Architecting User Identification (User-ID) Deployments

Strategies and Tactics guide

PANOS 5.0+

Table of Contents

SECTION 1: USER IDENTIFICATION SOFTWARE COMPONENTS 3

EXTERNAL SYSTEMS REFERENCED BY USER IDENTIFICATION 3
ACTIVE DIRECTORY 4
SINGLE ACTIVE DIRECTORY DOMAIN 5
MULTIPLE ACTIVE DIRECTORY DOMAINS, SINGLE FOREST 5
MULTIPLE ACTIVE DIRECTORY FOREST 6

SECTION 3: DESIGNING USER AGENT TOPOLOGY FOR USER TO IP MAPPING 7

THE USER-IDENTIFICATION AGENT 7
BASIC AGENT PLACEMENT THEORY – PROBING 8

SECTION 4: CAPTIVE PORTAL 9

CAPTIVE PORTAL MODES 9

SECTION 5: TERMINAL SERVICES AGENT 10

SECTION 6: GLOBAL PROTECT 11

SECTION 7: ADVANCED DESIGNS 12

TECHNIQUES 12
SCENARIOS 12
HARDWARE PAN-OS AGENTS 13
TARGETED / MULTIPLE SOFTWARE AGENTS 14
MICROSOFT LOG FORWARDING 15
DEDICATED HARDWARE AGENTS 16
SYSLOG LISTENER 17

REVISION HISTORY 18

Section 1: User Identification software components

There are two major jobs for the User Identification (User-ID) process on the Palo Alto Networks firewall.

1) User to Group mapping: This process learns group names then the users that are members of the groups. It allows the firewall to write policy and create reports based on groups rather then individual users.

2) User to IP mapping: This process associated a user name with a specific IP. A given user can be associated with many IP addresses, but a single IP address can only be associated with a single user. The only exception to this is the Terminal Server Agent, which is used for multi user Terminal Servers.

The Palo Alto Networks User Identification process consists of three separate components. The **Palo Alto Networks User-ID Agent**, the **Palo Alto Networks Terminal Server Agent** and the firewall running **PAN-OS**. Each component has specific functions and will require different topologies to function optimally.

- Palo Alto Networks User-ID Agent: This is a service that can be installed on any domain member system or run on any firewall device that runs version PAN-OS 5.0 and up. It is responsible for 5 mechanisms that map users to their corresponding IP addresses.
  - Server Monitoring – Active Directory Domain Controllers, Microsoft Exchange Servers and Novell eDirectory servers.
  - Session Monitoring – Windows servers.
  - Client Probing – Windows systems thru either the WMI or NetBIOS.
  - User ID API – Extensible interface to import user data for other external sources then the ones mentioned above. This can include user defined scripts as well as partner integrations (such as Aruba Clearpass and Splunk)
  - Syslog Listener (feature available in the 6.0 and higher version) – The agent runs a syslog listener on a designated port that is able to parse the syslog messages and convert the information in appropriate user-ip mappings

- Palo Alto Networks Terminal Server Agent: This is a Windows service designed to be installed on a Windows or Citrix terminal Server to map users to IP + Source Port tuple. This provides tight correlation of terminal server traffic to terminal server user. It is required on any terminal server supporting users.

- The firewall software performs three specific functions in the User Identification process.
  - User to IP Mapping via the hardware agent and also via Captive Portal – Both NTLM and Web Form methods are supported.
  - User to IP mapping using the Global Protect client in either internal or external modes.
  - User to Group mapping – Using LDAP the firewall will build a list of groups and the associated users for use in both policy and reporting.

External systems referenced by User Identification

Configuration and design of User Identification can be complicated due to the number of external systems that may be part of the network infrastructure. Common external systems that are referenced in this document include:

- Windows Active Directory Domain Controllers
- Windows Exchange Server Client Access Servers
- Novell eDirectory Servers
- Other more generic LDAP implementations (OpenLDAP being the most common)

Section 2: Designing User to Group Mapping

Active Directory

By far the most common directory in current deployment is Microsoft's Active Directory. This is a LDAP based directory that has an extensive interface layered over it. This has the effect of masking much of the underlying LDAP structure from the casual administrator. The following is a list of common terms that play a part in the design of user to group mapping within Active Directory (AD)

