and policy management.

Palo Alto Networks offers Cortex XDR endpoint protection to safeguard your endpoints across multiple platforms. It combines industry-best AI and behavior-based protection to block advanced malware, exploits, and fileless attacks.

Figure 3 shows a basic network architecture, demonstrating inline security with the PA-Series and VM-Series firewalls in the cloud, in the data center, and on campus. Delivered as a cloud service, Prisma Access is securing inline traffic flows for users and remote sites. The Cortex XDR agents provide protection for all endpoints, including PCs, servers, and mobile devices. GlobalProtect client software runs on end users' systems and allows access to network resources via the GlobalProtect portals and gateways on the firewalls or via Prisma Access. Cortex Data Lake logs all firewall and endpoint data.

*Figure 3 Prevention-based approach*




## INLINE PROTECTION FOR THE NETWORK AND CLOUD

To safeguard your environment, you need a multi-layered approach. You need inline protection on top of endpoint protection to stop outbound traffic from going to undesirable destinations and to stop threats that are inbound to your devices. Not all devices have endpoint protection, such as printers and IoT devices, and can be exploited if they aren't protected by next-generation firewalls that inspect traffic in and out of the network and across network zones.

### Next-Generation Firewall

For inline protection, the Palo Alto Networks next-generation firewall comes in physical (PA-Series) and virtual (VM-Series) form factors as well as a cloud service (Prisma Access), which provides secure access to internet and business applications hosted in SaaS, a corporate data center, or public cloud providers.

Organizations deploy the next-generation firewall at the network perimeter and inside the network at logical trust boundaries. All traffic crossing the next-generation firewall undergoes a full-stack, single-pass inspection, which provides complete context of the application, any associated content, and the identity of the user. With this level of context, you can align your security policy with your business initiatives.

The next-generation firewall enables complete visibility of the application traffic flows, associated content, and user identity, and it protects devices from known, unknown, and advanced persistent threats.

## Prisma Access

Prisma Access is a cloud-delivered, secure access service edge solution for branch offices, retail locations, and mobile users. Prisma Access enables you to secure access, protect users and applications, and control data—from anywhere. Prisma Access uses a cloud-based infrastructure, allowing you to avoid the challenges of sizing firewalls and computing resource allocation. This minimizes the risk of coverage gaps or inconsistencies across the distributed organization.

Prisma Access provides visibility into the use of applications on the network and provides the ability to control users' access to those applications. There are two options available as part of Prisma Access:

- **Prisma Access for networks**—Provides cloud-based security services, including App-ID™ and threat prevention, for your remote networks, safely enabling commonly used applications and web access. You connect remote networks to Prisma Access via an industry-standard IPSec VPN-capable device.
- **Prisma Access for users**—Provides cloud-based security services, including App-ID and threat prevention, for mobile users. Cloud deployment provides an alternative to the traditional, on-premises deployment of GlobalProtect.

## Security Policy

The security policy is important for prevention because it's the building block of a security posture to apply multiple traffic-filtering capabilities on the next-generation firewall. The threat examples covered later leverage components that are part of the security policy rules for inspecting traffic and denying or accepting it.

Palo Alto Networks firewalls are stateful, and all traffic passing through the firewall is matched against a session, which is then matched against a security policy rule. Policy rules allow you use multiple match criteria to control network traffic, including the zone, source address, destination address, port, application, URL category, source users, and host information profile.

Security policy rules consists of three type of rules: intrazone, interzone and universal. An *intrazone* rule applies to all matching traffic within a specified source zone and does not apply to traffic between different zones. An *interzone* rule applies to all matching traffic between specified source and destination zones but does not apply to traffic within the same zone. A *universal* rule applies to all matching traffic, both intrazone and interzone.

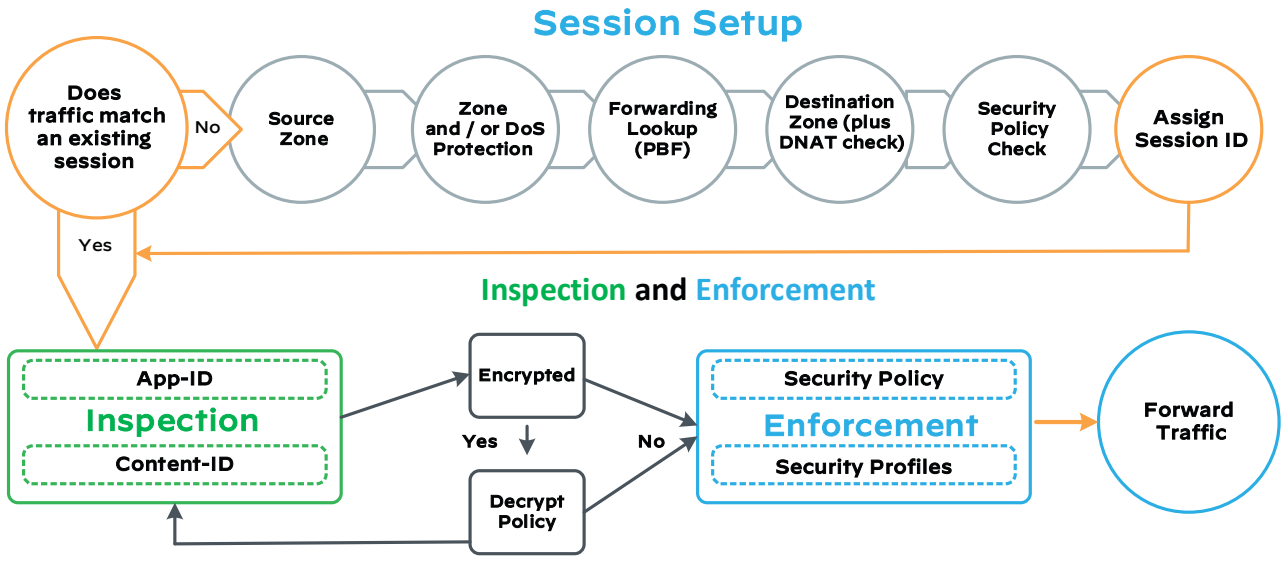
The firewall, by default, has implicit and explicit rules, such as allowing intrazone traffic and denying interzone traffic. It is best practice to define all your security rules after understanding your application and traffic flow requirements and then change the default behavior of the intrazone rule from allow-all to deny-all.

Logging should be added to all policy rules. Session logging is a useful troubleshooting tool that provides visibility and enables debugging of security policy rules. For logging, Palo Alto Networks offers a cloud-based logging service called *Cortex Data Lake*. After you start sending logs to Cortex Data Lake, you can start leveraging Cortex apps that analyze and report on your network, cloud, and endpoint data.

When configuring logging in a policy rule, for visibility, you should set the policy rule to log at session end, but for debugging on an as-needed basis, you can set the policy rule to log at session start. For regular logging, the best practice is to log at session end. When configuring policy rules, minimize the use of "any" in the columns, when possible, because granularity and specificity help prevent gaps in the security posture. After policy rules are active, the policy rule hit counts show how many hits each policy rule is getting.

Figure 4 shows traffic flow for a new or existing session through the firewall.

*Figure 4 Next-generation firewall traffic flow logic*




Multiple Palo Alto Networks threat-prevention capabilities are built into security policies in the form of *security profiles*. Security profiles are attached to security policy rules to provide capabilities such as antivirus, URL-Filtering and file blocking for example. The next-generation firewall also includes threat-prevention subscription services for added security. These features are covered in more detail in the individual threat case examples.

## ENDPOINT PROTECTION

Over and above inline protection with PA-Series firewalls, VM-Series firewalls, and Prisma Access, SecOps must also be able to secure endpoints and stop previously unknown threats. Threats can occur behind an inline firewall; users could possibly insert a compromised USB flash drive into their PC or bring in their own devices—such as smart phones, tablets, and laptops—that could be infected with malware. To overcome this, you need advanced endpoint protection.

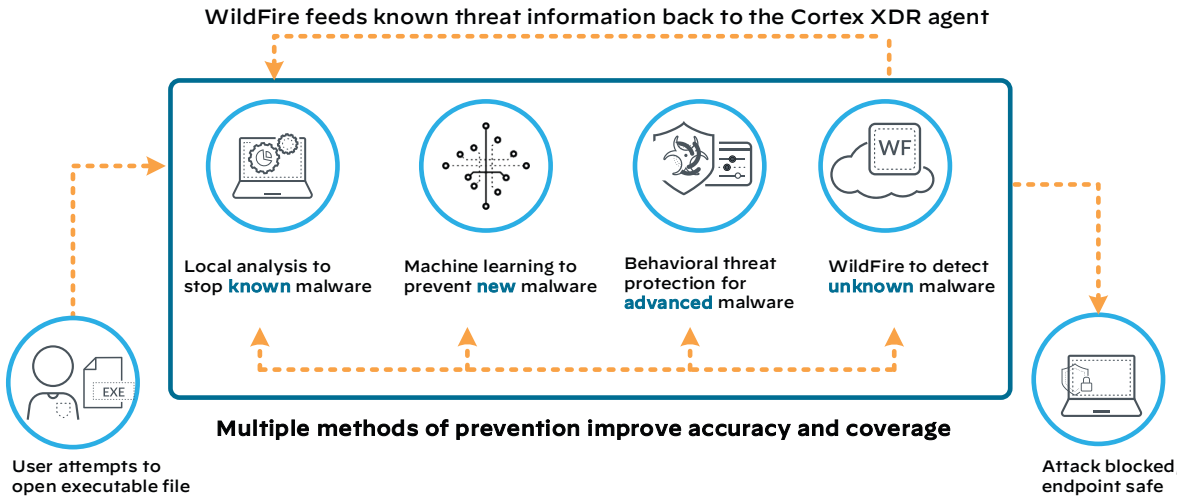
Palo Alto Networks provides the Cortex XDR agent (formerly *Traps™*) for endpoint protection in order to stop threats on the endpoint and coordinates enforcement with cloud and network security in order to prevent successful cyberattacks. The Cortex XDR agent minimizes endpoint infections by blocking malware, exploits, and ransomware. The agent prevents exploits by stopping the techniques themselves, requiring no prior knowledge, and stopping entire classes of attacks.

The Cortex XDR agent comes with two offerings—*Cortex XDR Prevent* for complete endpoint protection for the endpoint only and *Cortex XDR Pro* with comprehensive prevention, investigation, and response, which includes all the functions of Cortex XDR for the network, endpoint, and cloud.

The Cortex XDR agent provides multi-method protection, including local systems, local analysis, behavioral threat protection, and WildFire® in the cloud:

- **Local analysis via machine learning**—Local analysis runs on the endpoint and uses machine learning to keep false positives low. To achieve maximum protection, the Cortex XDR agent, trained by millions of known-good and known-malicious files, uses local machine learning to identify new threats. The agent sends all identified, new threats to the WildFire threat-intelligence cloud service for additional analysis.
- **Behavioral threat protection**—Sophisticated attacks use multiple legitimate applications and are much harder to detect. It's important to understand what is happening on the local endpoint, which is why behavioral threat detection is needed in order to monitor for malicious sequences of events across all processes, with the ability to terminate them when attacks are detected.
- **WildFire**—To provide faster verdicts and responses, the Cortex XDR agent leverages WildFire to quickly determine if the threat has previously been seen. *WildFire* is a cloud-based threat analysis service that acts as a threat-intelligence sandbox in the cloud. All unknown threats are converted to known threats, and the threat-intelligence data is fed back to the Cortex XDR agent as well as other Palo Alto Networks security products.

*Figure 5 Multi-method prevention*




The Cortex XDR agent is managed through the Cortex XDR app, where you can create agent installation packages, manage endpoints, create policy rules, and create profiles. You can create profiles for malware and exploits.

Now that you have an overview of the Palo Alto Networks components that are used for prevention, the next few sections cover the inline and endpoint threat-prevention capabilities for four different use cases—malware, credential abuse, lateral movement, and data exfiltration.

## PREVENTING MALWARE

*Malware* is typically a malicious program, file, or code aimed at causing harm or disruption to computer systems. Malware can be delivered in multiple forms, such as phishing attacks within the body of an email, delivery via USB drive, and downloads on a website, to name a few. Common types of malware include viruses, adware, rootkits, spyware, ransomware, trojan horses, worms, and key loggers.

Sophisticated malware attacks can use command-and-control (C2) servers that enable adversaries to communicate with infected systems, remotely control the system, and even exfiltrate data. Malware can be found as part of many PRE-ATT&CK techniques at MITRE. For more information, see MITRE Tactic TA0002, "Execution" (described as "the adversary is trying to run malicious code").

Malware can often use evasion techniques to avoid detection and security controls. Malware executables can be hidden in compressed or encrypted files, and it can deploy polymorphic techniques to constantly change the way it looks. In some instances, malware can deploy in a *staged approach*, where installed programs (downloaders, droppers) learn about the local environment before deploying specific malware.

To prevent malware, you stop all of your devices from going to undesirable places, such as insecure websites where they can be infected, and you stop any malware coming into your environment before it reaches any devices. If inline protection is not in the path to detect or block malware, such as a USB drive or a device introduced behind the firewall, you must be able to stop malware at the endpoint. Therefore, you need inline and endpoint protection in order to prevent malware.

The next section focuses on how you can prevent malware inline and on endpoints, using a prevention-based approach.

## Inline Protection from Malware

To provide inline protection from malware, you need to decrypt traffic (allowing the visibility needed to detect malware, which could be hidden in encrypted sessions), control which applications and websites users can access, and block known and unknown malware. The PA-Series firewalls, VM-Series firewalls, and Prisma Access have security features built in to prevent attacks such as malware. Some of these features include decryption, App-ID, User-ID™, URL Filtering, security profiles, and threat-prevention subscriptions.

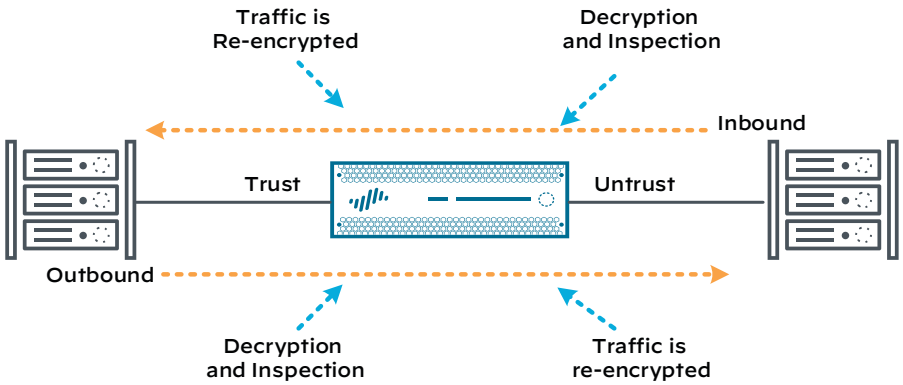
### Decryption for Visibility into the Traffic

Malware can be hosted anywhere, not just within known risky applications. It is therefore important that you are able to inspect all allowed traffic to ensure it does not contain malware. Because malware can be disguised in encrypted traffic, it is important to be able to decrypt traffic inline. Users can access a website that is hosting malware that they then download. You can't protect your network against threats you can't see, so decryption gives you complete visibility of all traffic entering and leaving your network.

After the traffic you specify in your decryption policy is decrypted, it is then subject to security policy enforcement to enable users to access the right data and applications while preventing known and unknown threats. The traffic is then re-encrypted before it exits the firewall.

Decryption capabilities allow you to enforce policies on encrypted traffic so that the firewall can prevent malicious, encrypted content from entering the network and prevent sensitive, encrypted data from leaving your network.

*Figure 6 Traffic decrypted to block malware*



For malware, the PA-Series firewalls, VM-Series firewalls, and Prisma Access have an SSL Forward Proxy Decryption policy rule. This policy rule decrypts and inspects SSL/TLS traffic from internal users to the web. By decrypting the traffic so that the firewall can apply decryption profiles, security policies, and security profiles to the traffic, SSL Forward Proxy decryption prevents malware that is masked as SSL-encrypted traffic from being introduced into your network.

In order to secure an SSL/TLS connection, SSL Forward Proxy decryption requires certificates to establish the firewall as a trusted third party and to establish trust between a client and a server.

Not all traffic should be decrypted, such as that protected by privacy laws or sensitive data. With decryption, you should decrypt everything except all sensitive and legally protected traffic. Decryption profiles are very granular, and you can create exception rules for specific zones, destination IPs, users, and URL categories.

## Blocking Risky Applications

To limit the exposure to known malware, you need to limit the risky applications that are being used, as well as limit the websites users can visit. Many organizations are not fully aware of which applications are in use, how frequently they are used, or who is using these applications. Previously, security policies for unapproved applications left you with two choices—either block everything in the interest of data security or enable everything in the interest of business. These choices left little room for flexibility.

You can strengthen security policies and reduce incident response times by knowing who is using each of the applications on your network and by knowing who might have transmitted a threat or is transferring files. On the PA-Series firewalls, VM-Series firewalls, and Prisma Access, you have the ability to leverage User-ID and App-ID to allow or block applications on the network, as well as have better visibility of the applications in use on your network.

User-ID enables you to leverage user information stored in a wide range of repositories. User-based policy controls can include application information, including its category and subcategory, its underlying technology, and its application characteristics. You can define policy rules in outbound or inbound directions in order to safely enable applications based on users or groups of users.

App-ID uses multiple identification techniques to determine the exact identity of applications traveling through your network, including those that try to evade detection by disguising themselves as legitimate traffic, by hopping ports, or by using encryption. App-ID is used in combination with security polices to match based on actual applications rather than port numbers.

App-ID even determines the difference between base applications and application functions. This level of visibility brings a complete understanding of the applications on your network and their value and risk to your organization. App-ID uses multiple techniques to identify traffic, including application signatures, TLS/SSL and SSH decryption, application and protocol decoding, and heuristics. The list of App-IDs is dynamically updated monthly, with new applications added based on market trends and on input from the Palo Alto Networks community (customers and partners).

*Figure 7 List of App-IDs*

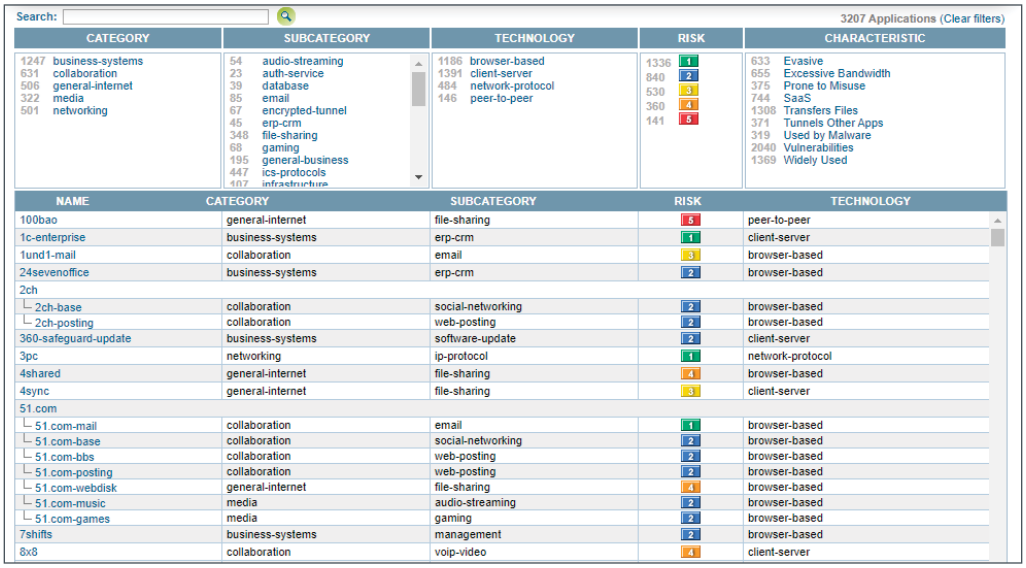



Figure 7 shows a list of App-IDs from Applipedia, an online resource from Palo Alto Networks where you can view applications and their characteristics, category, and risk level.

## Blocking Malware-Infected Websites, Files, and Exploits

To stop malware, you need to protect your users from websites hosting malware and stop known malware files. Palo Alto Networks URL Filtering service enables safe web access for all users. The cloud-based service uses a unique combination of static analysis and machine learning to identify, as well as automatically block, malicious sites and phishing pages.

