Network Discovery Overview

The following topics discuss network discovery:

About Detection of Host, Application, and User Data

The network discovery policy applies only to threat defense devices that are managed by the on-prem management center for events and analytics.

The system uses *network discovery* and *identity* policies to collect host, application, and user data for traffic on your network. You can use certain types of discovery and identity data to build a comprehensive map of your network assets, perform forensic analysis, behavioral profiling, access control, and mitigate and respond to the vulnerabilities and exploits to which your organization is susceptible.

Host and Application Data

Host and application data is collected by host identity sources and application detectors according to the settings in your network discovery policy. Managed devices observe traffic on the network segments you specify.

For more information, see Host and Application Detection Fundamentals, on page 2.

User Data

User data is collected by user identity sources according to the settings in your network discovery and identity policies. You can use the data for user awareness and user control.

For more information, see About User Identity.

Logging discovery and identity data allows you to take advantage of many features in the system, including:

- Viewing the network map, which is a detailed representation of your network assets and topology that you can view by grouping hosts and network devices, host attributes, application protocols, or vulnerabilities.
- Performing application and user control; that is, writing access control rules using application, realm, user, user group, and ISE attribute conditions.
- Viewing host profiles, which are complete views of all the information available for your detected hosts.

- Viewing dashboards, which (among other capabilities) can provide you with an at-a-glance view of your network assets and user activity.
- Viewing detailed information on the discovery events and user activity logged by the system.
- Associating hosts and any servers or clients they are running with the exploits to which they are susceptible.

  This enables you to identify and mitigate vulnerabilities, evaluate the impact that intrusion events have on your network, and tune intrusion rule states so that they provide maximum protection for your network assets
- Alerting you by email, SNMP trap, or syslog when the system generates either an intrusion event with a specific impact flag, or a specific type of discovery event
- Monitoring your organization's compliance with an allow list of allowed operating systems, clients, application protocols, and protocols
- Creating correlation policies with rules that trigger and generate correlation events when the system generates discovery events or detects user activity
- Logging and using NetFlow connections, if applicable.

Host and Application Detection Fundamentals

You can configure your network discovery policy to perform host and application detection.

For more information, see Overview: Host Data Collection and Overview: Application Detection.

Passive Detection of Operating System and Host Data

Passive detection is the system's default method of populating the network map by analyzing network traffic (and any exported NetFlow data). Passive detection provides contextual information about your network assets, such as operating systems and running applications.

If traffic from a monitored host does not offer conclusive evidence of the host's operating system, the network map displays the most likely operating system. For example, a NAT device may appear to be running several operating systems because of the hosts "behind" the NAT device. To make this most-likely determination, the system uses a confidence value it assigns to each detected operating system, and the amount of corroborating data among detected operating systems.

Note The system does not consider reported "`unknown`" applications and operating systems in its determination.

If passive detection inaccurately identifies your network assets, consider the placement of your managed devices. You can also augment the system's passive detection capabilities with custom operating-system fingerprints and custom application detectors. Or, you can use *active detection*, which is not based on traffic analysis, but instead allows you to directly update the network map using scan results or other information sources.

Active Detection of Operating System and Host Data

Active detection adds host information collected by active sources to network maps. For example, you can use the Nmap scanner to actively scan the hosts that you target on your network. Nmap discovers operating systems and applications on hosts.

In addition, the host input feature allows you to actively add *host input data* to network maps. There are two different categories of host input data:

- *user input data*—Data added through the system user interface. You can modify a host's operating system or application identity through this interface.
- *host import input data*—Data imported using a command line utility.

The system retains one identity for each active source. When you run an Nmap scan instance, for example, the results of the previous scan are replaced with the new scan results. However, if you run an Nmap scan and then replace those results with data from a client whose results are imported through the command line, the system retains both the identities from the Nmap results and the identities from the import client. The system then uses the priorities set in the network discovery policy to determine which active identity to use as the current identity.

Note that user input is considered one source, even if it comes from different users. As an example, if UserA sets the operating system through the host profile, and then UserB changes that definition through the host profile, the definition set by UserB is retained, and the definition set by UserA is discarded. In addition, note that user input overrides all other active sources and is used as the current identity if it exists.