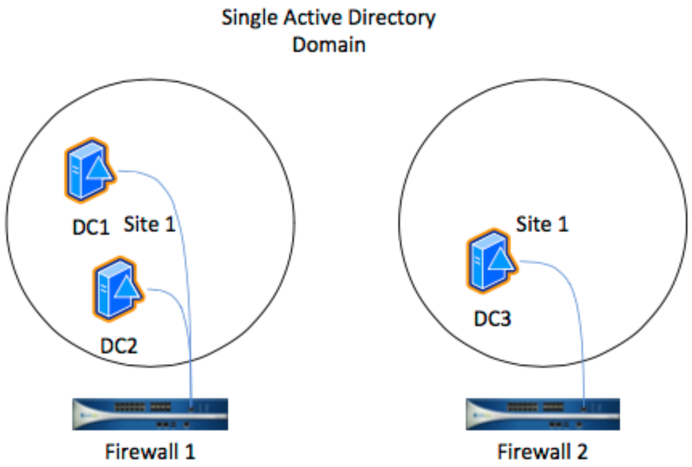
- Active Directory Forest: This is the term applied to a full LDAP tree. A forest is a set of domains that share a common LDAP schema and have implied trust relationships running throughout. It is most common to have a single forest in a single corporate environment. Mergers and acquisitions may create scenarios where there are multiple forests in a single customer environment. Domains in a forest may have different names but they share common configuration and schema. The forest derives its name from the first domain created in the forest. For example if the first domain created was corp.com the forest would be referred to as the corp.com forest.
- Active Directory Tree: This is a Microsoft term for a set of domains in a single forest that share a common naming space. For example asia.corp.com and corp.com would be in the same tree. Domains in the same tree are by definition in the same forest. This term is immaterial to User ID design.
- Active Directory Domain: This is the smallest unit of LDAP replication in the Microsoft Active Directory. All AD deployments must consist of at least one Domain. A domain is represented by a single common name within the forest. For example asia.corp.com could be the name of a specific domain in the corp.com forest.
- Active Directory Domain Controller: A Domain Controller (DC) is a server that contains part of the Active Directory LDAP database. All DC's in a single domain have identical databases. This database contains all attributes for all of the objects in the Domain. By default each of the DC's is able to make changes to this database and then replicate those changes to other DC's in their domain. Standard DC's contain no information regarding objects in other domains from the forest.
- Global Catalogue Server: The Global Catalogue Server (GC) is standard Domain Controller for one of the domains in the forest that carries an additional LDAP database replica. This additional replica contains pointers for all of the objects in all of the domains in the forest, but does not contain all of their attributes. This GC LDAP database runs on a different port then the server traditional domain database. Specifically this database contains useful information concerning Universal Group membership.
- NetBIOS Name: While Active Directory has supported DNS style names since 2003 it is still common to see users and domains using the 15 max character NetBIOS name. A user named John Smith in the corp.com domain could be identified with a name such as **jsmith@asia.corp.com** or they could be referred to as asia\jsmith. In this case the domain asia.corp.com is referred to using its NetBIOS name. It is important to note that a domains NetBIOS name is not always the left most portion of its fully qualified name. For example a domain named johnsonbrothers.com may have a NetBIOS name of johnson.

Single Active Directory Domain

In a single AD domain environment user to group enumeration is simple. The following objects must be defined on the firewall:

1) LDAP authentication server
2) User Identification Group Mapping Settings

In a single domain, all Domain Controllers will have the identical information. As a result the firewall should be configured to connect to the nearest or best connected domain controller to gather the user and group data. Additional Domain Controllers can be added to provide fault tolerance but should be added based on their proximity to the firewall.
Also a good consideration is to have a connection to the DR (Disaster Recovery) site Domain Controller (if any), to be able to get the information in case of a complete side down.