URLs are classified in benign or malicious categories, which can easily be built into a next-generation firewall policy for total control of web traffic. Newly categorized malicious URLs are immediately blocked upon discovery, requiring no analyst intervention. URL Filtering is a subscriptions-based service that you enable through security profiles in your security policy.

To block malware by using the PA-Series firewalls, VM-Series firewalls and Prisma Access, you use multiple security profiles within your security policy rules. Security profiles enable you to inspect network traffic for threats such as vulnerability exploits, malware, C2 communication, and even unknown threats, and security profiles use various types of threat signatures in order to prevent these threats from compromising your network. Default profiles are available, or you can create your own custom profiles.

The security profiles specific to malware prevention include the following:

- **URL Filtering profiles**—URL Filtering complements App-ID by enabling you to configure the next-generation firewall to control access to websites and protect your organization from websites hosting malware and phishing pages. You can control how users access websites over HTTP and HTTPS. The default profile is configured to block known malware sites, phishing sites, and adult content sites. URL Filtering helps to block access to known malicious sites to prevent malware delivery and access to C2 hosts. With URL Filtering enabled, all web traffic is compared against the URL Filtering database, PAN-DB, which is updated frequently and contains a listing of millions of websites that have been categorized into approximately 60-80 categories.
- **File blocking profiles**—These profiles allow you to identify specific file types that you want to block or monitor. For most traffic, including traffic on your internal network, you want to block files that are known to carry threats or that have no real use case for upload or download. Based on the specific matching file types and applications, file blocking profiles block prohibited, malicious, and suspect files in order to protect the end users from downloading or uploading known malware executables.
- **Antivirus profiles**—The next-generation firewall uses a stream-based malware prevention engine to protect against downloads of common malware types, such as viruses, worms, trojans, and spyware. The profile scans for a wide variety of malware in executables. You can use application exceptions to avoid false positives. These profiles provide protection against malware concealed in common file types, such as Microsoft Office documents and PDFs.
- **Vulnerability protection profiles**—At the network and application layers, these profiles detect and block exploit attempts and evasive techniques, including port scans, buffer overflows, remote code execution, protocol fragmentation, and obfuscation. They stop attempts based on threats that have patterns related to exploits' attacks on system vulnerabilities. For example, the profiles protect against buffer overflows and illegal code execution.
- **WildFire analysis profiles**—These profiles forward unknown files or email links to WildFire for analysis. You can specify forwarding based on the application, file type, and traffic direction.

Many of the security profile features discussed are offered as subscriptions. These are leveraged in the security profiles and delivered as frequent content updates (such as every 24 hours for antivirus, within 30 minutes for application and threat updates, and 5 minutes for WildFire).

## Endpoint Protection from Malware

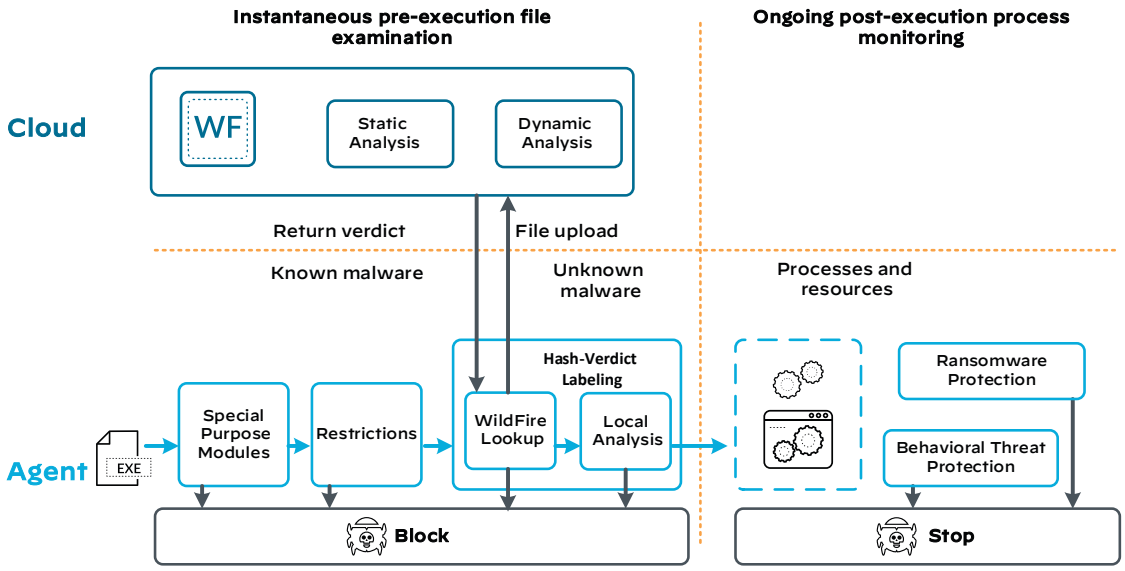
In combination with inline security, the endpoints need to be secured to prevent and provide visibility into malware attacks. This is important in case malware has been introduced through a USB flash drive, an infected device introduced behind the firewall, or devices not in the path of the firewall.

For blocking the techniques commonly used by malware, the Cortex XDR agent provides a wide variety of methods, known as Cortex XDR agent *malware protection modules*. This includes pre-execution file examination before a file or executable is run, which leverages local analysis and WildFire lookup, as well as ongoing post-execution process monitoring, which leverages behavioral threat protection and ransomware protection.

## Pre-Execution Modules to Stop Malware

Malware is typically delivered as an executable file that, once run, invokes other files and processes. On the endpoint, you need to be able to stop malicious files before they execute. *Pre-execution modules* are used to get a one-time pre-execution decision before a file is permitted to run. Their goal is to block potentially malicious files before they run. A sequential file-examination flow is used through protection modules for block or allow decisions. The protection modules can vary per platform operating system and include the Child Process Protection, Password Theft Protection, Reverse Shell Protection, and Execution Restrictions modules.

*Figure 8 Cortex XDR agent malware flow*




WildFire was designed to address weakness in traditional, signature-based anti-malware solutions. The Cortex XDR agent and the firewall send unknown files to WildFire for analysis. WildFire runs the files in a cloud-based sandbox environment with the following capabilities:

- **Static analysis**—Detects known exploits, malware, and any new variants
- **Dynamic analysis**—Detonation of files reveals zero-day exploits and malware
- **Heuristic engine**—Dynamically steers highly evasive suspicious files to bare metal
- **Bare-metal analysis**—Detonates evasive threats in a real hardware environment, entirely removing an adversary's ability to deploy anti-VM analysis techniques

The WildFire malware analysis service uses data and threat intelligence from the industry's largest global community and applies advanced analysis to automatically identify unknown threats and provide verdicts on the file. A verdict identifies samples as malicious, unwanted (grayware is considered obtrusive but not malicious), phishing, or benign.

### Post-Execution Modules to Stop Malware

On the endpoint, you need to be able to continuously monitor all applications and processes because they can start off as legitimate applications but then initiate malicious procedures. *Post-execution modules* continuously monitor currently running processes to detect abnormal behavior or malicious patterns in resource use. The protection modules can vary per platform operating system and include ransomware protection and behavioral threat protection.

The Behavioral Threat Protection (BTP) module is a post-execution protection module designed to stop script-based attacks. Script-based attacks are evasive, often difficult to detect, and they often use legitimate applications. The BTP module has the intelligence to detect patterns that consist of intentional events and malicious script operations, events that are outside of normal behavior. Detection is a continuous, ongoing surveillance activity, that determines the processes are calling on resources for a specific function or operation. A threat would typically consist of a few correlated activities that are outside of normal behavior. As part of the Cortex XDR agent, rules are based on behavioral analysis and associated with one of the following outcomes—Block, Report, or Silent. The Palo Alto Networks research team defines rules and updates them regularly through content updates.

The Cortex XDR agent also provides the ability to run file scans periodically or on-demand. Periodic scanning is supported only on the Windows OS. On-demand scanning is supported on Windows and Android platforms. On Windows, the Cortex XDR agent can scan drives (including removable drives) to check for *dormant malware*, malware that is not actively attempting to run.

The Cortex XDR agent quarantines a malicious file by moving it from its original location to a local quarantine folder and by changing its file format to a non-executable. The Cortex XDR agent supports two different ways to quarantine malware: automatically, by the Cortex XDR agent on Windows, and manually, by administrators on Windows and macOS.

## PREVENTING CREDENTIAL ABUSE

*Credential theft* is the first step of a credential-based attack, which involves unlawfully obtaining user account privileges through theft. Today, it's common for credential theft to start via phishing and malware attacks. Malware attacks are one of the most well-known credential theft techniques—for example, stealing credentials through techniques like keystroke monitoring. Other credential theft techniques include leveraging weak and default credentials for common applications where organizations have not changed the default user names and passwords, as well as brute-force attacks using bots to test millions of stolen username and password combinations on a targeted website or application. For more information, see MITRE Tactic TA0006, "Credential Access" (defined as "the adversary is trying to steal account names and passwords"). Multiple techniques are mentioned regarding how adversaries can either exploit software or install malware to obtain credentials.

Once obtained through credential theft, an adversary can sell compromised credentials or use them to access an organization's network. *Credential abuse* is the second step of a credential-based attack and consists of using the compromised credentials, typically with the malicious intent to steal data or cause damage within the organization. Compromised credentials can enable an adversary to bypass all security measures, operate undetected, and move laterally within the network in order to achieve their goals.

## Inline Protection to Prevent Credential-Based Attacks

Most adversaries take the easiest and fastest path to meet their objectives. Adversaries commonly use phishing for credential theft, because it is fairly inexpensive, easily implemented, and extremely efficient. Unlike malware and exploits, which rely on weaknesses in security defenses, phishing attempts to deceive users and is dependent on human interaction for its success.

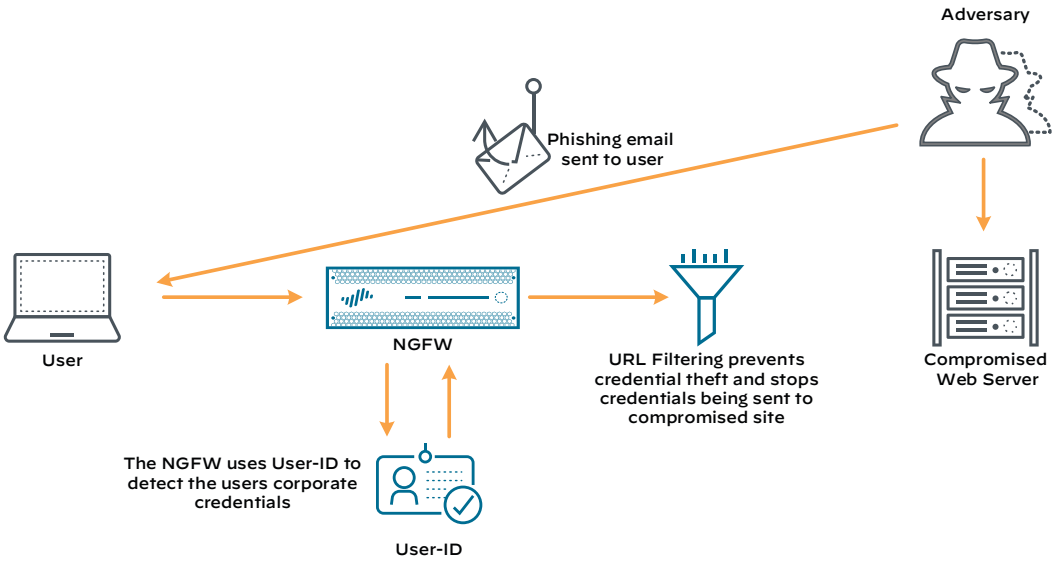
Adversaries can steal credentials more easily than they can locate and use a software vulnerability in an external-facing network. To avoid credential abuse, you first need to prevent the theft of any security credentials, and if they are stolen, you then need to prevent the invalid use of the stolen credentials. For preventing credential theft, the PA-Series firewalls, VM-Series firewalls, and Prisma Access leverage features such as User-ID and URL Filtering. To prevent credential abuse, the firewalls utilize authentication polices with multi-factor authentication (MFA).

### Credential Phishing Protection

To prevent credential theft and abuse, it's important to have visibility and control of the user credentials being used, especially when resources outside of your environment are using those credentials. User-ID provides visibility, policy control, and logging for your user's network activity and application usage. To avoid credential abuse, you can leverage User-ID to detect that the usernames being used on the network have valid credentials, as well as have visibility into which usernames are coming from a particular source IP address. You can configure the firewall to block or alert based on a network credential filter. You can configure User-ID and URL Filtering in order to prevent the use of corporate credentials on external websites or on websites that have been blocked. This prevents credential theft by restricting the user from sending their credentials externally to a malicious, blocked website.

With URL Filtering on the Palo Alto Networks next-generation firewalls and Prisma Access, you can prevent credential phishing by enabling the firewall to detect corporate credential submissions to external sites, and then you can control those submissions based on URL category. You can block users from submitting credentials to malicious and untrusted sites, warn users against entering corporate credentials on unknown sites or reusing corporate credentials on non-corporate sites, and explicitly allow users to submit credentials to corporate and sanctioned sites.

*Figure 9 Blocking credential leakage*




When URL Filtering is enabled, all web traffic is compared against the URL Filtering database. This database contains a listing of millions of websites that have been categorized into approximately 60-80 categories. You can use these categories as a match criterion in policies (Captive Portal, Decryption, Security, and QoS), or you can attach the categories as URL Filtering profiles in the security policy in order to safely enable web access and control the traffic traveling through your network. You can also create custom URL categories.

Figure 9 shows an example of a phishing attack where an adversary uses a phishing email to lure a user to a compromised web site. In this figure, the next-generation firewall leverages User-ID to detect the user's use of their corporate credentials, and based on the URL category, the use of URL Filtering on the firewall blocks the credential theft. On the next-generation firewall, URL Filtering is applied to a security policy rule with a URL Filtering profile.

## Preventing the Use of Stolen Credentials

If an adversary has stolen a user's credentials, to avoid credential abuse, you need to have another authentication layer in place to validate that the credentials being used are from a legitimate user. If a user's credentials have been stolen, MFA can prevent the use of the stolen credentials. Palo Alto Networks PA-Series firewalls, VM-Series firewalls, and Prisma Access support an MFA solution by integrating with multiple vendors through APIs. This does not need to be enabled on all of your firewalls, just in front of the resources you are protecting. MFA works well with the GlobalProtect agent on the client.

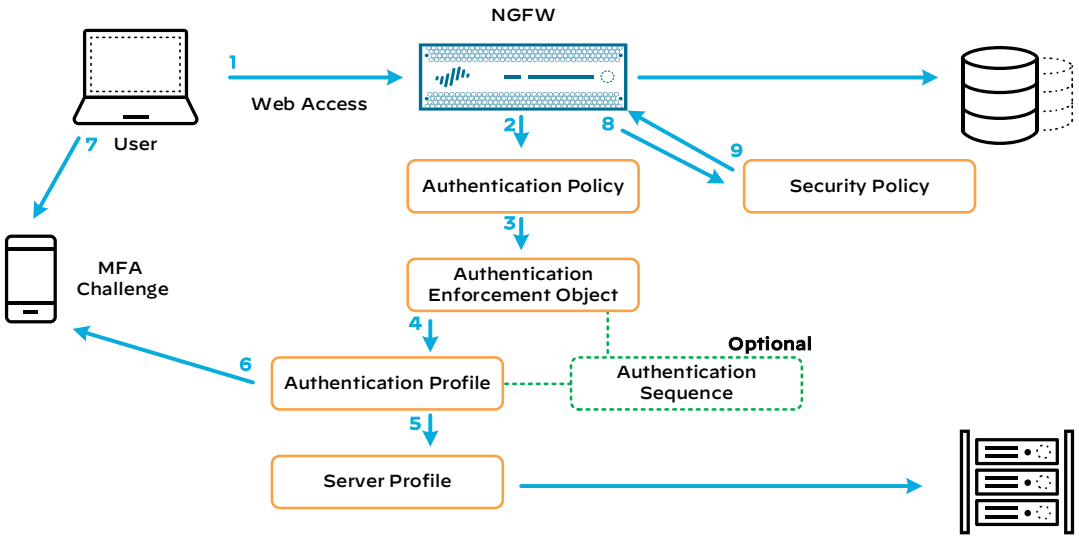
The Palo Alto Networks PA-Series firewalls, VM-Series firewalls, and Prisma Access support SMS, app push notifications, voice response, and a PIN code (app or key fob) as the secondary form of authentication. In order to use MFA on the next-generation firewall, you need to configure a captive portal with a web form for the first authentication challenge, or *factor*. When a user completes the first authentication factor, the firewall records a timestamp that it uses to evaluate timeouts, based on the authentication policy rules.

An authentication policy enables an administrator to selectively issue MFA challenges based on the sensitivity of the information stored on the network resource being accessed. For each network resource, you can configure the number and strength of the authentication factors.

When a user authenticates, the firewall records a timestamp for the first authentication factor and a timestamp for any additional authentication factors. When the user subsequently requests services and applications that match an authentication rule, the firewall evaluates the timeout specified in the rule relative to each timestamp.

In order for MFA to work, the firewall employs a series of steps and components to authenticate and authorize user access to web-based resources through the firewall, as shown in Figure 10.

*Figure 10 Firewall user authentication*



The process in Figure 10 consists of the following steps:

1. The user attempts to access a web resource in the organization's data center.
2. The firewall checks whether there is an authentication policy that requires the user to authenticate.
3. The authentication policy action field is configured with an authentication-enforcement object, which specifies the authentication captive portal information to use. The authentication-enforcement object also specifies an authentication profile.
4. The authentication profile defines which verification service to use, and the verification service defines which method to use for authenticating the user's credentials. The captive portal captures the user's first authentication factor and passes it to the authentication service.
5. The authentication profile specifies a server profile that defines connection settings to the authentication service. This example uses an LDAP server.
6. The user then receives an MFA challenge.
7. The user responds to the MFA challenge.
8. After the user has passed the MFA checks, the firewall checks the security policy to verify that the user is authorized to access the web-based resources.
9. The firewall then permits the traffic and forwards the traffic to the web resource.

User accounts can be defined in several account-authentication services. If this is the case in your environment, you can use an optional authentication sequence that specifies a set of authentication profiles for the firewall.

## Preventing Credential Abuse on the Endpoint

Sophisticated adversaries leverage exploit kits or malware to obtain users' credentials and gain access to data. Keyloggers and stealers are common malware tool types used to capture user credentials. *Keyloggers* covertly monitor and record users' keystrokes and can often be difficult to detect. *Stealers* are programs that wait for users to log into an OS or application and steal their credentials, log keystrokes, or run specific programs that dump password information stored in the OS and browsers.

Keyloggers and stealers are typically installed as malware or leverage known exploits in the OS or application in order to remain hidden while capturing your credential data. The two main components within the Cortex XDR agent are the Exploit Prevention and Malware Prevention modules, which look for activity related to exploit techniques and malware executables.

The malware example covered the Cortex XDR agent's Malware Prevention module. The next section covers the Exploit Prevention module because it's relevant to the credential abuse example, along with the Malware Protection module.

## Exploit Prevention

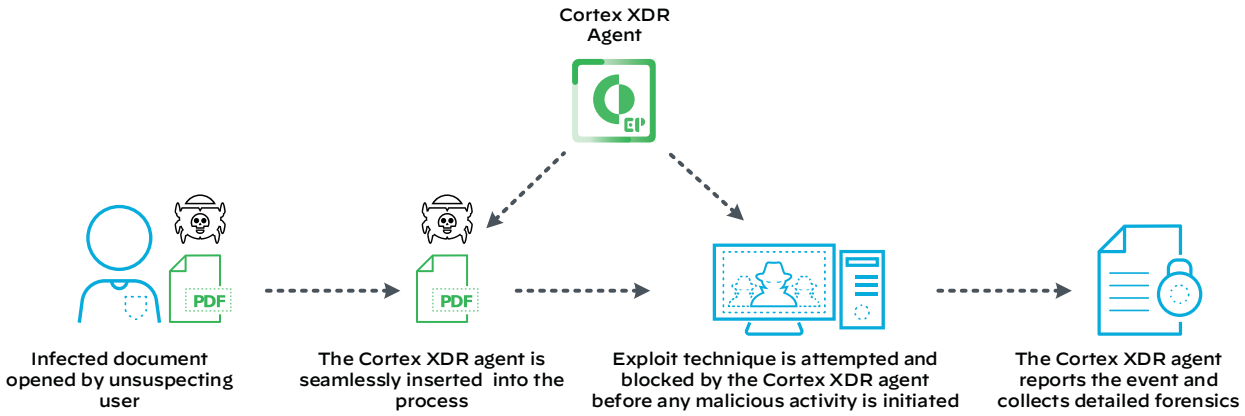
An *exploit* is a vulnerability in code that can be leveraged to cause an attack. Some attacks can start with an exploit before installing malware. An exploit typically uses several exploitation techniques. Palo Alto Networks focuses on the techniques rather than the exploits, because there are far fewer techniques than there are exploits. New exploits could be in the thousands per year, whereas there are zero to one new techniques per year.