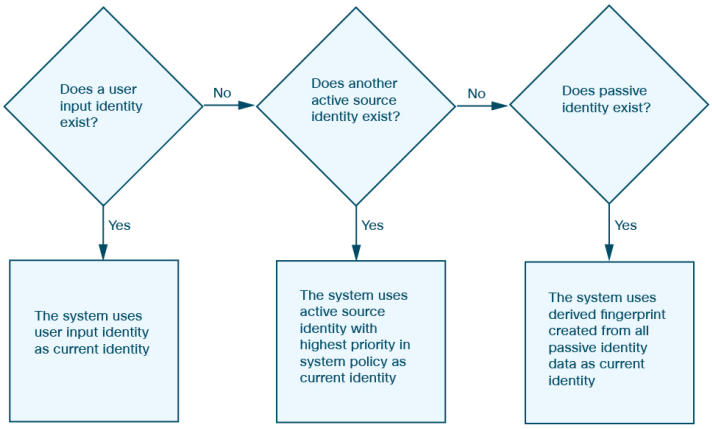
Current Identities for Applications and Operating Systems

The *current identity* for an application or an operating system on a host is the identity that the system finds most likely to be correct.

The system uses the current identity for an operating system or application for the following purposes:

- to assign vulnerabilities to a host
- for impact assessment
- when evaluating correlation rules written against operating system identifications, host profile qualifications, and compliance allow lists
- for display in the Hosts and Servers table views in workflows
- for display in the host profile
- to calculate the operating system and application statistics on the Discovery Statistics page

The system uses source priorities to determine which active identity should be used as the current identity for an application or operating system.



For example, if a user sets the operating system to Windows 2003 Server on a host, Windows 2003 Server is the current identity. Attacks which target Windows 2003 Server vulnerabilities on that host are given a higher impact, and the vulnerabilities listed for that host in the host profile include Windows 2003 Server vulnerabilities.

The database may retain information from several sources for the operating system or for a particular application on a host.

The system treats an operating system or application identity as the current identity when the source for the data has the highest source priority. Possible sources have the following priority order:

1. user

2. scanner and application (set in the network discovery policy)

3. managed devices

4. NetFlow records

A new higher priority application identity will not override a current application identity if it has less detail than the current identity.

In addition, when an identity conflict occurs, the resolution of the conflict depends on settings in the network discovery policy or on your manual resolution.

Current User Identities

When the system detects multiple logins to the same host by different users, the system assumes that only one user is logged into any given host at a time, and that the current user of a host is the last authoritative user login. If only non-authoritative user logins have been logged into the host, the last non-authoritative user login is considered the current user. If multiple users are logged in through remote sessions, the last user reported by the server is the user reported to the management center.

When the system detects multiple logins to the same host by the same user, the system records the first time that a user logs into a specific host and disregards subsequent logins. If an individual user is the only person who logs into a specific host, the only login that the system records is the original login.

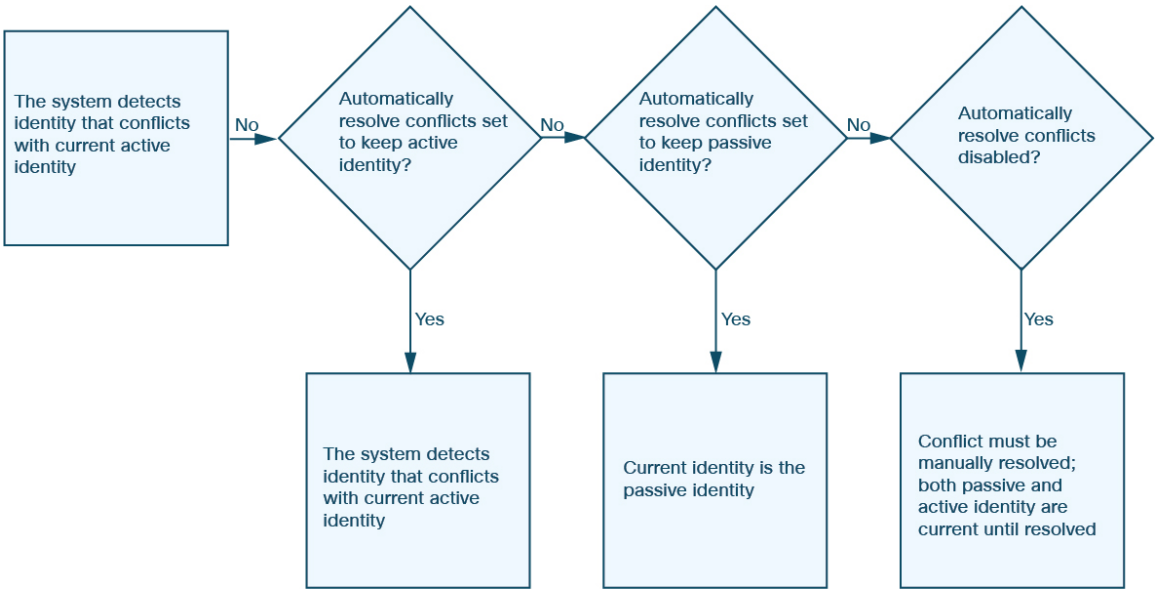
If another user logs into that host, however, the system records the new login. Then, if the original user logs in again, his or her new login is recorded.