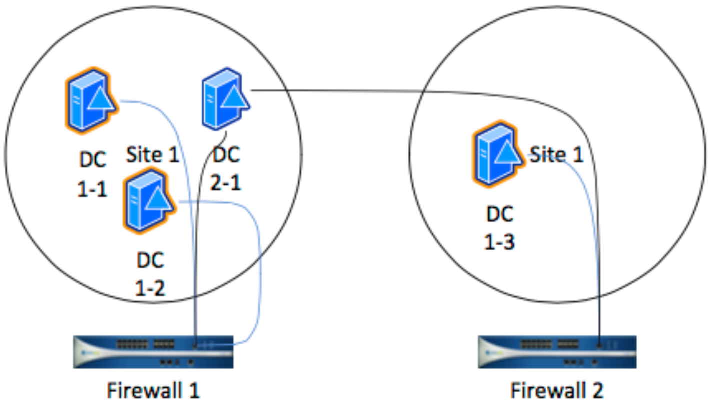


Multiple Active Directory Domains, Single Forest

In the case of a multiple AD Domain environment the firewall will need a separate LDAP server configuration for each domain. If Universal Groups are to play a significant role in firewall policy then an additional connection to a Global Catalogue server should be required as well. As in section 2.2 the firewall should connect to local instances of the domain controllers when possible. If no local instances are available for one of the domains then the firewall can be configured to connect to a remote domain controller with the update interval lengthened to reduce traffic across the WAN.

Required objects are:

1) LDAP authentication server for each domain and optionally for the Global Catalogue.
2) User Identification Group Mapping setting for each LDAP server.



Multiple Active Directory Forest

From the point of view of the firewall and group enumeration, a multiple forest environment is identical to the multiple domain environment. The additional concern with multiple forests is that there is no guarantee of unique names with respect to the other forest. Care will need to be taken that two identically named groups / users from two identically named domains are not encountered.

Section 3: Designing User Agent Topology for User to IP Mapping

The User-Identification Agent

To successfully deploy an identity aware security solution using Palo Alto Networks firewalls, users must be mapped to IP addresses in real time. In smaller and simpler networks, the deployment of a software or hardware user agent in a central location will suffice to map the majority of users. When planning agent placement it is important to consider the traffic generated by the agent. For the software agent the traffic generated is as follows:

- Log monitoring traffic: This traffic occurs between the agent and an Active Directory Domain Controller or Exchange CAS. The content of this traffic is the entirety of the servers Security Logs. Based on the "Security Log Monitor Frequency" value configured on the agent this traffic consists of an authenticated TCP based WMI connection to the server that fetches the new logs since the last check.
- Open server session traffic: This traffic occurs between the agent and an Active Directory Domain Controller or Exchange CAS. The content of this traffic is the entirety of the servers' current session table for file and print shares. Based on the "Server Session Read Frequency" value configured on the agent this traffic consists of an authenticated TCP based WMI connection to the server that fetches the current session table.
- Probing (WMI and NetBIOS) traffic: If probing is enabled the agent will make a connection using either the WMI or NetBIOS to each known IP address discovered through log or session monitoring. This connection will be used to verify that the last known user is still the current user. The list of known IP addresses is cycled through once each period defined by the WMI/NetBIOS Probing Interval as configured on the agent. The same connection will be used if the user is not known to the agent, as an active method to check the IP address the firewall requested from him.
- Firewall update traffic: when a user mapping is learned or updated the agent pushes this data to the firewall. In addition the agent refreshes all known mappings to the firewall every hour. This traffic consists of just the username and corresponding IP address as well as a time stamp.

The traffic between the Domain Controller and the Agent, or the Exchange Server and the Agent cannot be optimized. The amount of traffic is determined by the amount of logging activity on the target server. The only optimization is the frequency of calls sends to the servers, but the Agent will ask again for all the events since his last timestamp, so the amount of traffic will be exactly the same regardless of the frequency. As a result it is a best practice to place the agent closer to the monitored servers when bandwidth is a concern. The traffic between the Agent and the Firewall consists of the bare minimum data required for user to IP mapping and is better suited to traverse more impacted links.

Basic agent placement theory – Log Reading

It is usually a best practice to place Agent systems near by the Domain controllers they will monitor. "Near" being a relative term with respect to networking. In the simplest design, an agent is placed in each physical site that is separated by a WAN link and that contains DC's or Exchange servers. Since users could theoretically authenticate to any DC in the environment and since the security logs are not replicated between Domain Controllers, all DC's in the enterprise must be monitored. Domain Controller / Exchange log monitoring is the lowest overhead option the agent provides and should always be used as the base (primary) method for user to IP mapping. Each firewall in the deployment should receive updates from every agent. By placing agents across impacted WAN links we minimize the User Identification traffic over the links. The remote agent can query local DC's and then send the summarized data back to the firewall. This way not all the security events are going to be sent to the firewall and traverse the WAN, but just the needed (digested) mappings are going to be forwarded from the Agent to the Firewall using the WAN links.