An *exploit attack* is a malformed data stream that is embedded with a malicious payload, which tricks the legitimate application or kernel into running the adversary's code. An exploit delivers a small payload compared to malware, which is typically a large payload.

The goal of using an exploit is to cause an application to do something it is not supposed to do, such as run a piece of code supplied by the adversary. An exploit could be contained in a simple file, such as a PDF. After the exploit is set in motion, it can evade a local antivirus agent that might not have the capability within the agent to prevent exploits.

The Cortex XDR agent has Application Exploit Prevention modules to prevent unknown exploits that use known techniques. Each Exploit Prevention module targets and blocks a particular exploit type. Prevention of a technique in the chain blocks the entire attack. An adversary using an exploit to then install malware, such as a keylogger, is stopped before the exploit is successful. If the exploit manages to succeed, then the Malware Prevention module blocks the malware before it can be installed.

*Figure 11 Cortex XDR agent exploit prevention*




By default, the Cortex XDR agent stops known exploits and malware on the endpoint. The Cortex XDR app provides default security profiles that you can use out of the box in order to immediately begin protecting your endpoints from threats. You also have the capability to create your own custom Exploit and Malware Security profiles within the Cortex XDR app. You can create policy profiles for Windows, macOS, Linux, and Android. Windows, macOS, and Linux support both Exploit and Malware profiles, and Android platforms support just Malware profiles.

# PREVENTING LATERAL MOVEMENT

As the perimeter has expanded, attacks can come from any direction, from malicious insiders or external threats. Protecting your environment is essential. As part of an attack campaign, lateral movement is a technique adversaries use in order to cause damage or obtain and extract data. Typically, adversaries compromise a vulnerable host in a network and then use this as a starting point to move laterally through the network to explore the environment until they reach their desired target.

Lateral movement can avoid detection because technically it's not seen as malicious as it uses general tools and applications that can be common to the environment. Lateral movement is persistent, can take days to months, and often creates backdoors in the event one compromised system is discovered and locked down.

Often on the compromised host, tools are deployed for scanning the local network for open ports or known vulnerabilities on platforms. The adversaries can move from system to system, exploring the environment, scanning and listening to traffic before gaining access to the data they want to extract. For more information, see MITRE Tactic TA008, "Lateral Movement" (described as "the adversary is trying to move through your environment").

## Inline Protection from Lateral Movement

To reduce the attack surface, you should have a security strategy that focuses on identifying, controlling, and safely enabling applications while inspecting all content for threats. In addition, network segmentation is a best practice you can use to partition the enterprise network into manageable segments, which reduces the scope of compliance, limits data exfiltration, and reduces the attack surface.

### Segmentation to Reduce the Attack Surface

Palo Alto Networks provides segmentation capabilities with User-ID, security zones, and flexible deployment modes with Layer 1 (also known as *virtual wire*), Layer 2, and Layer 3, allowing you to easily insert the PA-Series and VM-Series firewalls into any existing architectural design. Tightly integrated technologies, such as App-ID and User-ID, are used to identify who is using which applications across different network segments or security zones.

User-ID and App-ID are important features for an effective security infrastructure. With User-ID you can gain visibility into who is using an application, rather than just their IP address, and with App-ID, you can see what application is in use. To prevent lateral movement, you can use security policies to specify which applications are allowed and by which users, with security rules to restrict nonstandard behavior such as port scans or windows remote desktop logins.

App-ID classifies all network traffic across all ports, enabling administrators to create logical application-based security policies. App-ID uses multiple identification techniques to identify applications passing through the network. This helps with lateral movement when adversaries try to mask applications' port numbers in order to evade detection. App-ID combined with User-ID gives you better visibility into what is traversing your network. You can treat unknown traffic as a separate category and handle it based on your organization's risk profile.

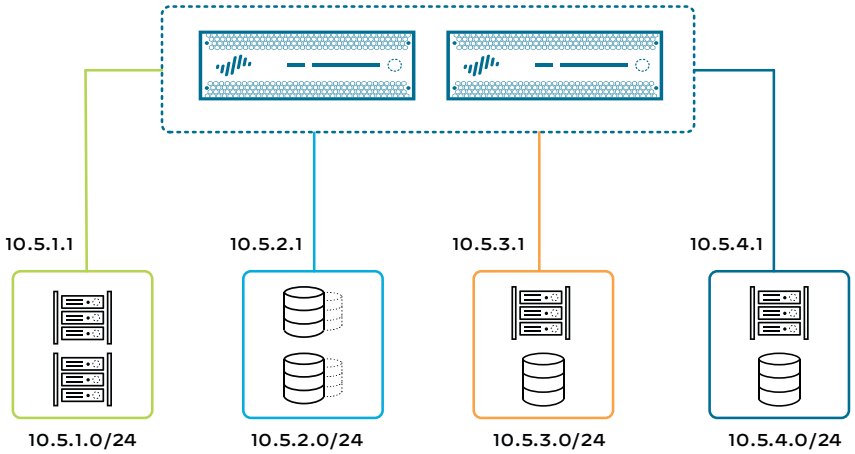
User-ID provides visibility into who is using applications on the network, which allows you to detect a threat, file transfer, or abnormal user behavior. By leveraging User-ID, you define which applications are permitted either inbound or outbound, and you restrict tools with a least privilege access policy (such as allow FTP) to only those users who need it between specific sources and destinations. User-ID gives you forensic capability for user-based analysis with reports.

User-ID communicates with an organization's existing directory, such as Microsoft Active Directory or LDAP, to gain information about users or user groups. With knowledge of the users in an organization, administrators can create dynamic contextual policies to secure access between zones for specific authorized individuals or functional departments. For compliance purposes, implementing User-ID means auditors see the identities of employees, contractors, and guests who have access to data—not just a bunch of IP addresses.

Segmentation of networks into separate and smaller zones helps reduce the attack surface. A larger, unsegmented network is much harder to protect and is more difficult to monitor and control. Because traffic and applications have access to the entire network, after an adversary enters a network, the adversary can move laterally to access critical data. Network segmentation is a critical part of an effective security strategy to reduce the risks and impact of attacks to the network. Segmenting the network prevents lateral movement between zones, which limits an adversary's ability to move through the network.

A *security zone* is a group of one or more physical or virtual firewall interfaces and the network segments connected to the zone's interfaces. You control and specify protection for each zone individually. You need to segment based on risk to the organization and sensitivity of the applications and data.

*Figure 12 Network segmentation*




Deploying lots of smaller zones provides better visibility into traffic and prevents lateral movement of malware more effectively because all traffic must be inspected before crossing zones. Therefore, the more granular you are with your zones, the more you reduce your attack surface and have greater control over access to sensitive applications and data, giving you more protection against lateral movement.

### Blocking Spyware

If hosts are infected on your network, they attempt to reach out to external C2 servers. You need to prevent this communication from happening. You can use anti-spyware profiles to block spyware on compromised hosts reaching out to external C2 servers. You can apply anti-spyware profiles to inspect all zone traffic, and you can apply various levels of protection between zones.

Compromised hosts try to access malicious sites. You can use DNS sinkholing to prevent access to these sites. You can enable DNS sinkholing within the anti-spyware profile in order to enable the firewall to respond to DNS queries for known malicious domains. The firewall makes the DNS query resolve to an IP address you specify, which helps to identify the infected hosts attempting to reach the DNS sinkhole address.

The Threat Prevention subscription bundles the antivirus, anti-spyware, and vulnerability protection features into one license. These are leveraged in the security profiles, which were covered in the malware example, but the anti-spyware security profiles are important for preventing lateral movement.

### Zone Defense to Stop Malicious Lateral Movement

To prevent lateral movement and DoS attacks, your defense requires a layered approach that includes a security policy, zone protection profiles, and DoS protection profiles. This guide has already discussed security profiles. Below we cover Zone and DoS protection profiles:

- **Zone protection profiles**—Increase network security and prevent lateral movement activities. Within a zone protection profile, you have the ability to configure reconnaissance protection. Zone protection profiles have configurable settings for flood protection, reconnaissance protection, packet-based attack protection, and protocol protection. Enabling lateral movement reconnaissance protection allows you to configure settings for TCP and UDP port scans as well as host sweeps. The settings include response actions based on configured time intervals.
- **DoS protection profiles**—DoS attacks are used to cause disruption to a targeted server or services. DoS protection profiles allow you to control the number of sessions between interfaces and zones to avoid an internal session attack from a comprised host within the network. DoS protection profiles protect the network and resources from a *flood attack*, where an adversary attempts to overwhelm network resources by sending too many packets or by using many hosts to establish multiple sessions. The profile supports settings for SYN, UDP, and ICMP floods.

## Endpoint Protection from Lateral Movement

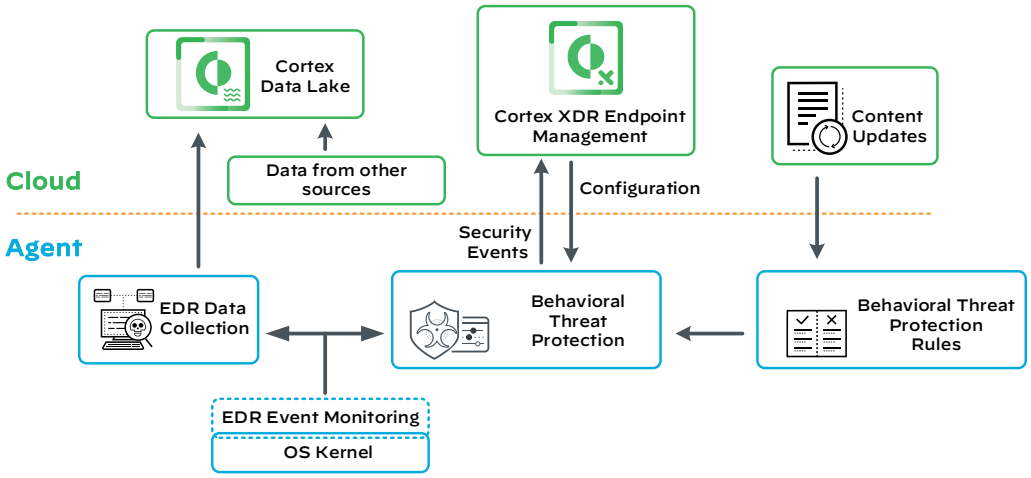
Lateral movement can often be hard to detect because the applications and tools being run might not look malicious. This is where behavioral threat protection built into the Cortex XDR agent as a post-execution module comes into play. It continuously monitors system events, looking for events that are part of the same chain of events, known as a *causality chain*. Each causality group has a *Causality Group Owner*, which is the process that initiated the chain of events that is part of the causality group.

When the Behavioral Threat Protection module triggers a prevention, it terminates all processes and threads that are part of the causality group. With malware, the BTP module is concerned with malicious script executions; with lateral movement, it is looking at misuse of legitimate operations for attacks and has the intelligence to detect attack patterns that consist of intentionally re-ordered events.

*BTP rules* are a predefined set of operations monitoring resources leveraged by adversaries. The Palo Alto Networks research team creates the rules and updates them by content updates. The rules can either block or report (which allows the activity but reports it to Cortex XDR).

Behavioral threat protection consumes *endpoint detection and response* (EDR) data. EDR is a category of tools or platforms that protect endpoints from threats.

*Figure 13 Behavioral threat protection consumes EDR data*




EDR data, combined with the Cortex XDR agent multi-method endpoint-protection technology, enables your security teams to automatically protect, detect, and respond to known, unknown, and sophisticated attacks. The "Detection and Investigation" section discusses Cortex XDR and how it uses machine learning and AI techniques to analyze data collected from multiple sources, including the endpoints, network, and cloud.

## PREVENTING DATA EXFILTRATION

Data exfiltration is normally one of the end objectives of a cyberattack. Data exfiltration can be classified as a security breach and can be achieved using various techniques. It's most commonly performed over the Internet or a network. The primary objective is gaining access to a network or device in order to copy specific data.

Data exfiltration can also occur through your employees, either by accident (such as placing data in an unprotected location) or deliberately with malicious intent, where employees are looking to harm the organization for their own personal gain or on behalf of someone else.

Because data exfiltration involves moving data within and outside the network and it often resembles typical network traffic, data exfiltration can be difficult to detect. Delay or failure to detect the exfiltration can result in substantial amounts of compromised or lost data, and the damages can be immense.

With stricter compliance regulations around data privacy, like GDPR and the California Consumer Privacy Act, the stakes for reporting data exfiltration events have also gotten much higher.

For more information, see MITRE Tactic TA0010, "Exfiltration" (described as "the adversary is trying to steal data"), which describes multiple techniques adversaries use, including techniques for stealing compressed and encrypted data.

## Inline Protection to Prevent Data Exfiltration

To prevent data exfiltration inline, you need to be able to control access to external locations, block files that contain sensitive information from leaving the organization, and prevent data transfers through file transfer protocols by unauthorized users and applications. Palo Alto Networks PA-Series firewalls, VM-Series firewalls, and Prisma Access provide several capabilities to block data exfiltration during an attack, including DNS security, data filtering, file blocking, URL filtering, and the use of App-ID and User-ID to prevent file transfer applications.

### Disrupting Attacks That Use DNS

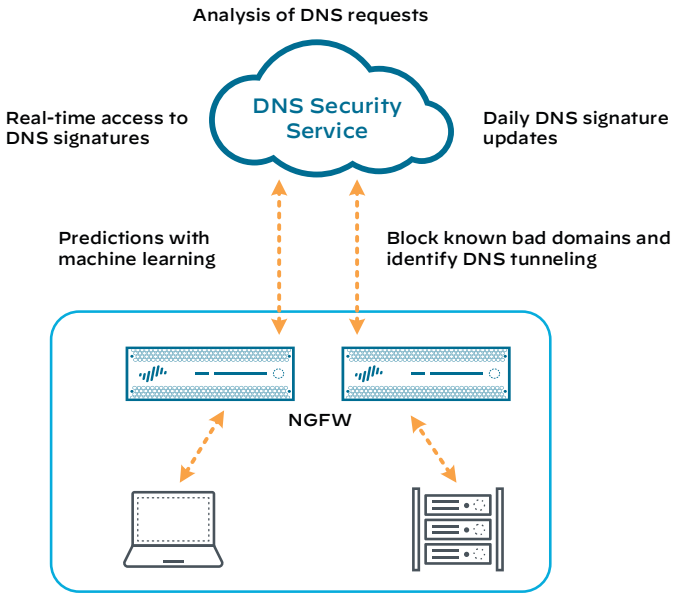
DNS is a massive and frequently overlooked attack surface that is present in every organization. Adversaries take advantage of DNS to abuse it at multiple points in order to steal data or establish connections with C2 servers. It's difficult for security teams to keep up with the high volume of malicious domains, let alone protect against advanced tactics such as DNS tunneling for stealthy data theft. Adversaries can use DNS tunneling to encode data of non-DNS based programs with DNS queries and DNS responses, which can often be used as a channel to exfiltrate data. DNS tunneling detection can detect a tunneling-based attack and block it with security polices, avoiding data theft.

The DNS Security service from Palo Alto Networks is a subscription-based service (on the next-generation firewall) that is designed to protect and defend your network from advanced threats that are using DNS. The DNS Security service leverages machine learning and predictive analytics to provide real-time DNS request analysis. The analysis enables production and distribution of DNS signatures that are specifically designed to defend against malware that uses DNS for command-and-control and data exfiltration.

The DNS Security service allows the next-generation firewall to sinkhole internal DNS requests, which basically allows the next-generation firewall to forge a response to a DNS query for a known malicious domain/URL and causes the malicious domain name to resolve to a definable, fake IP address that is given to the client. If the client attempts to access the fake IP address, a security rule is used to block traffic to this IP address and log the information.

The DNS Security service also uses techniques such as a domain generation algorithm (DGA) and DNS tunneling detection. Because malicious domains are frequently autogenerated by machines, a DGA analysis can determine whether a domain was likely generated by a person or machine. By reverse-engineering and analyzing other frequently used techniques, the DNS Security service can identify and block previously unknown DGA-based threats in real time.

*Figure 14 DNS security*




## Data Loss Prevention

As an organization, it is important to protect and secure company data across all repository locations. As part of your strategy, you need to have controls in place that prevent sensitive data, such as social security numbers, credit cards, and personally identifiable information, from leaving your network.

To prevent sensitive, confidential, and proprietary information from leaving your organization, you can use the data filtering profiles in security policy rules. With data filtering profiles, you can define data patterns for which you want to filter, such as credit card information, social security numbers, HIPAA data, and many more.

Within the profile, you can define whether you want to block, alter, or log the activity. You can use a default data filtering profile or customize your own. With a data filtering profile, you can avoid data theft by blocking the specific types of data (based on data patterns) that are not allowed to leave your environment. You can also block specific file types with file blocking profiles.

With file blocking profiles, you have the ability to block files, based on file type, from entering, traversing, or leaving your environment. You want to block file types that are known to carry threats or that have no real use case for upload or download. These include batch files, DLLs, Java class files, help files, Windows shortcuts, and torrent files. File blocking allows you to block potentially malicious files from entering or traversing your network as well as protect against data exfiltration, where data might be concealed in zipped files. File blocking profiles are applied to security policy rules just like the previously discussed security profiles, such as data filtering profiles.

The file blocking profile gives you control over the flow of different files by inspecting the payload in order to identify the true file type, as opposed to looking only at the file extension, to determine if your policy allows a file transfer. You can implement file blocking by type, user group, or on a per-application basis. For example, you could approve a specific webmail application (such as Gmail), allow attachments, and block the transfer of portable executable files.

You can leverage URL Filtering to block potentially high-risk file downloads from specific URL categories by creating a security policy with a file blocking profile attached.

### Preventing Malicious File Transfers

Having visibility over which applications are allowed on your network gives you more control over the applications used by adversaries to exfiltrate data by using file transfer applications that might not be used or required by the users. With App-ID and User-ID, you have visibility into and granular control over application use on the network, on a per-user basis. App-ID combined with User-ID gives you the ability to define which users are permitted to use a given application. Data exfiltration often uses file-copying programs that use FTP, SMTP, HTTP, DNS, and SMB, to name a few.

With security policy rules leveraging App-ID and User-ID, you can block the use of file-transfer applications, such as FTP and instant messaging, for all users and then allow only specific users who need access to these applications.

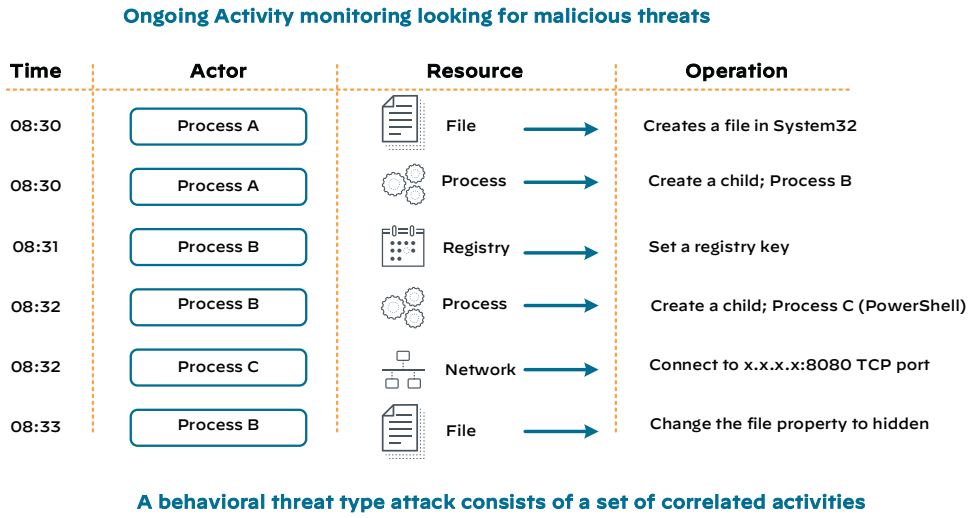
## Endpoint Protection to Prevent Data Exfiltration

Adversaries use exploits and malware to compromise the endpoint in order to deploy software and programs they can use to exfiltrate data out of your network. You ideally want to stop the security breach before an adversary compromises the endpoint, hence before exploitation or installation of malware.