Application and Operating System Identity Conflicts

An *identity conflict* occurs when the system reports a new passive identity that conflicts with the current active identity and previously reported passive identities. For example, the previous passive identity for an operating system is reported as Windows 2000, then an active identity of Windows XP becomes current. Next, the system detects a new passive identity of Ubuntu Linux 8.04.1. The Windows XP and the Ubuntu Linux identities are in conflict.

When an identity conflict exists for the identity of the host's operating system or one of the applications on the host, the system lists both conflicting identities as current and uses both for impact assessment until the conflict is resolved.

A user with Administrator privileges can resolve identity conflicts automatically by choosing to always use the passive identity or always use the active identity. Unless you disable automatic resolution of identity conflicts, identity conflicts are always automatically resolved.



A user with Administrator privileges can also configure the system to generate an event when an identity conflict occurs. That user can then set up a correlation policy with a correlation rule that uses an Nmap scan as a correlation response. When an event occurs, Nmap scans the host to obtain updated host operating system and application data.

Netflow Data

NetFlow is a Cisco IOS application that provides statistics on packets flowing through a router. It is available on Cisco networking devices and can also be embedded in Juniper, FreeBSD, and OpenBSD devices.

When NetFlow is enabled on a network device, a database on the device (the NetFlow cache) stores records of the flows that pass through the router. A flow, called a *connection* in the system, is a sequence of packets that represents a session between a source and destination host, using specific ports, protocol, and application protocol. The network device can be configured to export this NetFlow data. In this documentation, network devices configured in this way are called *NetFlow exporters*.

Managed devices can be configured to collect records from NetFlow exporters, generate unidirectional end-of-connection events based on the data in those records, and finally send those events to the management

center to be logged in the connection event database. You can also configure the network discovery policy to add host and application protocol information to the database based on the information in NetFlow connections.

You can use this discovery and connection data to supplement the data gathered directly by your managed devices. This is especially useful if you have NetFlow exporters monitoring networks that your managed devices cannot monitor.

Requirements for Using NetFlow Data

Before you configure the system to analyze NetFlow data, you must enable the NetFlow feature on the routers or other NetFlow-enabled network devices you plan to use, and configure the devices to broadcast NetFlow data to a destination network where the sensing interface of a managed device is connected.

The system can parse both NetFlow version 5 and NetFlow version 9 records. NetFlow exporters **must** use one of those versions if you want to export the data to the system. In addition, the system requires that specific fields be present in the exported NetFlow templates and records. If your NetFlow exporters are using version 9, which you can customize, you **must** make sure that the exported templates and records contain the following fields, in any order:

- IN_BYTES (1)
- IN_PKTS (2)
- PROTOCOL (4)
- TCP_FLAGS (6)
- L4_SRC_PORT (7)
- IPV4_SRC_ADDR (8)
- L4_DST_PORT (11)
- IPV4_DST_ADDR (12)
- LAST_SWITCHED (21)
- FIRST_SWITCHED (22)
- IPV6_SRC_ADDR (27)
- IPV6_DST_ADDR (28)

Because the system uses managed devices to analyze NetFlow data, your deployment must include at least one managed device that can monitor NetFlow exporters. At least one sensing interface on that managed device must be connected to a network where it can collect the exported NetFlow data. Because the sensing interfaces on managed devices do not usually have IP addresses, the system does not support the direct collection of NetFlow records.