Basic agent placement theory – Probing

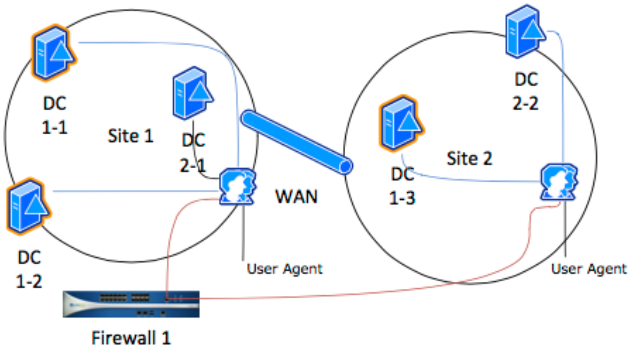
Agent probing can be used to verify the known user is still at an IP address that was learned from another method and for learning the user at an unknown IP address referenced by the firewall. Probing generates network traffic based on the total number of learned IP addresses of the network. Probing is most useful in networks with a high turnover of user to IP address. The probes can time out or update an IP mapping before the cache timer is hit. This can give a tighter correlation of current user to IP status.
WMI probing is always preferred to the legacy NetBIOS option. The hardware PAN-OS agent does not support NetBIOS, so if used, WMI probes will be send from the Firewall to the IP addresses that needs to be probed.

Advantages of WMI probing are:

1. WMI probes are authenticated. The agent will use its account to authenticate itself to each end point.
2. WMI is a more reliable method of gathering user data then NetBIOS
3. WMI is more likely to be allowed within the network then NetBIOS

Probing is best used on a well-connected network. It is not appropriate for WAN or congested network segments. In cases where the IP to user mappings are relatively static, probing is most likely unneeded, but it can be seen as useful in case the user is not seen in the DC logs, because the WMI probe will be sent to the unknown IP address directly "asking" for the user that is logged on at that moment.

The diagram below shows a simple topology using a network with 2 sites and 2 domains. Each site has an agent that is responsible for monitoring the Domain Controllers in that site. The single centralized firewall receives updates from both agents. The total number of agents in this design is 2.



Section 4: Captive Portal

Captive portal provides an active method to authenticate an unknown user. It will only be triggered when traffic from an unknown IP matches a captive portal policy. Since it is only invoked for this unknown traffic it provides a low cost option for identifying any IP not covered in the log or session monitoring. Most designs can benefit from the addition of captive portal to provide mapping for users that slip through the cracks of the other methods. Captive portal is only effective on web traffic. For firewalls deployed in a region of the network that does not encounter web traffic often, this method is less useful.

Captive Portal modes

Captive portal can function in 2 modes. Depending on the mode supported by clients the design consideration for captive portal are different.

- **Browser Challenge (NTLM Authentication):** If the client systems are running Windows operating system that are part of the domain, then NTLM authentication can be used. NTLM provides a transparent authentication method that will not impact the user experience. In environments where this is an option it is a best practice to enable it. For environments that are Microsoft and the browser usage can be controlled this method provides a seamless backup to the log monitoring. It will not work in environments where the users authenticate to local systems with a different account then the one used in the Active Directory.
- **Web Form Authentication:** In all environments other than described above, captive portal will use web form. Web form can also be enabled for all use cases if desired. An advantage to web form is that the user can be authenticated using any of the available authentication sources. A disadvantage is that this method interrupts the users work to prompt for credentials. This method is ideal for kiosk systems where the local user is not a user from the enterprise directory and where many different users may utilize the same system without performing any network log on. This method is also the desired method for mobile devices since they are not using NTLM.

In the Captive Policy configuration setup there is also a third action called **no-captive-portal**. This “mode” is used in case some traffic needs not to be user-identified.

PAN-OS 5.0 and later firewalls can be configured to redistribute this data to other firewalls in the environment and server as agents to those firewalls.