As mentioned in previous sections, the Cortex XDR agent uses a multi-method prevention approach to stop known and unknown threats such as exploits and malware. Data exfiltration is typically performed over a command-and-control channel, which is often disguised by using an application that does not typically act or behave in this manner.

The Cortex XDR agent detects abnormal behavior through the BTP module, which continuously monitors endpoint activities in order to identify causality chains. During the stages of the cyberattack, the Cortex XDR agent is evaluating the behavior and determining if any of the events are part of the causality chain. The BTP module leverages machine learning to consciously monitor activities on resources, detecting attack patterns that consist of intentionally re-ordered events. After a threat is detected, it is blocked.

*Figure 15 Activity monitoring by the BTP module*




To determine if malicious activity is present, the BTP module monitors endpoint activities, such as registry access, process executions, file access, network operations, DLL retrievals, macro executions, and many other activities. The Cortex XDR agent obtains and consumes the data from the EDR Event Monitoring process and not the OS.

## PREVENTION SUMMARY

Palo Alto Networks offers a comprehensive preventative approach that leverages inline security capabilities with PA-Series firewalls, VM-Series firewalls, and Prisma Access and leverages endpoint-protection capabilities with Cortex XDR. These components from Palo Alto Networks give you increased visibility and attack protection from malware, exploits, credential abuse, data exfiltration, and many other attack types. Reducing the attack surface and having a strong preventive approach benefits the SOC by reducing the quantity of alerts they need to process on a daily basis.

The next section describes components for rapidly detecting and investigating threats that can be harder to detect and prevent without a platform in place such as Cortex XDR, which enables the SOC to obtain alert details that have more visibility and context from multiple data sources, such as the network, endpoints, and cloud. Cortex XDR provides the SOC with detection, investigation, and response capabilities.

# Detection and Investigation

After you have prevented all the threats you can, you then need to rapidly detect and investigate any threats that were not prevented. To find important security events without generating too many low-value alerts that require lots of analysts' time, attention, and manual remediation, the SOC needs an efficient way of handling incidents.

The SOC is challenged with manual tasks and having to switch through multiple screens to understand and obtain alert data from multiple, disparate sources, making it difficult to understand the full effect of an attack at any given point in time. This is time-consuming for the analysts, and dealing with multiple false positives means that real threats can take longer to discover and resolve. Being able to determine the root cause of an alert is often difficult without a clear visual representation of what has occurred and how incidents are related as part of a single attack.

For the SOC to be effective, a solution must provide more context of what is happening in their environment and get to the root cause of an attack much faster and more accurately than they do today. Having access to rich data, obtaining better visibility, and gaining insight and context have the benefit of reducing the time spent investigating incidents and reducing the skills and experience required in the SOC.

Analysts shouldn't need to manually analyze or correlate the growing amount of collected data in order to identify threats. They must be able to quickly confirm and uncover stealthy attacks and obtain rich investigative data with actionable alerts. They can achieve this with AI and ML, which can learn the unique characteristics of your organization's behavior, create a baseline for normal behavior, and then detect abnormal behavior, which is important for finding zero-day attacks.

AI and ML are important components that can facilitate detection and investigation capabilities by understanding and examining security events in an organization, with high fidelity, by stitching together data from multiple sources.

Based on the challenges discussed and the capabilities needed, a new approach is required to solve current security operations challenges. The approach is one that eases every stage of security operations, from threat detection and threat hunting to triage, investigation, and response. Palo Alto Networks understands these challenges facing the SOC and addresses them with Cortex.

## INTRODUCTION TO CORTEX

For detection and investigation, customers can leverage Cortex. Cortex is Palo Alto Networks' AI-based continuous-security platform, which continuously evolves to help the SOC stop the most sophisticated threats. Among the components that currently comprise Cortex:

- **Cortex Data Lake**—Cortex Data Lake is a cloud-based logging infrastructure that normalizes and stitches together your enterprise's data from the network, endpoint, and cloud and allows you to centralize the collection and storage of logs from your log data sources. Cortex Data Lake is built to simplify security operations by collecting, transforming, and integrating your enterprise security data. The benefit of using Cortex Data Lake goes well beyond scale and convenience when tied into the Palo Alto Networks Cortex platform. Cortex Data Lake acts as the exclusive, customer-specific data store used by the Cortex platform and the associated apps.
- **Cortex XDR**—Cortex XDR is a cloud-based detection and response product that natively integrates network, endpoint, and cloud data to stop attacks. Cortex XDR uses continuous ML-based detection and automated root-cause analysis, and it has been designed to help organizations secure their digital assets and users while simplifying operations.
- **Cortex XDR agent**—The Cortex XDR agent is a component of Cortex XDR. Cortex XDR Prevent provides endpoint prevention only, and the Cortex XDR Pro offering includes analytics and machine learning for investigation and response for all alerts from network, cloud, and third-party logs.

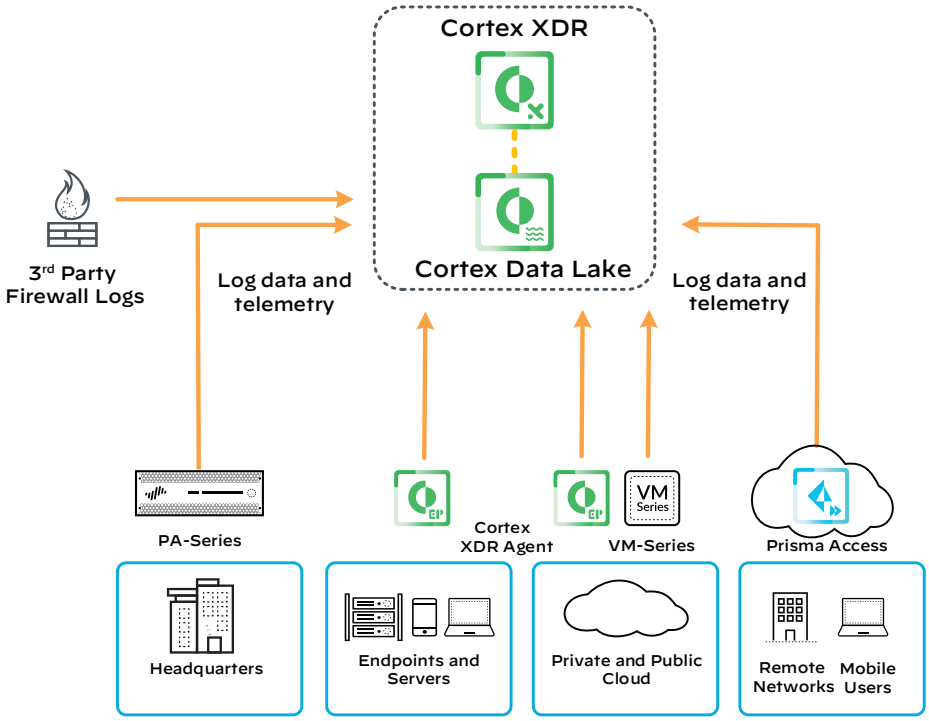
## CORTEX XDR

Palo Alto Networks manages Cortex XDR through a cloud-hosted app. The Cortex XDR app provides complete visibility of network traffic, user behavior, and endpoint activity. Cortex XDR simplifies threat investigations and helps you identify the root cause of alerts. Cortex XDR achieves this by using data from Cortex Data Lake, which stitches together multiple logs across your different log sensors to derive event causality and timelines. The Cortex XDR app provides detection, investigation, and response capabilities as well as an endpoint section to fully manage all your Cortex XDR agent endpoints.

Organizations have typically relied on endpoint detection and response (EDR) as a trigger for cybersecurity incidents. As attacks have become more sophisticated, the time it takes to detect, identify, and contain breaches has increased. Cortex XDR leverages endpoint, network, and cloud data, surpassing the traditional EDR approach of using only endpoint data to identify and respond to threats. Cortex XDR applies machine learning across all your endpoint, network, and cloud data. This approach enables you to rapidly detect and stop targeted attacks, detect insider abuse, and remediate compromised endpoints.

The main components of Cortex XDR are the analytics engine and the sensors. The analytics engine uses network and endpoint data to detect and report on threats based on alerts it receives from multiple sensors, such as firewalls and endpoints. The analytics engine does this by identifying good (normal) behavior on your network, so that it can detect bad (abnormal) behavior.

*Figure 16 Cortex XDR*




The analytics engine alerts on a subset of the attack tactics defined by the MITRE ATT&CK knowledge base of tactics, including execution, persistence, discovery, lateral movement, command and control, exfiltration, and impact. A baseline of normal behavior is created so that when abnormal behavior is detected, an alert can be raised. Each detector raises an alert when abnormal behavior is identified. Alerts are grouped into *incidents,* reducing the amounts of alerts to view.

The alerts within the incident consist of key artifacts that Cortex XDR extracts and uses as criteria for other alerts that have the same or similar artifacts so that they can be grouped together. Artifacts can include items such as software processes, signers, files, file hashes, domains, and IP addresses. The app analyzes the type of alert and determines whether there are any other alerts related as part of a causality chain. The key artifacts are used to determine which incident has the highest correlation with the alert, and the Cortex XDR app then groups the alert within that incident.

To detect advanced, targeted, and insider attacks, in addition to non-malicious risky behavior, Cortex XDR uses AI and ML to learn from the data in Cortex Data Lake in order to gain insight on activity across the environment. Cortex XDR establishes a baseline of normal behavior and then alerts on high-fidelity anomalies. For example, Cortex XDR uses ML models to detect whether or not an administrative connection to a server is expected behavior.

For Cortex XDR, Palo Alto Networks researchers create the *ML models*, which is the algorithmic representation of what a machine-learning system has learned from the training data, and the ML models detect anomalous activities. An ML model is essentially a more refined equation for the algorithm, one in which critical parameters used by the algorithms have been determined. In the example of a compromised network user, the parameters might include trusted or untrusted applications, ports being used, or the time of day of use.

Cortex XDR identifies a forensic timeline of aggregated events, creating a sequence of events called a *causality chain*. The causality chain is built from processes, events, insights, and alerts associated with the activity. When investigating why an alert occurred, you should review the entire causality chain, which consists of all the alerts and the order in which they occurred. Each causality chain has a *Causality Group Owner*, which is the process that caused or is responsible for the activities that initiated the alert (for example, opening an email).

To identify threats, in individual rules you create, you can define specific indicators on which you want Cortex XDR to raise alerts:

- **Indicators of compromise (IOCs)**—As determined by a computer forensics process, *IOCs* are known artifacts that are considered malicious or suspicious. IOCs are static and based on criteria such as SHA256 hashes, IP addresses and domains, file names, and paths. You can create IOC rules based on information that you gather from investigations within Cortex XDR or through various other threat-intelligence feeds that you gather.
- **Behavioral indicators of compromise (BIOCs)**—These enable you to alert and respond to behaviors (tactics, techniques, and procedures). Instead of hashes and other traditional indicators of compromise, BIOC rules detect the behavior of processes, registry, files, and network activity. Cortex XDR comes with multiple BIOCs already available.
- **Rule exceptions**—You can use rule exceptions to create a rule to take action on specific behaviors while excluding one or more indicators from the rule, such as a hash, username, or path name.

The Cortex XDR features for detection, investigation, and response—including rules, BIOCs, and IOCs—are covered in more detail in the individual threat examples.

## DETECTING AND INVESTIGATING MALWARE

Detecting malware and investigating malware-infected hosts is a common task for the SOC. New zero-day attacks might get through your existing defenses, requiring you to be able to detect and investigate these threats as quickly as possible. For these types of threats, analysts might have to look for artifacts that are indicators of a malware infection, rather than the malware itself, in order to locate the new malware.

Advanced malware can use *polymorphism*, where each iteration of the malware takes on a new characteristic, making it much harder to identify and detect. With these techniques, adversaries are often able to bypass detection.

The Cortex XDR agent, PA-Series firewalls, VM-Series firewalls, and Prisma Access protect the endpoints from known and unknown malware and exploits as part of a preventative approach. There can be cases where malware can be undetected or dormant, and the ability to detect, investigate, and respond to malware threats is paramount. The Cortex XDR app provides multiple features and investigative capabilities over and above the endpoint management capabilities discussed previously.

Cortex XDR consumes data from Cortex Data Lake and then correlates network, endpoint, and cloud data across your detection sensors. The act of correlating logs from different sources is called *log stitching*. The Cortex XDR app detects malware and grayware by using a combination of network activity, Cortex XDR agent endpoint data, WildFire analysis of suspicious files, and Pathfinder scans of your endpoints. Pathfinder is deployed as a virtual machine that you install on your network for the purpose of investigating your network hosts, servers, and workstations for malicious or risky software and other artifacts. If the firewalls detect malware, Cortex XDR correlates that activity with endpoint logs in order to assess the impact of the malware, identify the cause, and identify the origination point. Reported information includes details about the execution process that generated the malware and all the context, including the host and user who executed it.

The Cortex XDR app offers the SOC multiple functions, including viewing the top incidents, creating and viewing reports, investigating alerts, responding to threats, managing the Cortex XDR agents, and creating rules. Within the **Investigation > Incidents** section of the Cortex XDR app, you are able to view all incident types based on incident severity (Low, Medium, or High). Each incident has an incident description and illustrates which host the incident affected.

An attack event can affect several hosts or users and raise different types of alerts caused by a single event. You have the ability to visually view all alerts or filter on specific events or alerts of interest, such as malware. You can investigate an incident, assign the incident to an analyst, change the severity, or even view or assign an incident status (such as **Under Investigation**).

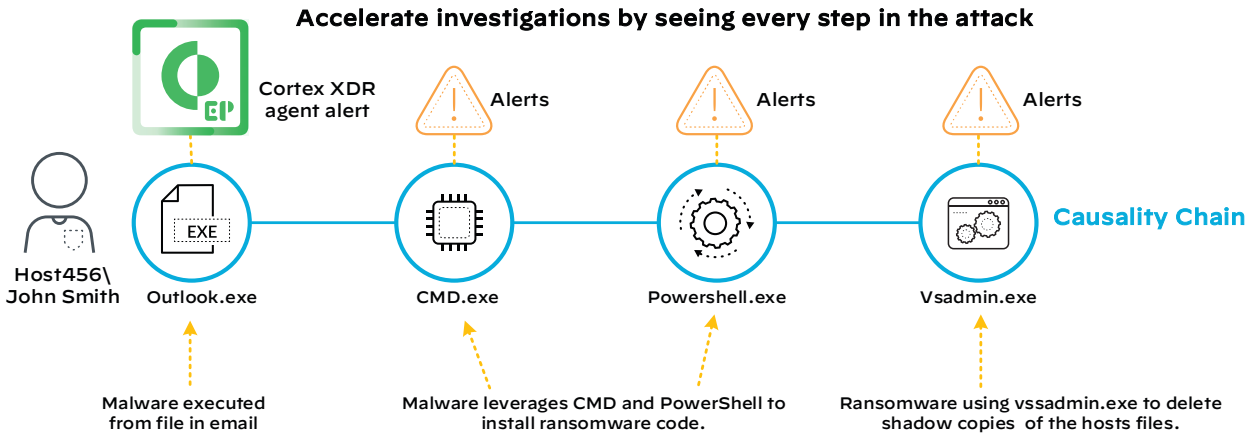
During your investigation, when viewing an incident such as malware, you can also perform additional management of alerts related to the incident, in the Cortex XDR app **Investigation > Incidents** section, which includes the following:

- **Analyze**—Allows you to understand the full context of the alert from all data stitched together to give you a causality view. The alert provides a lot of detailed information around the malware alert, including if it was prevented or just detected.
- **View the alert causality**—The causality view provides information on the alert that was generated and the entire process execution chain that led up to the alert being generated. You have the ability to see what happened before and after the malware file was executed.
- **View the alerts in a timeline**—A forensic timeline shows the sequence of events, alerts, and informational BIOCs that are involved in the attack over time.
- **Copy alerts**—You can copy the URL of the alert record or values of the alert to paste into an email to share with a team member or record in a document or report.
- **Build an alert exclusion policy from alerts in an Incident**—After review or investigation, you might find some alerts that you want to suppress from appearing in the future. You can create an exclusion policy to prevent certain alerts from triggering incidents.

For any incident you want to investigate, you can analyze the incident and obtain additional details on the malware, any affected hosts, when and how it occurred, if the malware successfully executed or if it was blocked, as well as the path location where it executed. If multiple alerts were related to this one malware incident, then they are listed, and you can view and analyze them individually or as one whole causality chain.

Analyzing the event gives you a lot of useful information such as the username that was used to run the file, as well as a graphical representation of how the malware file was executed and what applications or processes were then executed from the initial executable, which could have been opened in an email in Microsoft Outlook, for example.

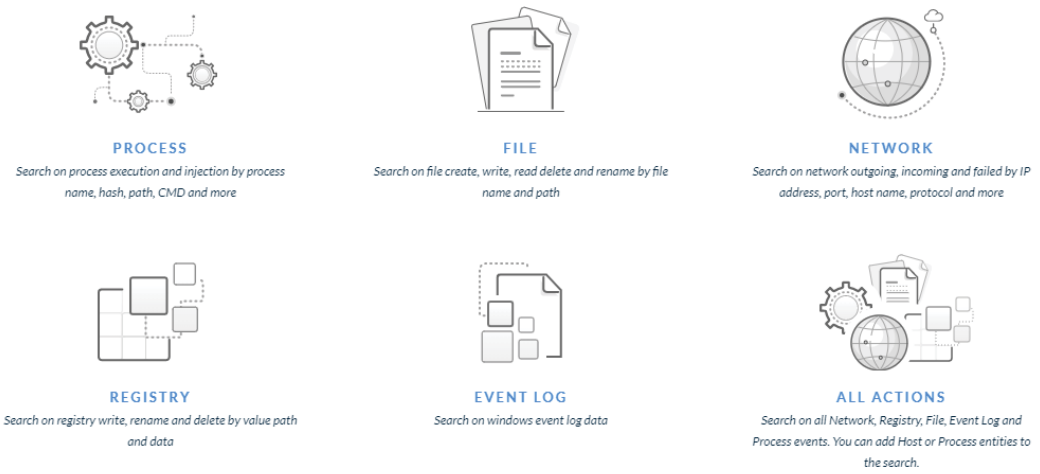
*Figure 17 Incident analysis*




The previous figure shows a basic example for malware-delivered ransomware on a host machine. The ability to see the chain of events that led to the attack visually helps with investigation, remediation, and response to similar type activities in the future. Within the GUI, you can view the child or parent processes for each node in the causality chain. You can also click on each event of the chain to understand the details of the alert, including what commands were run, the file path, and if the file was malicious.

When investigating an incident such as malware, you can use a feature within the Cortex XDR app called *Query Builder*. Query Builder is a powerful search tool that you can use to investigate any type of incident or indication of an attack very quickly, understand the root cause of an alert, determine what damage has occurred, and hunt for threats from all your data sources. Query Builder allows you to build complex queries for entities and entity attributes so that you can surface and identify connections between them. Query Builder searches for all the entities and attributes you specify from the raw data in the logs received by Cortex Data Lake.