Note that the Sampled NetFlow feature available on some network devices collects NetFlow statistics on only a subset of packets that pass through the devices. Although enabling this feature can improve CPU utilization on the network device, it may affect the NetFlow data you are collecting for analysis by the system.

Differences between NetFlow and Managed Device Data

The traffic represented by NetFlow data is not directly analyzed. Instead, it converts exported NetFlow records into connection logs and host and application protocol data.

As a result, there are several differences between converted NetFlow data and the discovery and connection data gathered directly by your managed devices. You should keep these differences in mind when performing analysis that requires:

- Statistics on the number of detected connections
- Operating system and other host-related information (including vulnerabilities)
- Application data, including client information, web application information, and vendor and version server information
- Knowing which host in a connection is the initiator and which is the responder

Network Discovery Policy versus Access Control Policy

You configure NetFlow data collection, including connection logging, using rules in the network discovery policy. Contrast this with connection logging for connections detected by managed devices, which you configure per access control rule.

Types of Connection Events

Because NetFlow data collection is linked to networks rather than access control rules, you do not have granular control over which NetFlow connections the system logs.

NetFlow data cannot generate Security Intelligence events.

NetFlow-based connection events can be stored in the connection event database only; you cannot send them to the system log or an SNMP trap server.

Number of Connection Events Generated Per Monitored Session

For connections detected directly by managed devices, you can configure the access control rule to log a bidirectional connection event at the beginning or end of a connection, or both.

In contrast, because exported NetFlow records contain unidirectional connection data, the system generates at least two connection events for each NetFlow record it processes. This also means that a summary's connection count is incremented by two for every connection based on NetFlow data, providing an inflated count of the number of connections that are actually occurring on your network.

Because the NetFlow exporter outputs records at a fixed interval even if a connection is still ongoing, long-running sessions can result in multiple exported records, each of which generates a connection event. For example, if the NetFlow exporter exports every five minutes, and a particular connection lasts twelve minutes, the system generates six connection events for that session:

- One pair of events for the first five minutes
- One pair for the second five minutes
- A final pair when the connection is terminated

Host and Operating System Data

Hosts added to the network map from NetFlow data do not have operating system, NetBIOS, or host type (host vs network device) information. You can, however, manually set a host’s operating system identity using the host input feature.

Application Data

For connections detected directly by managed devices, the system can identify application protocols, clients, and web applications by examining the packets in the connection.

When the system processes NetFlow records, the system uses a port correlation in `/etc/sf/services` to extrapolate application protocol identity. However, there is no vendor or version information for those application protocols, nor do connection logs contain information on client or web applications used in the session. You can, however, manually provide this information using the host input feature.

Note that a simple port correlation means that application protocols running on non-standard ports may be unidentified or misidentified. Additionally, if no correlation exists, the system marks the application protocol as `unknown` in connection logs.

Vulnerability Mappings

The system cannot map vulnerabilities to hosts monitored by NetFlow exporters, unless you use the host input feature to manually set either a host's operating system identity or an application protocol identity. Note that because there is no client information in NetFlow connections, you cannot associate client vulnerabilities with hosts created from NetFlow data.

Initiator and Responder Information in Connections

For connections detected directly by managed devices, the system can identify which host is the initiator, or source, and which is the responder, or destination. However, NetFlow data does not contain initiator or responder information.

When the System processes NetFlow records, it uses an algorithm to determine this information based on the ports each host is using, and whether those ports are well-known:

- If both or neither port being used is a well-known port, the system considers the host using the lower-number port to be the responder.
- If only one of the hosts is using a well-known port, the system considers that host to be the responder.

For this purpose, a well-known port is any port that is either numbered from 1 to 1023, or that contains application protocol information in `/etc/sf/services` on the managed device.

In addition, for connections detected directly by managed devices, the system records two byte counts in the corresponding connection event:

- The **Initiator Bytes** field records bytes sent.
- The **Responder Bytes** field records bytes received.

Connection events based on unidirectional NetFlow records contain only one byte count, which the system assigns to either **Initiator Bytes** or **Responder Bytes**, depending on the port-based algorithm. The system sets the other field to 0. Note that if you are viewing connection summaries (aggregated connection data) of NetFlow records, both fields may be populated.

NetFlow-only Connection Event Fields

A small number of fields are present only in connection events generated from NetFlow records.