Section 5: Terminal Services Agent

For Windows and Citrix servers a special agent must be used. These systems multiplex users behind a single IP address. Traditional User ID methods cannot address this type of user traffic. The TS Agent is a small footprint agent that must be installed on each terminal server. The agent will control the source ports allocated to each user process and report this to the firewall. It is the only user ID component required for terminal servers.

From the point of view of design there is only one way to deploy the TS Agent. It must be installed on every Terminal Server and then added to each firewall that may encounter traffic from the terminal server and will require user data.

Section 6: Global Protect

The Global Protect client software plays a role in the Palo Alto Networks remote access, mobility and User Identification solutions. With respect to User ID Global Protect acts as a trusted client process that can report the user and IP of the endpoint to PAN-OS firewalls acting as gateways.

If an environment requires 100% correlation of user to IP at all times then the Global Protect client with Internal Gateways is the best solution. This does not need to be pushed out globally. It is possible that only a sub set of the client systems require this tight correlation. The rest of the user base may be serviced by log monitoring and Captive Portal.
The Global Protect Option of IDing users is also available for mobile devices (ios and android) and mac/linux/unix devices.

Section 7: Advanced Designs

When dealing with larger or more complex networks, additional techniques may be employed to cover user to IP mapping. The strategies covered here can be mixed and matched to address any customer environment. It is important to realize that in most customer networks, multiple strategies may need to be employed to reach an acceptable level of user to IP mapping. This is especially true for environments that are very dynamic and users change machines very often (example are the PoS-Points of Sales, in a big shop chains, where the shop assistants do not have dedicated PC's and very frequently log-on and log-off from the devices)

Techniques

The following techniques will be covered in this guide:
1) Central deployment of a hardware PAN-OS User agent for distributed environments
2) Deployment of software agents for targeted segments within the network
3) Use of Microsoft Log forwarding for highly distributed networks
4) Use of smaller (PA-200) firewalls as dedicated User ID appliances
5) Use of a syslog listener to read the log-on events from a proprietary application/system

Scenarios

The following broad descriptions represent the most common customer networks that require more sophisticated User Identification solutions:

1) **Highly distributed, low density Domain Controllers**. – Networks containing many Domain Controllers but having no clear hub sites. Sets of one or two DC's connected to other sites over a WAN. No logical location for centralized agent placement. Example: A regional financial institution with DC's at each branch.

2) **Multiple high latency / low bandwidth / heavily subscribed links**. – Networks containing sites with Domain Controllers that are poorly connected to central locations. Example: Oil platform with satellite link

3) **Large WiFi segments with high user turnover**. – Segments where users are ephemeral and the turnover is high. Login events are not frequent enough to map users as they move through the network, Example – Large campus WiFi

4) **Distributed sites with agent HA requirements** – Networks with a number of remote sites (10+) that require agent High Availability. This requires a large number of software agents to be configured.

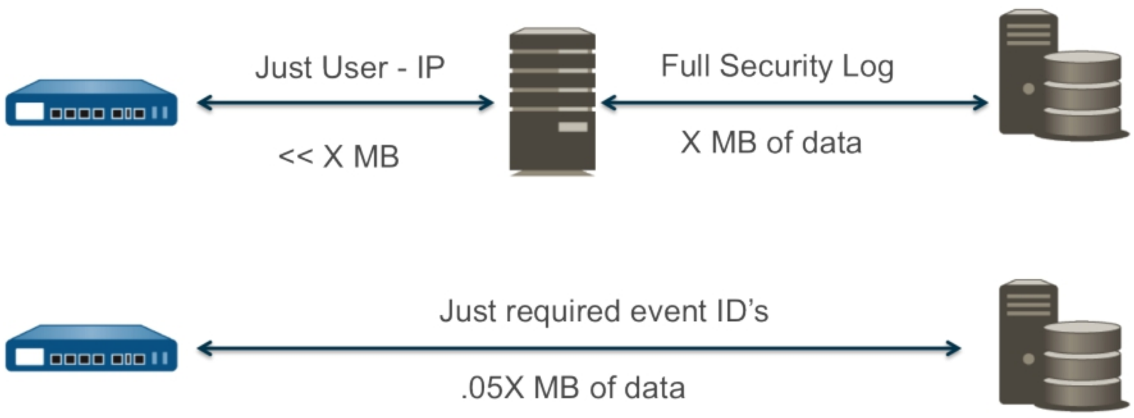
5) **User ID and multiple VSYS** – User agents are not shared between virtual systems. A network with 10 user agents and 5 VSYS would effectively have 50 agents configured. The hard limit of 100 agents per firewall limits the design options for large VSYS deployments.