*Figure 18 Query Builder*



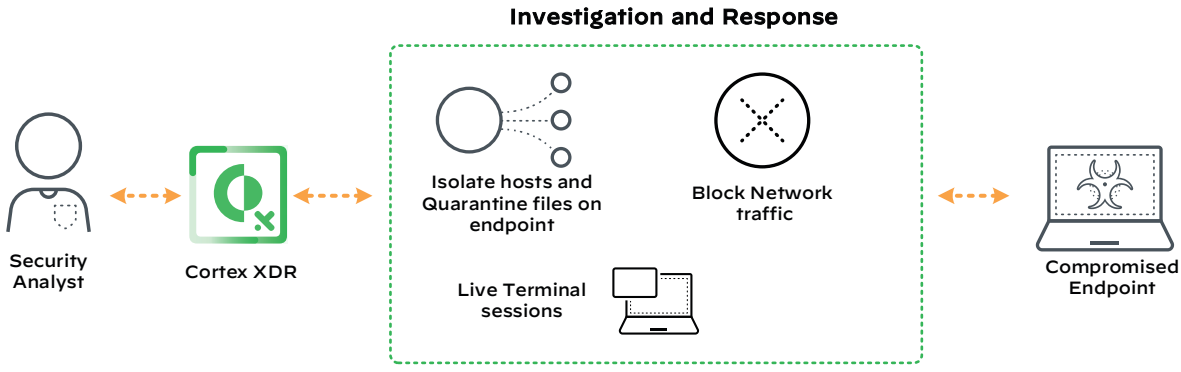

You can search for multiple object types, including processes, files, networks, registries, event logs, and all actions (runs a query across all the mentioned entities). If you suspect a host is infected with malware, you can run a search on all actions and select the host option, where you can search by host name, host IP, or host OS, for example. All queries you create are located in the Query Center so that you can rerun them or modify them for any future queries. You also can also create scheduled queries. Query Builder gives you the flexibility to run queries based on threat intelligence where a confirmed threat has known behaviors, characteristics, and file types it uses.

Within the Response section of the Cortex XDR app, there are multiple response features that you can use for malware incidents:

- **Action Center**—You can leverage the **Response > Action Center** section to quarantine hosts, run a malware scan, isolate an endpoint, retrieve support files, and even create whitelists and blacklists based on multiple hash values. The Action Center also lists and tracks the progress of all investigation, response, and maintenance actions you performed on your Cortex XDR endpoints.
- **External dynamic lists**—You can create *external dynamic lists* (EDLs) for the firewalls to block files, applications, or URLs related to malware. You can use a Cortex XDR EDL with a Palo Alto Networks firewall to provide an integrated response to malicious network activity. With a Cortex XDR EDL as the source of a firewall EDL, the firewall can control user access to IP addresses and domains that the app has found to be associated with an alert.
- **Live Terminal**—A powerful response capability is *Live Terminal*, which allows you to remotely connect to hosts and isolate an endpoint from everything else except access via Cortex XDR.

  As part of investigating and responding to security events related to endpoints, you can use Live Terminal to initiate a remote connection to an endpoint. The locally installed Cortex XDR agent handles the connection by using a remote procedure call. With Live Terminal, you can manage remote endpoints, investigate, and perform response actions, including the ability to explore and manage files in the file system, manage active processes, and run commands. This is ideal when you suspect a host has been compromised and further investigation is needed, including running malware scans, viewing quarantined files, and stopping and starting processes.

*Figure 19 Live Terminal*




With Live Terminal, you can contain, respond, and remediate. Security analysts are able to view alerts remotely, upload any tools they need, and run Python or PowerShell commands or scripts for in-depth forensic investigations. With a complete audit trail and without disrupting end users, security analysts can eliminate threats by terminating and deleting processes in a live environment on any host.

Live Terminal maintains an audit of all interactions between the user and the host, logging connections, file and tool uploads, and any items marked to be documented against the user. Also, any data downloaded from the host is hashed, creating a defensible process that meets basic forensic requirements.

## DETECTING AND INVESTIGATING CREDENTIAL ABUSE

Corporate credential theft is usually a targeted effort undertaken by adversaries who search social media sites such as Facebook and LinkedIn, searching for specific users with certain role types whose credentials grant them access to specific data. After they have a list of individual targets, they typically target them with phishing emails to redirect them to websites in order to obtain their credentials. Adversaries put a great deal of effort into making these phishing emails and websites look legitimate and nearly identical to authentic corporate communications.

After they steal credentials, adversaries either sell them or use them to gain access to systems and resources. After adversaries have obtained stolen credentials, they can appear and behave as a legitimate user and go unnoticed. At this point, it is often very difficult to identify an intruder and confirm that a user is the person their credentials represent them to be.

Credential abuse can be often hard to detect, so it's important to have complete visibility of user behavior and how and when credentials are used. The Cortex XDR analytics engine examines logs as they are streamed to Cortex Data Lake and analyzes data from sensors as soon as it arrives. The analytics engine establishes a baseline for normal user behavior. When abnormal behavior occurs, such as credentials being used on a different host, at a different location, or at a different time, the analytics engine raises an alert.

Using behavior analytics, Cortex XDR can find stealthy threats by pinpointing attacks before any damage is done. To detect abnormal activity that is indicative of an attack, Cortex XDR continuously analyzes user and device behavior. By examining rich data built exclusively for analytics, Cortex XDR can detect multiple attack types that are typically very difficult to identify, such as credential theft. Therefore, identifying credential theft requires identifying multiple actions that have been used in combination, which are indicative of abnormal behavior.

The types of actions and abnormal behavior related to credential theft could include brute force login attempts against a service or sending random data to an application. Sending random data to an application or authentication service often is an attempt to look for vulnerabilities, known as *fuzz testing*.

Multiple BIOC rules exist within Cortex XDR. The Palo Alto Networks research team defines the rules, and you can find them in the Rules section of the Cortex XDR app. The rules are delivered through regular content updates. You can edit, update, import, and export the BIOCs, and you also have the option to add additional BIOC rules for your network or endpoints. When creating your own BIOCs, you need to ensure you test the behavior of the rule because a general rule that isn't sufficiently defined can create too many alerts.

BIOCs are displayed within the Cortex XDR app with the following main fields:

- **Modification date**—The date BIOC was created or modified
- **Name**—A description of the BIOC and what it does
- **Type**—The Type field includes BIOCs attack tactics, such as credential access, privileged escalations, evasion, tampering, reconnaissance, and exfiltration.
- **Behavior**—Indicates whether the BIOC is a network, process, file, or registry behavior, such as the credential access example:

```
Process [ action type = execution AND name = gsecdump.exe  ]
```

- **Severity**—BIOC severity defined when the BIOC was created, such as informational, low, medium, and high
- **Number of hits**—The number of hits (or *matches*) on this behavior
- **Source**—Who created the BIOC, such as Palo Alto Networks or a SOC analyst

The BIOC behavior example above is a rule that looks for credential dumping via gsecdump.exe, which is a hacking tool used within a command-line interface to dump the Windows security account management database, cached domain credentials, local security authority details, and active logon sessions.

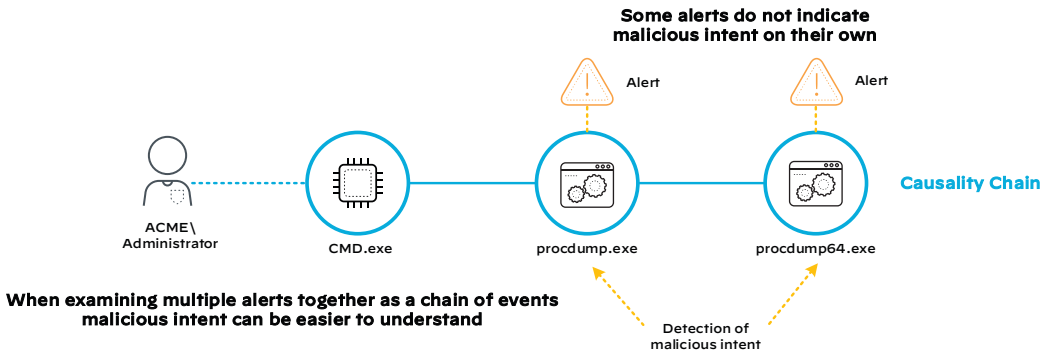
Within the Cortex XDR app's **Investigation > Incidents** section, the description of each incident indicates the incident's type, such as the following credential abuse example:

```
Dumping lsass.exe memory for credential extraction' alerts detected by XDR
BIOC on host msft-rmsclient-w10 involving user ACME\Administrator
```

In this example, we can determine that lsass.exe on the host msft-rms-client-w10 with the user Administrator from the Acme domain is attempting to dump credentials.

Analyzing the incident then lists all the alerts associated with the incident, including the alert sources, such as an XDR Analytics BIOC, XDR BIOC, PAN Next-Generation Firewall (NGFW), XDR IOC, XDR analytics, or XDR agent. You can analyze each alert individually, or if you click on one of the related alerts, you can access the causality view (shown in the following figure).

*Figure 20  Causality view of credential abuse*



In this figure, you can see the causality chain, which includes alerts as part of the chain, and two of the alerts indicate some potentially malicious behavior. You can view the details for each specific alert by clicking on each alert in the causality view. This provides visibility of all the processes, files, and modules used during each alert. In this incident example, Cortex XDR has discovered that Procdump, a command-line utility that is normally used for monitoring CPU spikes and creating crash dump files, is being used in this example for dumping user credentials to a local file.

A SOC analyst at this point could remediate the host and continue to investigate the incident to determine if other hosts were affected by this threat. The analyst could also create additional BIOC rules to look for similar types of behaviors. The SOC analyst could also use the Query Builder tool to search on process execution with Procdump. The query could focus on the executable file, hash, or username to determine the extent of the credential theft within your environment over a particular time period.

## DETECTING AND INVESTIGATING LATERAL MOVEMENT

Lateral movement is a technique that can be used to identify, gain access to, and exfiltrate sensitive data. The adversary uses different tools and methods to gain higher privileges and access, allowing them to move laterally between devices and applications and move through the network in order to map the environment and identify key targets, eventually reaching the organization's electronic assets.

As mentioned in the Prevention section, adversaries can use lateral movement techniques such as port scans, vulnerability scans, network mappers, and remote access tools to gain access to additional resources within your network for accessing and exfiltrating your data. Lateral movement can be difficult to detect without the correct detection and investigation tools in place to provide context, visibility, a forensic timeline, and insight. This is due to the fact that many lateral movement activities might be from legitimate users or from known and supported applications. This is where behavioral analysis and understanding the full chain of events is important.

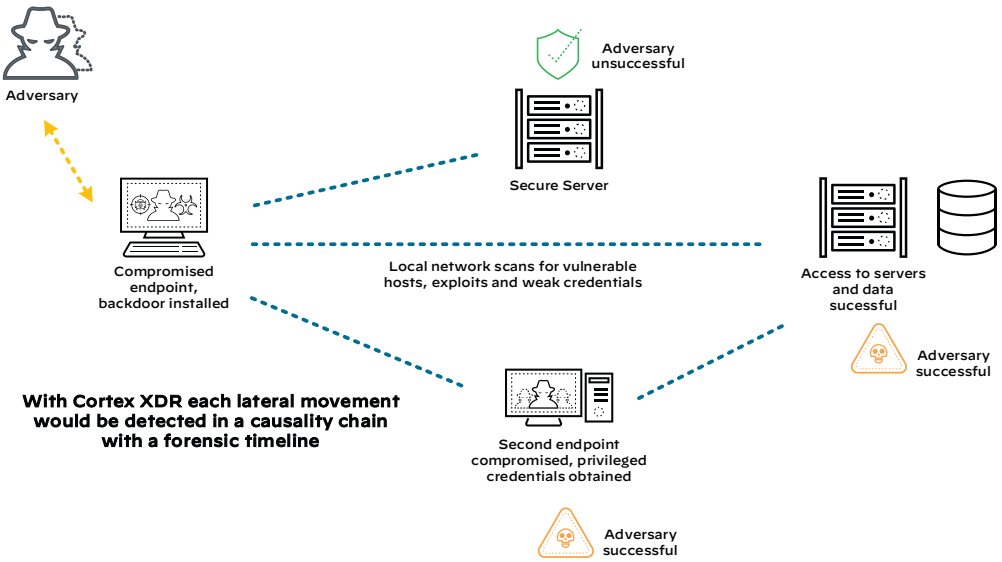
Cortex XDR correlates the data from your detection sensors across the network, endpoints, and cloud in order to provide a complete account of the activity surrounding an event. If the firewalls detect malicious network activity, Cortex XDR correlates that activity with endpoint logs in order to assess the impact of the activity and identify the cause. Reported information includes details about the process that generated the network behavior and its context, including the application and user who executed it. This is achieved, as mentioned earlier, through log stitching, which streamlines detection, helps with investigation and threat hunting, and reduces overall response times by eliminating the need for manual analysis across separate tools and platforms for each of the individual data sensors.

As discussed previously in the malware example, within the Cortex XDR app, viewing incidents and associated alerts is a quick way to easily assess and understand the most urgent and threating events. In a single view, the Cortex XDR app highlights the open incidents and any malicious activity requiring investigation.

During the investigation of incidents within the Cortex XDR app, you might discover after analyzing an incident and viewing the causality chain that one of the individual alerts within that chain indicates that malicious network activity is present as part of a sequence of events. The malicious activity could include some form of lateral movement after an endpoint was compromised and indicate suspicious network activity, such as port scans and reconnaissance activities.

The lateral movement could be indicated by network activity such as a nonstandard application connecting to the network or an application that a given user would not normally use. You can analyze the alert or investigate its timeline to see how it relates to other malicious activity that is part of the overall attack. This quick analysis allows you to act on the information rapidly to block or stop the threat.

*Figure 21 Lateral movement*




The types of alerts that Cortex XDR generates based on lateral-movement detection are activities such as port scans, failed connections, high connection rates, DNS tunneling, and abnormal session types such as SSH to a remote endpoint. Often when a host is compromised, an adversary is connecting to another host or hosts. Activity such as multiple consecutive sessions and high connection rates are key indicators of lateral movement activity. The Palo Alto Networks Cortex XDR Analytics Alert Reference guide provides detailed information on all the alert types available in the Cortex XDR app.

When working with alerts within the Cortex XDR app, you have the option to right-click on an alert and create an EDL to block the network traffic. EDLs allow the enforcement of security policy on firewalls, which they can use to block IP addresses and domains from being accessed. To enforce policy on the entries included in the EDL, you must reference the list in a supported policy rule or profile. As you modify the list, the firewall dynamically imports the list at the configured interval and enforces policy without the need to make a configuration change or commit on the firewall. You can also add EDLs within the **Response > EDL** section of the Cortex XDR app.

As discussed earlier, the Cortex XDR Query Builder can search on multiple entities, including network activity. Using the Query Builder, you can search on network activity by using All, Incoming, Outgoing, and Failed, based on the following search parameters:

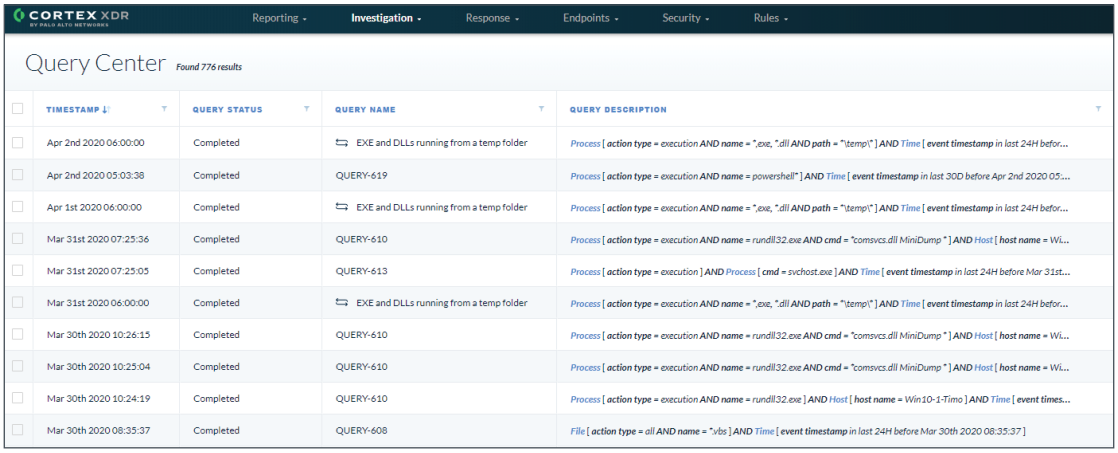
- **Local IP address**—The local IP address or range of the local host(s) that triggered the alert
- **Local port number**—The port number of the local host(s) that triggered the alert
- **Remote IP address**—The IP address or address range where network traffic is going
- **Remote country**—Country where network traffic is going
- **Remote host**—Remote host to which traffic is going
- **Remote port number**—Remote port number to which traffic is connecting
- **Protocol**—TCP or UDP port of connection

The Query Builder offers powerful search queries that make it easy for analysts of all experience levels to find evasive threats without needing to learn a new query language. The query options are very granular and can include a search timeline such as the last 24 hours, last seven days, last month, or custom start and end times.

When running a query, you also have the option to include processes and hosts within the search in order to narrow the results. After the query has completed, you can view detailed information on the network activity. With the resulting information, you can respond to the threat or continue to hunt for related threats in your environment. You can save queries to run them again, modifying them as needed.

You can find all previously run queries in the Query Center. The Query Center lists all the details surrounding previously run queries, including when they were run, who ran them, a query description, and the number of results the queries found. The following figure shows a view of the Query Center. Right-clicking a query gives you the option to view saved query results as well as the ability to rerun or schedule the same query. Often if a query is giving you too many results, you can be more granular on specific network search settings, such as source and destination IP addresses, ports, and protocols.

*Figure 22 Query Center*



# DETECTING AND INVESTIGATING DATA EXFILTRATION

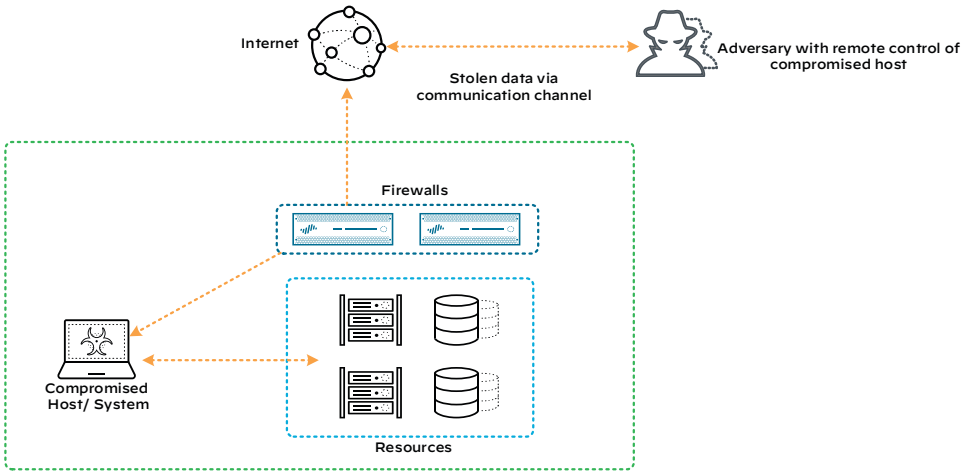
Sophisticated adversaries often remain undetected for a long period of time in an organization's network, even while actively hunting for valuable data. After an adversary has gathered enough data, they then attempt to exfiltrate it. An adversary's main objective during exfiltration is to send as much key data as possible out of the network as quickly as possible. This often leads to them implementing stealthy techniques and using common applications to minimize the chance of being detected.

Data exfiltration is often detected and acknowledged too late, after adversaries have achieved their impact or malicious damage. The last thing an organization needs is to be headline news for a data breach, as has been the case for many large corporations.

Data exfiltration can often be the effect of the high volume of alerts the SOC receives on a daily basis; too many false positives allow potential threats to be overlooked or investigated long after the event occurred.

Similar to lateral movement, detecting data exfiltration activities is difficult without the right visibility tools that have built-in intelligence, machine learning, analytics, and the ability to triage alerts from multiple sources. Because adversaries often use standard applications and session types that you would normally see in your environment, alerts can be missed after an adversary has infiltrated your environment.

*Figure 23 Data exfiltration*




After an adversary has compromised and gained access to a system within your network, they typically then establish a communication channel that enables remote interaction with a host. Often, DNS can be used to transfer files or facilitate command and control with a compromised host, especially in environments where other methods might be more easily detected. For example, if an adversary can issue external DNS queries from a compromised host within an organization's network, then they will likely be able to exfiltrate data by tunneling it over DNS.

The transmission of encrypted data over an unencrypted channel can be a strong indicator of suspicious activity, might indicate that the network has been compromised, and should be investigated further.

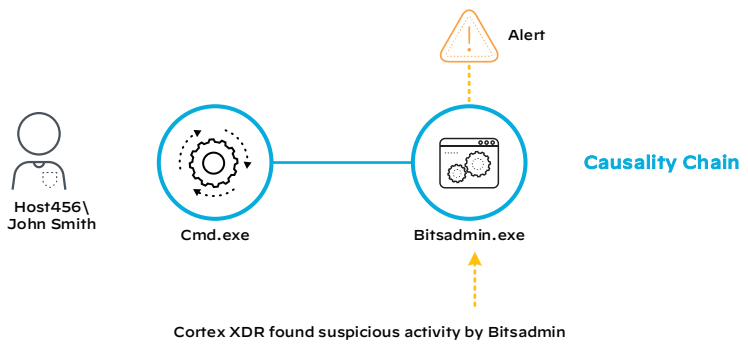
Cortex XDR uses behavioral analytics to profile normal versus abnormal behavior and find hidden threats. Cortex XDR can identify data exfiltration activities by examining multiple criteria, including the activity surrounding all your outbound connections, with a key focus on some particular areas of interest, such as the following:

- **The volume of data being transferred**—Cortex XDR detects when excessively large data transfers are occurring that are using data transfer protocols such as FTP, SMTP, and HTTPS. Adversaries often use these protocols and many others to exfiltrate large volumes of data. At a host or user level, Cortex XDR can analyze whether a large data transfer is abnormal behavior.
- **Remote destinations**—Cortex XDR detects data being transferred to remote destinations outside of your network, including non-corporate locations, different countries, and remote sites.
- **DNS tunneling**—Through detection, you can determine if there is an unusually high number of DNS transactions occurring in your environment. Cortex XDR detects abnormal endpoint activity, such as uncharacteristically sending and receiving DNS queries and responses, which can indicate command and control activity through DNS tunneling. Adversaries use DNS tunneling to encode data in DNS queries and responses in order to bypass firewalls and HTTPS traffic rules, allowing the adversary to execute command-and-control instructions or exfiltrate data.
- **Unusual upload session types and protocols**—Cortex XDR detects uploads and sessions types that are unusual and that are not normally seen in the environment, with a focus on detecting abnormal behavior for users or hosts.

By analyzing normal activity, Cortex XDR can identify different types of users and devices, what protocols the devices typically use, and to which destinations the devices typically connect. By profiling hundreds of types of behavior, Cortex XDR can understand the expected behavior. Using patented behavioral analytics and machine learning, Cortex XDR can identify hidden threats with precision.

Figure 24 shows an example of a data exfiltration incident, where a genuine application called Bitsadmin is used to exfiltrate data from a host, which is an activity that was detected with Cortex XDR. Cortex XDR found suspicious activity, including various file reads and file writes, it found additional modules that were loaded in path C:\windows\System32\, and it identified exfiltration activities initiated by Bitsadmin.

*Figure 24 Data exfiltration incident*



Cortex XDR is able to detect events like this because it tracks hundreds of different types of behavior, including traits that are nearly impossible to determine from traditional threat logs. It then profiles user and device behavior and detects any anomalies compared to previous behavior to uncover malicious activity, such data exfiltration.

The BIOC rules mentioned previously look for multiple attributes around data exfiltration behavior patterns. Table 1 lists a few examples of current BIOC rules created by Palo Alto Networks, but you can also create many of your own rules.

*Table 1 BIOC rule examples*

| Name | Type | Behavior |
|---|---|---|
| Execution of a CLI file transfer/ copy utility | Exfiltration | Process [ action type = execution AND name = robocopy.exe , ftp.exe , sftp.exe , scp.exe , pscp.exe , copy.exe , xcopy.exe ] |
| Unsigned process makes connections over DNS ports | Exfiltration | Network [ action type = Outgoing , Failed AND remote port = 53 ] AND Process [ signature = N/A , Unsigned , Invalid Signature , Weak Hash , cgo signature = N/A , Unsigned , Invalid Signature , Weak Hash ] |
| Commonly abused AutoIT script connects to a remote host | Exfiltration | Network [ action type = Outgoing , Failed ] AND Process [ signature = Signed , Weak Hash AND signer = Autoit* AND signer = Autoit* ] |

The BIOC rules in Table 1 are:

- **Execution of a CLI file transfer**—This BIOC rule is looking at specific patterns related to file transfer applications that are commonly used by adversaries to exfiltrate data.
- **Unsigned process makes connections over DNS ports**—This BIOC rule looks at unassigned processes making connections over DNS ports. This one is included here to highlight that the fact that many adversaries are known to use the DNS protocol to exfiltrate data and avoid detection.
- **Commonly abused AutoIT script connects to a remote host**—This BIOC rule looks at AutoIT scripts (a freeware programming language for Microsoft Windows), which have legitimate uses but are often abused by malware to execute in a signed process context.

Over and above these example rules, you can also run queries in Query Builder to look for activities related to data exfiltration.

## DETECTION AND INVESTIGATION SUMMARY

This section discussed the challenges the SOC faces today and covers how Palo Alto Networks can address them with Cortex XDR. These challenges include the need for reducing the number of alerts, which cause alert fatigue; getting visibility into what is occurring; understanding the context of alerts from multiple data sources; reducing the number of false positives; and getting to the root cause of an attack as soon as possible. Addressing these challenges enables the SOC to be more effective and efficient. Palo Alto Networks addresses these challenges with Cortex XDR.

Cortex XDR provides detection, investigation, and response capabilities and reduces risk by lowering mean-time-to-response and mean-time-to-detection, increasing the efficiency of your security team. Cortex XDR speeds up your investigation time by having the right data, integrated across the network, endpoints, and cloud, with all the context needed to provide automated root-cause analysis.

Within the Cortex XDR app, from the incident management view, analysts can right-click an alert to understand the root cause. Automated root-cause analysis dynamically associates parent and child processes into an easy-to-understand causality chain.

The Cortex XDR causality view is a fast way to determine graphically the chain of events related to an attack. Cortex XDR provides multiple response options from quarantine and endpoint isolation, firewall access controls with EDLs, and a powerful Live Terminal feature, providing the flexibility to extend investigations to the endpoint and quickly isolate and remediate threats.

# Response with Orchestration and Automation

The SOC faces multiple challenges in reducing their mean-time-to-respond, putting pressure on them to be effective. As a result, the SOC requires a solution that orchestrates tasks across different security platforms with built-in automation that minimizes repetitive tasks and enables real-time threat response, reducing the overall incident response time. With an effective security orchestration, automation, and response (SOAR) solution, it's possible to achieve more, in less time, while still allowing for human decision-making when it's most critical.

This section covers how you can assist the SOC with the orchestration, automation, and response capabilities offered with Palo Alto Networks Cortex, specifically focusing on the Cortex XSOAR Platform.

With Cortex XSOAR, you can ingest Cortex XDR incidents for playbook-driven enrichment and response. Playbooks allow SOC analysts to focus on other tasks, such as threat hunting.

## WHAT IS SOAR?

The first component of SOAR is *security orchestration*, which involves controlling and activating multiple and multi-vendor security products from a central location. SOAR products do this through either runbooks or playbooks.

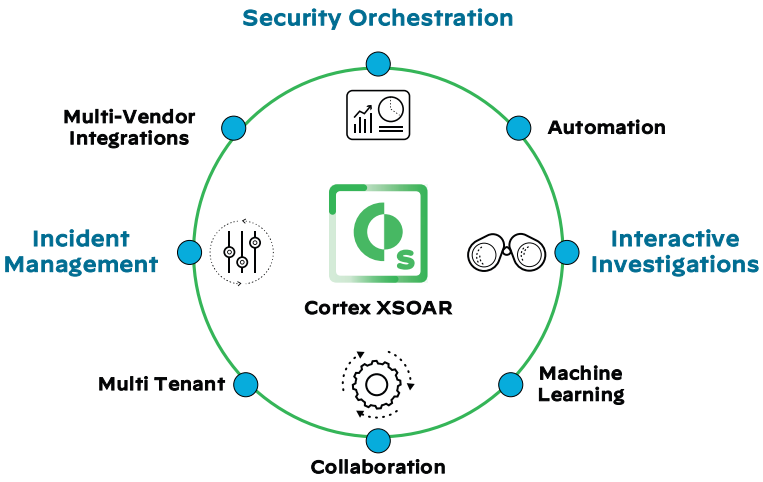
The second component of SOAR is *security automation*. Automation is a subset of orchestration and involves finding repeatable tasks and executing them at machine speed. To accomplish this, SOAR products have automation scripts and extensible product integrations.

The final component, *security response*, involves maintaining oversight of an incident as it goes through the incident lifecycle. Within SOAR products, this includes case management, collaboration during investigation, and analysis, remediation, and reporting after incident closure.

In the industry, the terms *security orchestration* and *security automation* are often used interchangeably, but these terms are different. Security automation is making machines do human-based tasks with machine-based execution. Security orchestration is making both security and non-security products integrate with each other and automating different tasks across products through workflows, while also allowing for end users to oversee and interact with these tasks.

Figure 25 shows a high-level overview of the components of SOAR, with Cortex XSOAR at the center of it, which is covered in the next section.

*Figure 25 SOAR*




Security orchestration combines the interaction of people, processes, and technology to improve and secure an organization's environment. Security orchestration achieves this by unifying security processes, connecting independent security products, and balancing security automation with human intervention. With security orchestration, security analysts are empowered to effectively and efficiently carry out security operations and incident response.

Additional capabilities include case and incident management features; the ability to manage threat intelligence, dashboards and reporting; and analytics that you can apply across various functions. SOAR tools significantly enhance security operations activities like threat detection and response by providing machine-powered assistance to human analysts to improve the efficiency and consistency of people and processes.

## CORTEX XSOAR

Cortex XSOAR is a SOAR solution that manages alerts, standardizes processes, and automates responses, giving your security teams more time to be more proactive and effective by eliminating multiple trivial manual tasks. Cortex XSOAR combats security challenges facing security operations with three main areas of focus: workflow automation, incident management, and collaboration.

## Respond and Automate

Cortex XSOAR makes workflow automation possible through an extensible integration network with hundreds of security and non-security products. These integrations are powered by thousands of actions that can be remotely executed within Cortex XSOAR, either as automated tasks or in real time. For example, you can automate tasks such as looking up a URL's reputation, quarantining an endpoint, detonating a file in a sandbox, and sending an email.

All these tasks can be coordinated by using a drag-and-drop visual playbook editor that allows for playbook reuse, nesting, and a combination of automated and manual tasks. Cortex XSOAR's workflow automation helps you respond to incidents with speed and scale.

Cortex XSOAR has multiple integrations with Palo Alto Networks platforms, including WildFire, Cortex XDR, Panorama, NGFWs, AutoFocus™, and MineMeld™, and also integrates with hundreds of products from third-party vendors. Some of these are discussed later, in the individual threat examples.

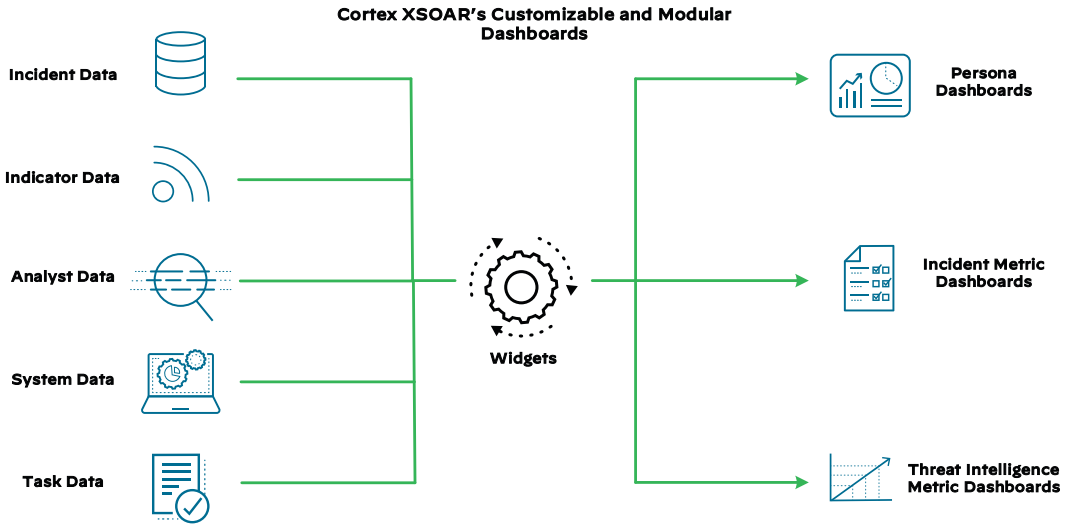
## Manage Incidents

The second area of focus is incident management and ticketing. Cortex XSOAR can ingest alert data across a range of different sources, including SIEMs, network security tools, email inboxes, vulnerability management, and cloud-security solutions. You can intuitively search these alerts and query according to various parameters, helping you analyze and segment the data at your disposal.

You can assign custom SLAs and link them to playbook execution, resulting in an accurate measurement of your key performance indicators. Case management helps you standardize processes across products, teams, and use cases, while also providing the flexibility needed to adapt to emerging threats. The full incident management suite in Cortex XSOAR enables end-to-end oversight of incidents and is fully customizable, allowing you to create your own incident types, incidents, and summary layouts. Incident management also offers the following capabilities:

- **Incident repository**—A Cortex XSOAR database consisting of incidents from multiple sources. The database has full search-and-query capabilities, includes details and context, and provides visualized cross-sections of data.
- **Evidence board**—For root-cause discovery, an evidence timeline allows you to reconstruct attack chains and piece together key pieces of verification.
- **Dashboards and reports**—With Cortex XSOAR, you can fully customize dashboards and reports with a user-driven widget library. This allows visualization of customized metrics both in real time and for the future. The default dashboards provide visibility into metrics at the analyst, incident, and business levels. You can also create your own dashboards by using a fully customizable widget editor, enabling the creation of customized visuals representing the underlying data.

*Figure 26 Visibility and reporting*




## Collaborate and Learn

The third focus area is real-time collaboration and investigation. To accomplish this, each Cortex XSOAR incident has a War Room view, which is a shared workspace where security analysts can chat with one another and conduct joint investigations.

The War Room also enables third-party product actions to be executed in real time through a command-line interface, ensuring that you take advantage of your security product stack with minimal screen-switching and dead time. All actions and commands are automatically documented in the War Room, preventing the need for manual post-incident documentation, where important information can fall through the cracks. The Cortex XSOAR War Room helps you improve investigation quality by working the way humans are meant to work—together.

## Cortex XSOAR Deployment

To serve security teams across the incident lifecycle, Cortex XSOAR is a comprehensive SOAR technology that you can deploy both on-premises and as a hosted solution in the cloud. You can access Cortex XSOAR via the hub. Cortex XSOAR supports full multi-tenancy with data segregation and scalable architecture.

Cortex XSOAR automates alert ingestion from Cortex Data Lake, real-time execution of response actions within Cortex XSOAR, and unified activation of your security product stack through task-based playbooks.

## WORKFLOW: RESPONDING TO MALWARE

Security teams often spend valuable time performing repetitive tasks because they coordinate between many security tools, causing them operate multiple consoles simultaneously. If SOCs use different solutions for endpoint protection, malware analysis, and threat intelligence, the incident-related information is often very fragmented. With fragmented information, it becomes very challenging and time consuming to obtain a centralized view of incidents and their overall impact.

To address this challenge, Cortex XSOAR gives you a single centralized and consolidated view for incident management, malware analysis, and threat intelligence. Cortex XSOAR automates the creation of a ticket for the malware incident and the execution of different tasks through pre-defined playbooks, including tasks such as automated remediation and response actions. Depending on your preference or requirement, playbooks can be fully automated (without SOC analyst involvement) or semi-automated (with reduced SOC analyst involvement). As the SOC deals with more and more of the same type of incident, they then feel more comfortable automating the response without human intervention.

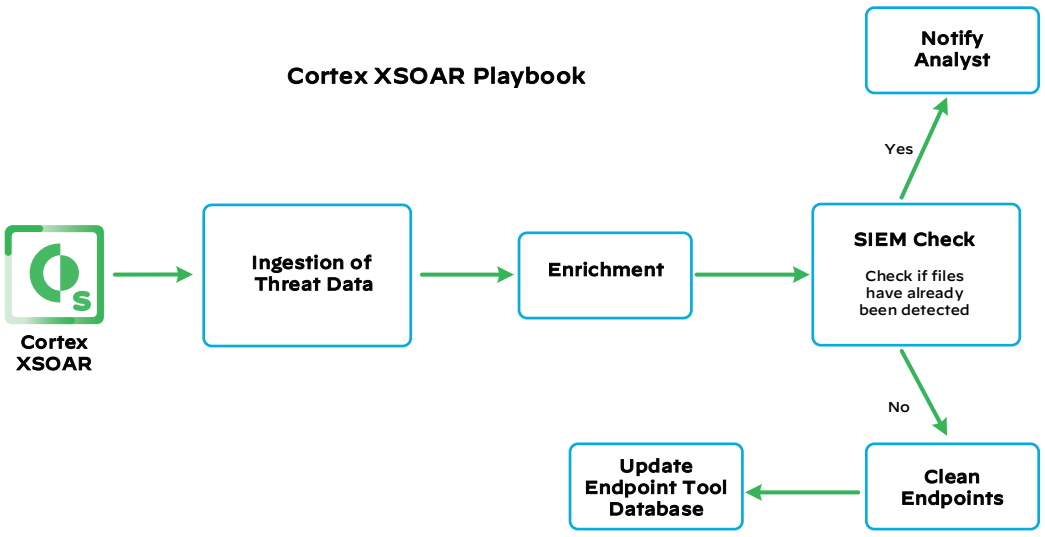
With Cortex XSOAR deployed, you can ingest multiple alerts, including malware-related alerts from Cortex XDR as well as other detection sources that can automatically trigger playbooks. These playbooks have pre-defined automation scripts that can enrich incidents by gathering additional data, create an incident ticket, assign incidents to a SOC analyst, or invoke a response action, such as quarantining or isolating an endpoint.

Cortex XSOAR comes with multiple pre-built playbooks available, and you can also create many of your own custom playbooks. To give your organization the best chance at combating malware, you can use these task-based playbooks for malware analysis, threat intelligence, incident enrichment, and endpoint protection. Playbooks can be used across Cortex XDR and the multiple product integrations that Cortex XSOAR supports, including third-party vendors' endpoint-protection tools and multiple SIEM platforms.

When using playbooks in Cortex XSOAR, a malware incident might have the following workflow:

1. Threat data is ingested via the integration APIs of Cortex XDR, SIEMS, and third-party vendors' endpoint-protection tools.
2. The IOCs are extracted from the data received and enriched with additional data from Palo Alto Networks Cortex XDR, WildFire, SIEMs, and other threat-intelligence and malware-analysis tools. IOCs can include similar patterns, such as file hashes, file types, IP addresses, and domains.
3. The playbook checks whether any of the indicators are identified as malicious. If they are, the playbook then increases the incident severity, opens a ticket, and notifies a security analyst for review and investigation.
4. The playbook checks the SIEM in order to determine if these files have already been detected, resolved, and removed. If they have, Cortex XSOAR notifies the analyst via an email stating these files were detected and previously known by the SIEM.
5. If these are new indicators and they are determined to be a threat, then a trigger executes a task to clean the endpoints. For malicious indicators, you can automate multiple cleaning actions on the endpoints, which could include activities such as running queries to kill any malicious processes, removing any malicious files, or quarantining and isolating the infected endpoints.
6. The playbook updates the Cortex XDR and other endpoint tool databases with the new file information, in order to avoid any future events with the same malware.
7. If the indicators are not identified as malicious, the playbook then records the context for future reference and automatically closes the incident. The playbook does not notify a security analyst, which reduces the number of false positives requiring human review and saves the SOC time.