6) **Non-domain or guest users that need to be mapped** – Segments of the network where there are users that connect to machines that are not part of the domain. These are usually guest or even domain users that are using phones, tablets or even BYOD which are not

generating logs to the DC’s, but can potentially generate authentication logs to another system (like RADIUS, WLC, or some other VPN solution….)

Hardware PAN-OS Agents

While both the stand-alone software User ID agent and the “agentless” PAN-OS processes perform the same basic tasks they use different underlying protocols. This difference makes each one more appropriate for different environments.



The software agent uses MS RPC to query the Domain Controller and Exchange Server logs. This method requires the full log to be transferred to the agent where it is then filtered for the required events. The hardware-integrated agent uses the WMIC library and only transfers the required log events to the agent process.

As a result the hardware agent is appropriate for reading remote Domain controllers where the software agent is appropriate for reading local Domain Controllers. The drawbacks to the hardware agent are the following:

1) Resources for the agent process come from the Management Plane. Significant User ID activity can impact other management plane features such as reporting, log querying and management.

2) There is no way to increase the resources for the hardware agent as the User ID environment grows.

3) The WMIC can have a higher impact on the target Domain Controllers CPU. This is mostly noticeable on 32 bit low RAM servers.

This technique can allow a significant reduction in the total number of agents required by allowing the agent process to sit in a central location rather then in multiple remote sites.

Performance of the hardware agent is determined by the firewall model performing the service.

The following table shows preferred numbers for each platform. If there are significant other requirements on the management plane the number of DC's should be reduced.

Platform	Supported DC's
PA500, PA2000, PA4000	10
PA200	25
PA3000, PA5000, PA7000	100

Use of the hardware agent is suggested for the following scenarios:
-Highly distributed, low density Domain Controllers.
-Multiple high latency / low bandwidth / heavily subscribed links.
-User ID and multiple VSYS.

Targeted / Multiple Software Agents

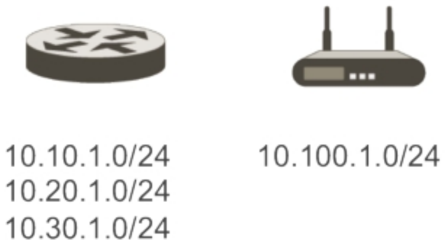
All settings on the dedicated User-ID Agent are global. There is no facility for different agent setting values based on the network range, or target Server. In some cases different network segments would benefit from different agent settings. To accomplish this multiple agents will need to be installed and their "Include / Exclude" network lists need to be configured so that they divide the customer environment into the needed segments.

There are a number of valid reasons to install more than one agent in a single site.

1. **Fault Tolerance**: Multiple agents monitoring the same domain controllers provide redundancy for the firewall, should one of the agents fail. The multiple agents do not need to be aware of each other (and they are not). The agents singularly send all known IP to user mappings to the firewall. The firewall is aware of the multiple feeds coming its way, so he will normalize them and remove any duplicate data, making use only of what he needs to get the correct mapping.
2. **Requirement of different agent settings for different network segments**: Agent global settings such as the enablement of probing or cache time out may not be ideal for all subnets at a given site. For example a common wireless network at a main site may have a very short DHCP lease time and hence a fast turnover of users to IP addresses. The rest of the network may be wired and have weeklong lease times. In this case a cache time out of 120 minutes for the wired network would be too long for the wireless. In addition the wireless network may benefit from WMI probing while it would be a significant traffic increase to probe the full, wired network, with very little benefit. Since both settings are global to the agent the only way to provide probing and a 15 minute time out to the wireless network while disabling probing and assigning a 120 minute time out to the wired network is to install two agents. Each agent would have their Include / Exclude list configured to only cover their portion of the sites IP address space. Each agent would be configured to monitor all of the same DC's and Exchange servers. This setup also

is recommended to be used with the fault tolerance setup, so that there is no single point of failure for that part of the network.