*Figure 27 Malware incident example*



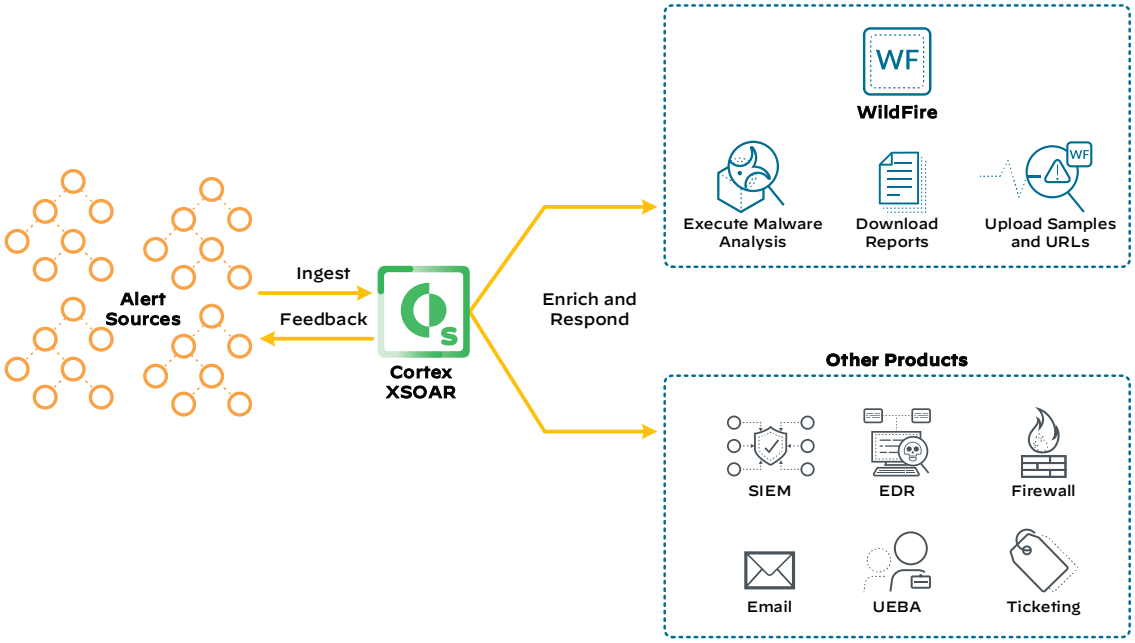
Security orchestration playbooks for malware or any other incident types can unify processes across SIEMS and endpoint tools in a single workflow, providing automation of repetitive and time-consuming steps before having to involve an analyst.

Cortex XSOAR offers automated malware analysis and response through its integration with Palo Alto Networks WildFire. WildFire provides detection and prevention of zero-day malware by using a combination of malware sandboxing and signature-based detection and blocking. With Cortex XSOAR, your SOC can use a standardized playbook that runs automatically and queries WildFire for malware analysis. These playbooks can perform checks to start triage, detonate files, and submit reports to the analysts for follow up and further investigation.

The advantage of using playbooks is saving the SOC time and eliminating repetitive and redundant tasks by automating triage and detonation tasks. This allows the analyst to invest their time and effort into more advanced investigative actions and assists with standardizing responses, reducing error rates, and ensuring that no alert is overlooked.

Figure 28 highlights Cortex XSOAR integration with WildFire and how WildFire provides feedback and highlights how Cortex XSOAR can enrich and respond to other products, such as SIEMs and EDR tools.

*Figure 28 Automated malware analysis and response*




In certain situations, just using playbook orchestration might not be a sufficient response to an incident. Attack investigations often require some additional real-time tasks, such as gathering critical evidence from multiple suspicious indicators, looking at similar patterns between incidents, and coming to a resolution. Normally, running these commands forces analysts to switch between consoles during the investigation, which makes it difficult to provide complete documentation at the end of an investigation. Often a SOC analyst needs to work with another analyst to discuss what they are seeing or work with someone who is experienced with a particular type of threat activity. Collaboration is key for the SOC.

After playbook execution, analysts can conduct joint investigations in the Cortex XSOAR War Room by using the ChatOps collaboration environment and running WildFire-specific commands in real time. For example, if the playbook for a particular incident extracted a hash, analysts can run the file command to obtain more information for that hash.

Sometimes, SOC teams might want to get verdicts for a large list of hashes. You achieve this by running the **wildfire-get-verdicts** command, which returns the verdict for all the hashes in the list as either good or bad. As an example, if two of the hashes have a verdict of bad, the SOC could run the **wildfire-get-sample** command to retrieve relevant samples and add them to incident context.

SOC teams can also run commands from hundreds of other integrated products in the War Room, from a single integrated platform for collaboration, investigation, and documentation of all actions. The benefit of the War Room is that all analysts have full task-level visibility of the process that was followed and are able to run and document commands from the same console. This eliminates the need to collect and merge information from multiple sources for documentation purposes.

Having a full audit trail of the tasks and actions taken to remediate a particular malware threat makes it easier for the next SOC analyst who works on a similar incident.

## WORKFLOW: RESPONDING TO CREDENTIAL ABUSE

Sophisticated attacks often involve compromising user credentials in one form or another. Adversaries can deceive and bypass traditional security products in order to launch large-scale attacks by using stolen credentials. They can use existing tools in the target environments and leverage SSL to disguise malicious activity. Security teams need a platform that provides insightful and real-time network intelligence and utilizes the information to drive actions across security environments.

Adversaries can impersonate employees to compromise and exfiltrate business data. Stolen credentials can enable an adversary to steal data without being detected because the user session appears legitimate. To meet these challenges, Cortex XSOAR can combine the network detection and investigation capabilities of Cortex XDR as well as integrations with other solutions such as C2Sec and Awake Security.

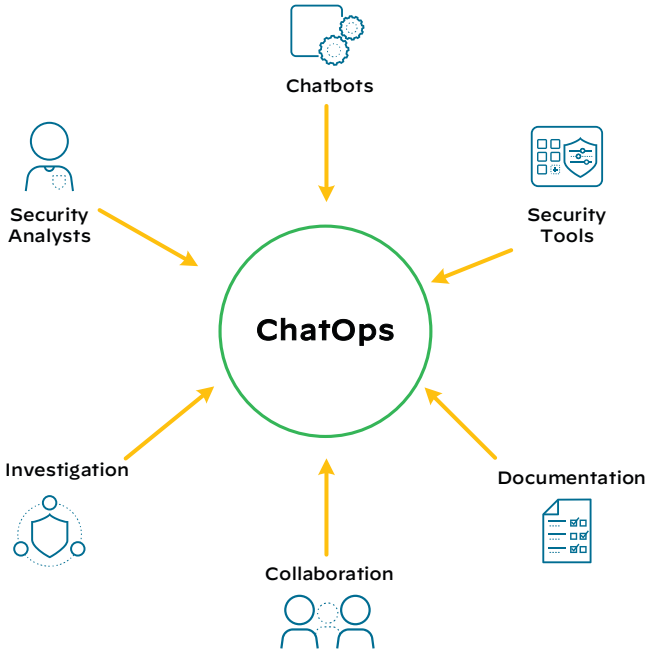
The integrations from multiple sources provide triaged data for the detection of compromised credentials and automatic data enrichment with playbooks. Over and above the orchestration and automation capabilities with playbooks, the Cortex XSOAR War Room assists the SOC by speeding up collaboration, investigation, and resolutions of credential abuse activity. Often with advanced threats, you need an experienced analyst to work on the incident, and without machine learning, it's hard to determine who is best suited for the task.

Unfortunately, as SOCs scale, they end up following a "first analyst available" approach for assigning incidents to analysts. Not only can this lead to uneven distribution of the workload amongst analysts, but it doesn't take analyst expertise into account when assigning incidents. This eventually results in incidents that are not efficiently handled. When assigning analysts to incidents, the Cortex XSOAR machine-learning capability evaluates previous incidents in the system, including incident types and a variety of fields. These details are cross-referenced against the analysts' current workload to suggest which analysts are best assigned to the incident.

The Cortex XSOAR machine-learning capabilities help to increase response productivity, enabling learning and more efficient security operations. *DBot* is a security chatbot and a component of Cortex XSOAR that helps analysts by providing suggested indicators and incidents that might be connected or relevant to the incident on which they are working.

With ChatOps, analysts can perform all three actions—investigation, collaboration, and documentation—using the same solution, without having to change windows and while leveraging the power of chatbots and other security tools.

*Figure 29 Security ChatOps*




DBot tracks information and educates itself with machine learning to provide automation capabilities, and it interacts with analysts by providing information and enabling War Room collaboration for incidents that require human intervention.

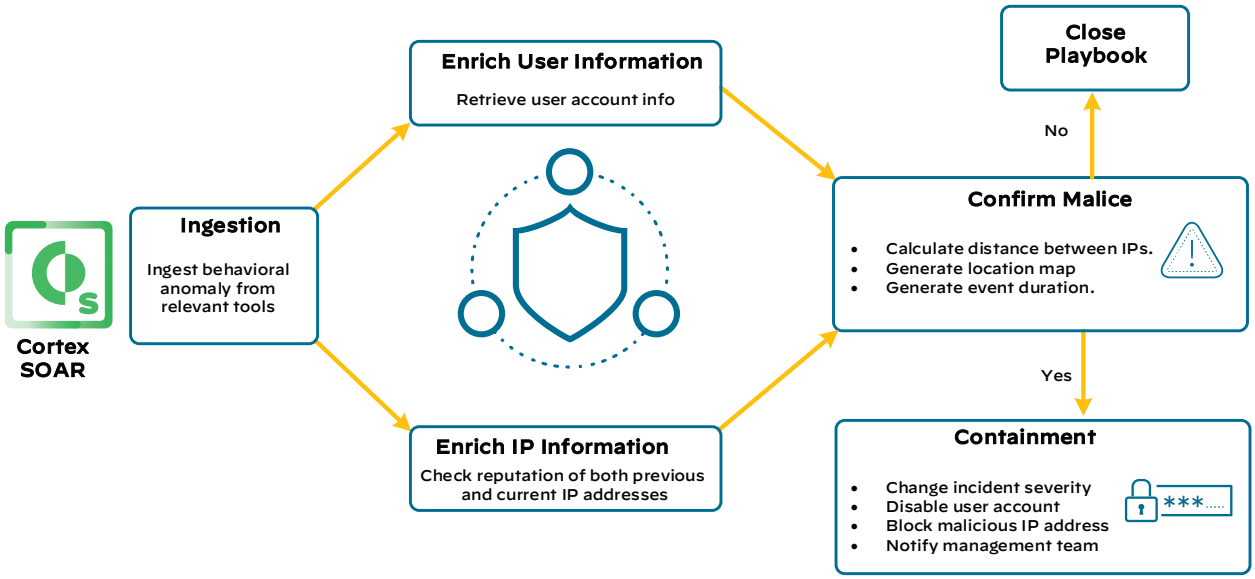
By running different security commands in the Cortex XSOAR War Room, analysts can improve their visibility and learn new, actionable information about the attack. For example, if playbook results bring up a set of artifacts, analysts can run commands to access other devices that match the same set of artifacts, within a specific time interval. The analyst can use a rich set of integration-supported queries to view the set of matching devices, domains, or activities.

From a common window, the War Room allows analysts to quickly adjust and run specific commands relevant to incidents in their environment. All analysts have full task-level visibility of the process and are able to run and document commands from a single console. This helps eliminate false positives and improv response times by collating information from multiple sources of documentation.

Multiple analysts can collaborate and work on a credential-theft incident in the War Room to determine how the credentials were obtained, if the stolen credentials have been used already, and what systems have been compromised. In the War Room, you can ask an analyst for help or ask questions to DBot, such as Microsoft Active Directory details for the given user whose credentials were compromised.

The following example use case of credential abuse leverages Cortex XSOAR and a pre-built playbook called the *Impossible Traveler Playbook*. This playbook looks for activity where two IP address are tied to the same user but are geographically a great distance apart, indicating stolen and compromised credentials.

*Figure 30 Impossible Traveler Playbook*




The playbook can run automatically or be manually run by an analyst. The playbook creates an incident when a user logs in from a great distance away from their previous point of login, and the time between those log-in sessions is short, such as 5 minutes. Enriched IP address information ingested from multiple sources would then show the user had a previous IP address in location A and now the same user has a different IP address logging in from location B. The physical distance between the two sessions is calculated, and they're a few thousand miles apart. In this case, it would then be determined that the user's credentials have been exploited.

After any credential abuse has been discovered, the immediate priority is containment. With Cortex XSOAR, DBot can disable the user account in Active Directory, update the firewall rules to block the IP address, and the notify the management team with a detailed report on the incident, especially if more than one user is affected.

After the incident has been completely resolved, you can create a PDF report that details the entire investigation from start to finish, including all the activities, artifacts, analyst interactions, and DBot involvement related to the incident.

In addition to the out-of-the-box Cortex XSOAR reports, you can create custom reports that allow you to specify widgets, data sources, graph types, and more. *General* reports display statistic data, such as the number of open incidents, and *Investigation* reports display investigation-specific data such as an investigation summary. Cortex XSOAR provides reports in DOC, CSV, and PDF formats.

Reports can be related to time intervals or incident data. Time-based reports are incident summaries over a period of time, such as the last 24 hours or last 30 days. Incident data reports are filtered summaries of incident data, such as open, late, or critical incidents. A report of incidents related to credential abuse can provide a baseline to implement new security policies.

## WORKFLOW: RESPONDING TO LATERAL MOVEMENT

After you detect and investigate lateral movement in the network, response normally requires blocking IP addresses, URLS, and applications.

Because Cortex XSOAR ingests data from multiple sources through the many supported integrations such as Cortex XDR, you have the opportunity to gain more insight into malicious activities on the network that might not be detected by individual security sensors such as your firewalls and endpoints. Enriched data from these multiple sources and Cortex XSOAR machine learning helps you detect the stealthy threats that are often overlooked on individual sensors.

With Cortex XSOAR, SOC analysts can choose to investigate certain incidents that you might not want to automate until completing further analysis, possibly as part of a new threat-hunting campaign to find a larger threat. Having data from multiple sources and historical data correlated and available in a single console makes it easier to detect stealthy attacks that are part of a larger attack campaign, such as lateral movement. An analyst using Cortex XSOAR can correlate and analyze indicators across incidents.

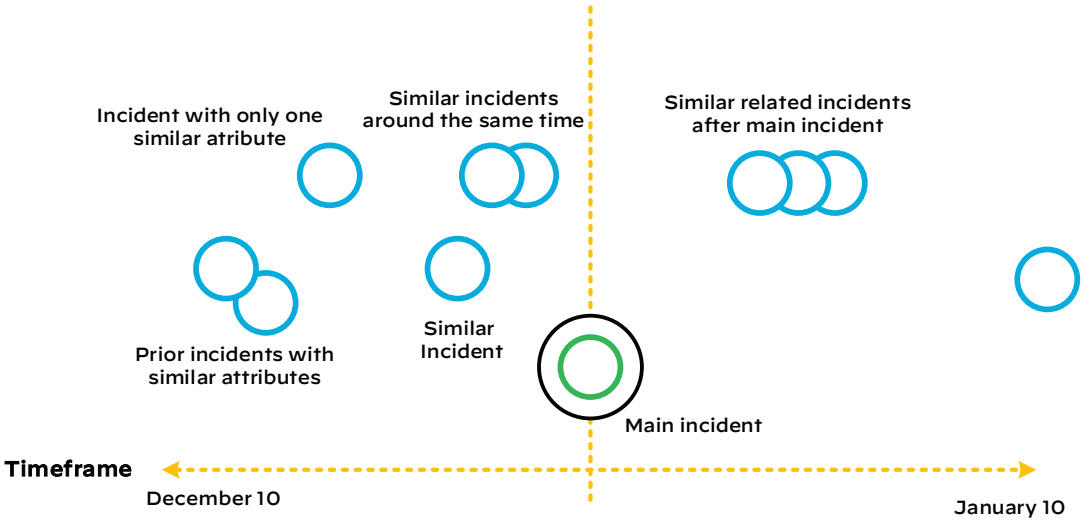
An example of a lateral movement activity could be a network port scan. After you have detected it, you can stop it through automation, or you can choose to investigate further. Through automation, the set tasks in a playbook can block the session on the firewall, but in case it is a legitimate action within the network by an authorized user, you might want to do additional verification first.

In our port scan example, an incident type could come from multiple sources, such as Cortex XDR or a third-party platform. The incident has certain indicators, such as IP addresses, URLs, or malicious domains. Within the incident, you have the opportunity to perform auto-remediation, view related incidents, assign the incident to another analyst, or go into the War Room to investigate the incident with other analysts. Cortex XSOAR offers a lot of flexibility because parts or all of these tasks could be run as playbooks.

Within the Cortex XSOAR GUI, you can customize your own dashboards with multiple views and, when selecting the incidents themselves, view the incidents in a table view or summary view. Each incident has a unique incident identifier number, name, type, severity, owner, time of occurrence, and playbook (if set on the incident type). You can find detailed information on each incident type by clicking it. The incident provides detailed information, including the indicators that are relevant to this incident. Indicators also show if they are related to other incidents, which can help with investigating related events.

Within an incident, an analyst can view graphically all related incidents and how close together the incidents are within a certain timeframe. Incidents are considered related if they share common attributes and indicators. In the GUI, the large dot in the center indicates the main incident, and each smaller dot represents a related incident. When you hover your mouse over a dot, more information about the incident appears. The x axis indicates the related incidents' proximity to the main incident in time, and the y axis indicates the related incidents' similarity to the main incident.

*Figure 31 Related incidents*

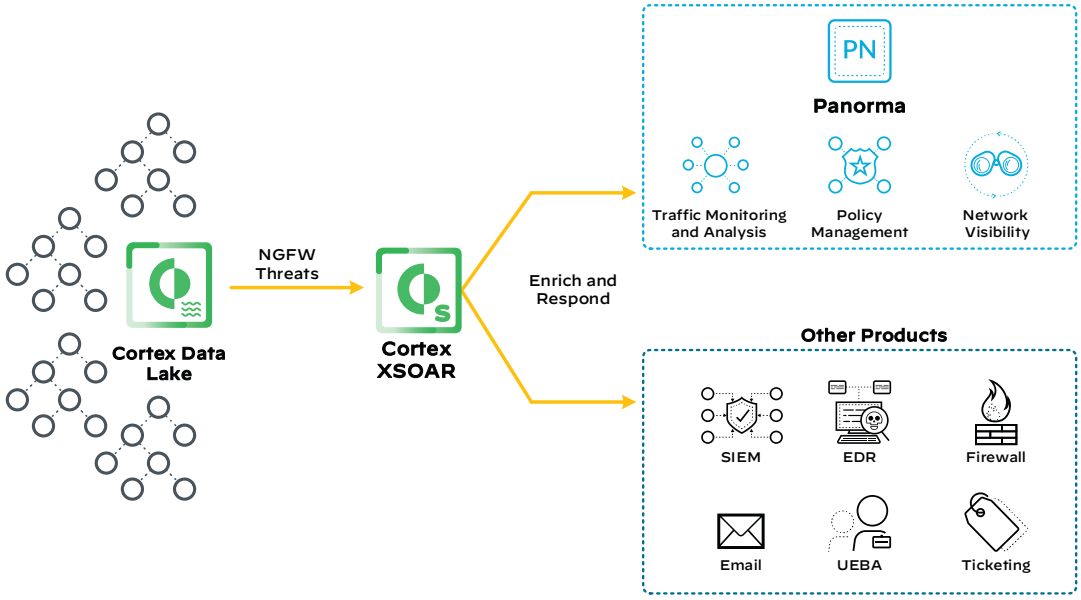



The related incidents view is a powerful way to find all incidents related to the one you are investigating, which provides additional information about the active threat, enabling faster resolution. In this example, if you find that there are multiple port scans occurring from one or more hosts, you can then use a playbook to automate actions in order to reconfigure the firewalls through the Firewalls platforms integration with Cortex XSOAR. The next example demonstrates using the Panorama integration with Cortex XSOAR.

Panorama offers easy-to-implement and centralized management features to gain insight into network-wide traffic and threats and to administer your firewalls everywhere. The integration with Cortex XSOAR and Panorama offers the following capabilities:

- **Object handing**—Automate actions on Panorama objects, which include IP addresses, address groups, services, URLs, and custom URL categories.
- **Security rule management**—Apply, edit, and update rules for Panorama objects through playbook-driven tasks within Cortex XSOAR.
- **Log and PCAP access**—Search and query log and packet-capture data from Palo Alto Networks firewalls within Cortex XSOAR.
- **Alerts**—Get alerts from a next-generation firewall using Cortex Data Lake by using a log-forward object within Cortex XSOAR.
- **Response**—Dynamically add IP addresses to Panorama address groups within Cortex XSOAR, without committing the instance each time.
- **Integrations**—Leverage hundreds of Cortex XSOAR product integrations in order to enrich data from Panorama and coordinate responses across security functions.
- **Execution**—While collaborating with other analysts and the Cortex XSOARs chatbot, run thousands of commands (including for Panorama) from the ChatOps interface.

*Figure 32 Panorama integration with Cortex XSOAR*



SOCs can integrate Panorama with Cortex XSOAR for both ingestion of alerts and playbook-driven response actions. Through integration with Cortex Data Lake, Cortex XSOAR can ingest NGFW alerts that trigger playbooks responsible for orchestrating and automating a variety of critical and repeatable actions during incident response. Highlighting a lateral-movement remediation example, Cortex XSOAR playbooks can dynamically add IP addresses to Panorama address groups without needing to manually commit the instance each time.

The Panorama example is one of many Cortex XSOAR integrations where either playbook automation stops lateral movement activity upon detection or the SOC first investigates the threat.

## WORKFLOW: RESPONDING TO DATA EXFILTRATION

Data exfiltration is normally the objective and the last stage of a cyberattack. The risk of a high-impact data exfiltration increases with the amount of time the adversary spends in the network. Effective incident response in Cortex XSOAR uses a combination of automation and proactive threat-hunting techniques in order to eradicate and isolate threats as quickly as possible, limiting the impact of the adversary's access.

To minimize impact to your information assets, a rapid response to an intrusion is extremely important. Although SOC analysts and incident responders deal with so many incidents and task-related investigations daily, their time-to-respond and time-to-resolution can never beat an automation with task-based workflows. Automation is also the answer to the shortage of security professionals and the time needed to respond to the barrage of alerts the SOC faces.

If data exfiltration has been detected and an incident has been created, you can use an existing playbook or create your own in order to prevent or block an existing data exfiltration activity that can be applied to an existing incident or a new incident that you have created. With the integration of Palo Alto Networks PAN-OS®, you can apply playbooks directly to the NGFWs or to Panorama if you are using Panorama for firewall management. Based on the playbook input, the playbook commits the configuration to the NGFW or pushes the configuration from Panorama to predefined device groups of firewalls.

To stop an exfiltration incident, you can automatically respond with multiple types of actions. For example:

- Create custom security rules in Palo Alto Networks PA-Series firewalls, VM-Series firewalls, and Prisma Access, such as application filtering, data filtering, and file blocking.
- Creating, deleting, and updating address objects, address-groups, custom URL categories, and URL filtering objects. Address objects can consist of FQDNs or IP addresses and IP address ranges.
- Add or remove sites from a custom URL category. A *custom URL category* is where you can create a custom list of URLs and use it in a URL Filtering profile or as match criteria in policy rules.
- IP addresses can be blocked by using registered IP tags from PAN-OS, without having to commit the PAN-OS instance.
- Create, edit, and delete EDLs.
- Run commands supported by the PAN-OS API, such as **show**, **get**, **set**, **edit**, and **delete**.

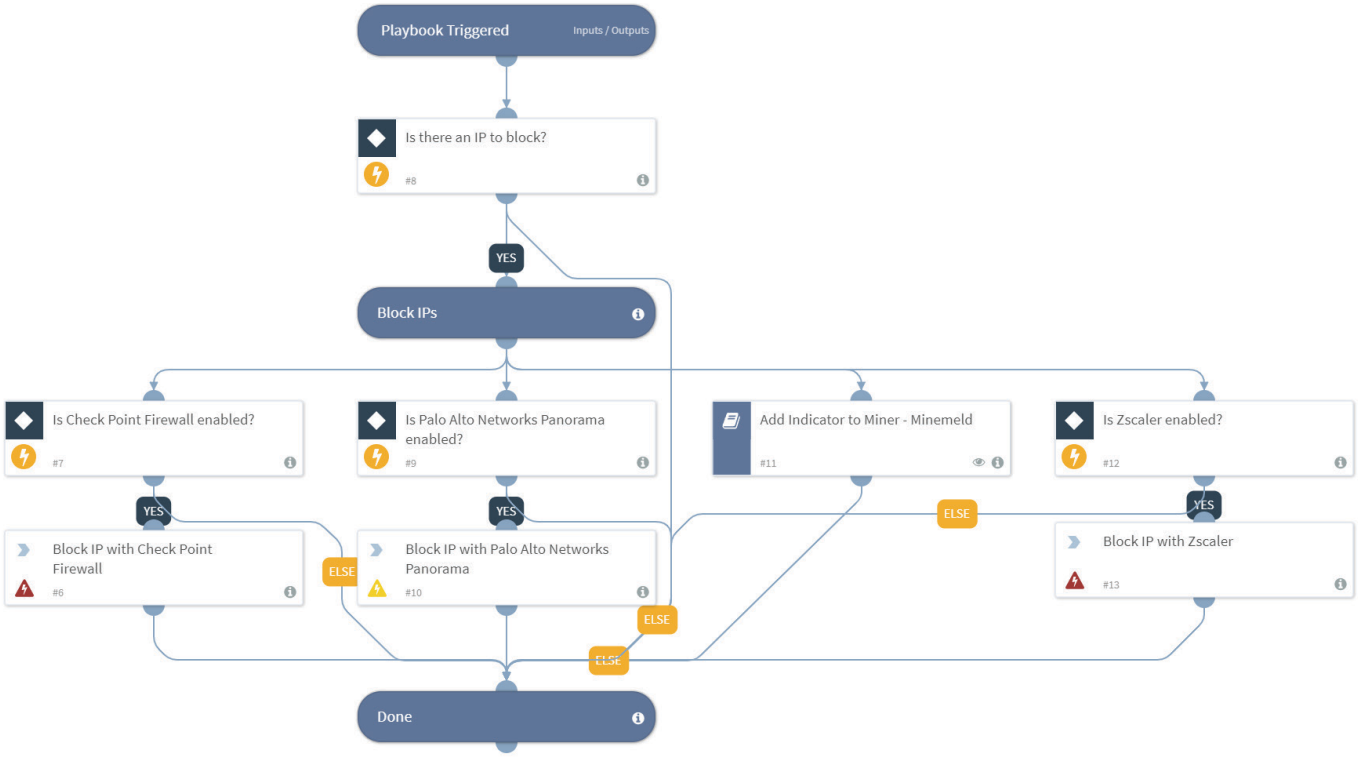
A single playbook can include support for multiple integrations. For data exfiltration, you can block a session that is uploading data to an external site or IP address. Figure 33 shows the blocking of an IP address, using one of the built-in playbooks called the *Block IP Playbook*, which includes support for multiple third-party firewalls that are integrated with the platform. In this example, the Cortex XSOAR platform has added a few of the integrations for the following firewalls:

- Palo Alto Networks firewalls
- Check Point firewalls
- Zscaler firewalls

Cortex XSOAR supports firewall integration from other vendors, and after integration into the Cortex XSOAR platform, these firewalls can also be part of the playbook. As previously mentioned, there are many built-in playbooks, and you also have the ability to easily and flexibly edit or create your own custom playbooks. Cortex XSOAR offers a visual playbook editor, which allows you to build your own custom playbooks by using a drag-and-drop interface. This makes it easier because you don't have to have scripting skills to create or modify playbooks.

Figure 33 shows a snapshot of a real playbook and how it is visually represented in the Cortex XSOAR GUI. You can click each task in the playbook to view specific details, including the actual commands and conditions for each action.

*Figure 33 Block IP Playbook*



You can run the playbook manually based on an analyst's incident investigation or automatically based on incident types received. The playbook runs through a sequence of tasks until the malicious IP address has been blocked across all the firewall platforms.

Often if data exfiltration is detected after it happens, then you could already be running against the clock in regard to regulations and compliance related to a breach.

## Regulations and Compliance

With multiple industry standards present today (such as NIST, CERT, and SANS) and regulatory requirements like the *General Data Protection Regulation* (GDPR) and *California Consumer Privacy Act* (CCPA), incident management platforms cannot afford to be tied to a single standard anymore. Apart from dedicated templates aligned to popular standards, SOCs need the flexibility to enable their platform to be able to customize templates for evolving standards as required.

GDPR is an EU regulation that protects the information and privacy of individuals. The GDPR introduces the requirement for a personal data breach to be reported to the competent national supervisory authority and in certain cases, to communicate the breach to the individuals whose personal data has been affected by the breach.

The California Consumer Privacy Act is a state statute intended to enhance privacy rights and consumer protection for residents of California, United States. CCPA essentially applies to any for-profit entity doing business in California that collects, shares, or sells the personal data of Californian consumers.

For an organization to keep up to date with regulations and compliance, you need to ensure that you don't require staff to create all regulation-related material from scratch and that you don't lock in to a standard, as regulations are subject to evolution. Service support, templates, and training material should be provided to get users informed and up to speed with using the applicable standards, and the platform should be flexible enough to morph as standards are updated, without sacrificing overall fidelity and product quality.

Cortex XSOAR is fully customizable to adapt to regulations and compliance, and by using playbooks, you can automate notification of data breaches per the GDPR example in the next section.

## Responding to a Data Breach

A *data breach* is a confirmed incident in which sensitive, confidential, or otherwise protected data has been accessed or disclosed in an unauthorized fashion. Data breaches can involve personal health information, personally identifiable information, trade secrets, or intellectual property. Data breaches occur through a data exfiltration process.

Many organizations treat customer notification of a data breach as a mandatory compliance obligation, and due to this approach, customers often feel as though the notification is a reluctant afterthought on the part of the organization. The unfortunate result is that the organization can lose an important opportunity to redress the data breach with their customers at a critical point when communication is key and customer trust is fragile.

It's imperative that your incident response plan includes a process to notify customers of a data breach. If it does not, it's missing an important part of the communication that informs customers of a breach concerning their personal data and enables them to take appropriate action to protect and prevent damage from the breach.

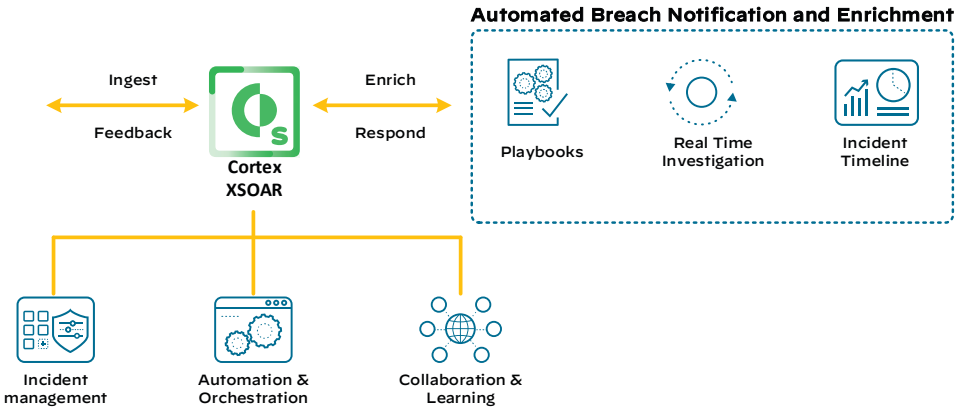
Cortex XSOAR comes with a GDPR playbook, which you can adjust to meet your organization's needs. Within Cortex XSOAR, there is an incident type called *GDPR Data Breach*, which is linked to the GDPR incident fields and the GDPR Breach Notification playbook. Within the Cortex XSOAR GUI, you can manually create an incident that has the type GDRP Data Breach.

After you create a GDPR Data Breach incident, the playbook is triggered. The playbook executes the following steps. Some steps are sequential, and others run concurrently and independently:

- An email that includes an investigation form is sent to the Data Protection Officer (DPO) to complete. The form includes questions about the nature and scope of the breach. A DPO is the person in your organization that is responsible for notifying the Data Protection Authority if personal data has been breached.
- The playbook proceeds according to the answers to the investigation form. If necessary, a report is generated, and the DPO reviews the report and either approves or declines sending the report to the Data Protection Authority and affected data subjects.
- If the DPO approves sending the report to the supervisory authority, the *GDPRContactAuthorities* playbook runs. The script receives the country in which the organization is located and returns the relevant contact information for the local supervisory authority. Based on the script results, an email that includes the data breach report is sent to the local supervisory authority.
- If the DPO declines sending the report to the supervisory authority, a manual task in the playbook requires the DPO to edit the report and approve sending the edited report to the supervisory authority.
- Depending on the DPO's answers to the investigation form, the playbook concurrently sends the data breach information to the affected data subjects. This information includes whom to contact for additional information.
- All data breach information is documented in the incident summary in Cortex XSOAR. At any point, you can create an incident summary report. Finally, the investigation is then closed.

Beyond the Palo Alto Network integrations, Cortex XSOAR also has integrations with third-party platforms (for example, Digital Shadows) that focus on detection of breached records.

*Figure 34 Digital Shadows*




With Cortex XSOAR and the Digital Shadows playbook, you can find all the breached records for a company domain, find unique breach identifiers, and get specific details for a breach you are investigating. With the integration, you can ingest Digital Shadows breach data into Cortex XSOAR, create incidents, and trigger playbooks tied to those incidents, such as blocking the exfiltration activities by isolating hosts and systems. You can also automate enrichment of data connected to the breaches by running playbook tasks that gather details such as data records, domains, usernames, IOCs, and network ports used.

Regardless of the threat type, with Cortex XSOAR, you can always do real-time investigations in the previously mentioned War Room to do the following tasks:

- Run real-time security actions through the CLI without switching consoles. The CLI is built into Cortex XSOAR.
- Run existing and new security playbooks, scripts, and commands.
- Collaborate with others for joint investigations and execute remote actions across many integrated products.

With Cortex XSOAR, you are able to standardize and automate processes for any security use case, and you can use playbooks in order to easily orchestrate and automate response actions across more than 350 third-party products.

Security-focused case management allows the SOC to efficiently adapt to any alert, and the unification of alerts, incidents, and indicators within a single framework expedites incident response. Using Cortex XSOAR, you can take control of the threat intelligent by aggregating data from disparate sources, customizing and scoring feeds, and matching indicators against the environment. You can then leverage playbooks in order to quickly and confidently drive instant action.

# Summary

This guide has discussed the Security Operation Center and the challenges they face with the high volume of alerts received, high false positives, skills shortage, and the lack of automation for tasks and procedures, requiring time-consuming, manual workflows. The Palo Alto Networks three-pronged approach for the SOC includes the following:

- **Prevention first**—Leveraging built-in prevention capabilities in Palo Alto Networks next-generation firewalls and the Cortex XDR agent, you prevent all the attacks you can and reduce the number of alerts going to the SOC. This is the starting point, and it is paramount to get it right first.
- **Detection and investigation**—With Cortex XDR, the SOC gains better visibility and insight into detecting stealthy attacks from data triaged from multiple sources, including the network, endpoints, and cloud. Cortex XDR provides threat-hunting capabilities and graphical causality views that help you better understand the sequence of events that occurred during an attack campaign.
- **Orchestration and automation**—Cortex XSOAR provides automated responses through playbooks and integration with hundreds of security tools and platforms, speeding up response times. Cortex XSOAR provides workflow automation, incident management, and collaboration capabilities.

Palo Alto Networks provides an integrated solution for SecOps and creates efficiency with a comprehensive product suite for prevention, detection, response, and automation. This approach gives you complete context for multiple use cases for known and unknown threats, speeding up response times and enabling the SOC to handle more alerts, faster and more efficiently.

Palo Alto Networks offers a comprehensive product suite for security operations that reduces false positives, detects stealthy attacks, automates responses, and is more efficient and proactive. The overall business benefits are reducing security risks and data breaches, reducing costs, and getting the most out of your existing security investments.

# Additional Resources

This section describes useful resources that provide additional information beyond the scope of this guide.

## BEST PRACTICES

It's important to understand other available resources, for example, fully hardening the NGFW with a day-one configuration or deploying specific features through templates called *skillets*. Besides configuration for security hardening, the NGFW provides visibility features that provide insight into what is happening on the network, such as telemetry and Application Command Center (ACC).

### Skillets

Palo Alto Networks offers best-practice configuration templates for hardening of your NGFW. For example, IronSkillet is a set of day-one configuration templates for PAN-OS that enable alignment with security best practices. The purpose of the IronSkillet project is to provide day-one best-practice configuration templates that you can load into a Palo Alto Networks next-generation firewall or Panorama management platform.

### Telemetry

*Telemetry* is the process of collecting and transmitting data for analysis. When you enable telemetry on the firewall, the firewall collects and forwards data to Palo Alto Networks, including information on applications, threats, device health, and passive DNS. The data that each telemetry participant shares benefits all Palo Alto Networks users, enabling a community-driven approach to threat prevention.

Telemetry is an opt-in feature, and, for most telemetry data, you can preview the information that the firewall collects. Palo Alto Networks does not share your telemetry data with other customers or third-party organizations.

### ACC

The ACC included with the GUI of the firewall or with Panorama is an analytical tool that provides actionable intelligence on activity within your network. The ACC uses the firewall logs in order to graphically depict traffic trends on your network. The graphical representation allows you to interact with the data and visualize the relationships between events on the network, including network usage patterns, traffic patterns, and suspicious activity and anomalies. The ACC provides graphical views on network activity, threat activity, and blocked activity.

# THREAT-INTELLIGENCE RESOURCES

There are multiple threat-intelligence tools available for the SOC, including some from Palo Alto Networks, such as WildFire and AutoFocus, as well as multiple third-party organizations, such as SANS and the Department of Homeland Security. This section covers Palo Alto Networks Unit 42 and VirusTotal.

## Unit 42

One great threat intelligence resource is Palo Alto Networks Unit 42. *Unit 42* is a threat-intelligence team at Palo Alto Networks that is a recognized authority on cyber threats and provides free threat intelligence to the community. Unit 42 is backed by the Palo Alto Networks Engineering and Critical Response teams, offering years of experience detecting and preventing attacks.

Unit 42 follows a traditional intelligence cycle that starts with direction from their leadership regarding critical intelligence requirements. These requirements help the analysts determine what data is necessary to answer specific questions about threats to Palo Alto Networks and customers. Unit 42 collects that data from internal and external sources and runs it through a detailed threat analysis process. The process uses automated systems that correlate incoming data and uses expert human analysis to interpret the data, identify patterns, develop hypotheses, and evaluate those hypotheses against the data set. By doing this, Unit 42 can put threats into context and help others determine how to best defend against future attacks.

Unit 42 provides in-depth research on adversaries, malware families, and attack campaigns. The team documents and shares adversary behaviors as playbooks. Multiple playbooks and a regular podcast are available on their website.

Unit 42 is a great research resource to gain insight into attack types with real examples of known adversaries, the various commodity malware tools (such as AgentTesla, KeyBase, LokiBot, Pony, Zeus), and the various attack methods used for each of their campaigns.

For more information about Unit 42, see their web page.

## VirusTotal

VirusTotal is a free service used to analyze files and URLs for malicious content such as worms, viruses, and trojans. Their goal is to make the internet a safer place by enabling collaboration between researchers, members of the antivirus industry, and end users. Many antivirus engines, website scanners, file and URL analysis products, and users directly contribute to VirusTotal's aggregated data, and the file and URL characterization tools can be used for a wide range of purposes, including heuristic engines, known-bad signatures, metadata extraction, and identification of malware indicators. Many prominent security companies, governments, and Fortune 500 companies are members of the VirusTotal community, which has more than 500,000 registered users.

You can use the feedback form to send comments about this guide.

## HEADQUARTERS

Palo Alto Networks
3000 Tannery Way
Santa Clara, CA 95054, USA
http://www.paloaltonetworks.com

Phone: +1 (408) 753-4000
Sales: +1 (866) 320-4788
Fax: +1 (408) 753-4001
info@paloaltonetworks.com



B-000225P-20a