3. **Load Balancing**: In a large data center there could be a very high concentration of Domain Controllers or Exchange servers with very high levels of logging. The available monitored servers could be split up between multiple agents to minimize the amount of logs each agent was required to process. This is also very useful for environments where there are geographically distributed domain controllers that serve users in the dedicated network. Placing an User-ID agent in each location where user identification is needed, will lower the amount of data traversing the WAN links. Also here a combination with fault tolerance is a recommended option.



Agent 1	
Probing Settings	OFF
Include /Exclude List of configured Networks	10.10.1.0/24,10.20.1.0/24, 10.30.1.0/24

Agent 2	
Probing Settings	ON
Include /Exclude List of configured Networks	10.100.1.0/24

The agent values most likely to be targeted in this way are the WMI Probing Setting and the Cache time out.

The useage of targeted software agents is suggested for the following scenarios:

-Large WiFi segments with high user turnover.

Microsoft Log Forwarding

Microsoft servers support a publish/subscribe service for forwarding event logs from one server to another. We can use this service to the required security logs from remote Domain Controllers to a server more convenient for the software or hardware agent to monitor. This is a built in feature from Microsoft and does not require any additional software. This can be used to effectively reduce the total number of Domain Controllers that need to be monitored. This service will batch the logs over 30 seconds for efficient transport. It does introduce this additional latency into the user ID process. MS Log Forwarding is supported on all versions of Windows Server.

-Advantages of MS Log forwarding:
1) Reduces the number of servers that User Agents need to monitor.
2) Publish logs from remote or poorly connected sights.
3) Only the required logs need to be sent to the central server as opposite to all security logs being sent and read.
4) This configuration can be made standard on new DC's mitigating the need to keep the agent up to date on all existing DC's in production.
5) The agent only needs rights on the subscribing server, which does not need to be a Domain Controller.

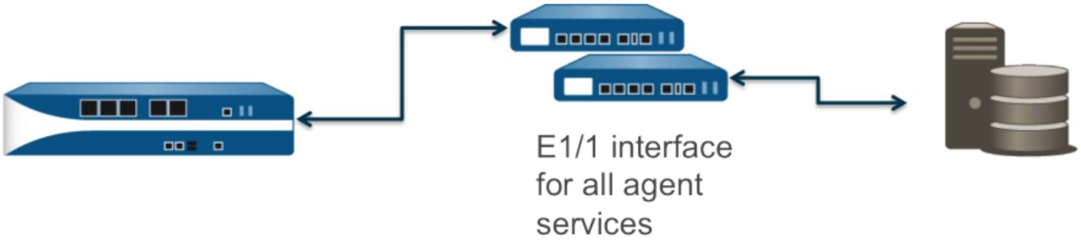
-Disadvantage of MS Log Forwarding
1) Can introduce up to an additional 30 seconds of latency into the process.
2) Server session reading is no longer useful as the agent is not connecting to the actual Domain Controller.
3) Reconfiguration needs to be done on the Server centrally collecting the logs, since by default all the logs are send to the ForwardedEvents.evtx, and the agents are only reading the Security.evtx.

The use of MS Log Forwarding is a possible recommendation for the following scenarios:
-Highly distributed, low density Domain Controllers.
-Multiple high latency / low bandwidth / heavily subscribed links.
-Distributed sites with agent HA requirements.
-User ID and multiple VSYS

Dedicated Hardware Agents

As an extension of the PAN-OS on board agent solution, dedicated platforms can be used to provide Agent services. Most commonly we deploy PA-200's in this manner but VM series systems could be used as well.

Dedicating hardware to this process has the following advantages:
1) Platforms can be deployed in HA pairs giving fault tolerance without adding additional agents to the configuration.
- This can drastically reduce the total number of agents configured on the traffic forwarding firewalls.



2) The management plane CPU and RAM are used fully for the User ID service allowing the platform to monitor the maximum number of DC's recommended.
3) Upgrades and configuration of the hardware agents can be done centrally though Panorama.
4) In some cases the removal of the Windows server requirement would be seen as beneficial.

The use of Dedicated Hardware Agents is recommended for the following scenarios:
-Distributed sites with agent HA requirements.
-User ID and multiple VSYS

Syslog Listener

Starting with PANOS 6.0, the Firewall has the capability to use the Software and the Hardware Agent as a syslog listener on which they can collect the syslog messages from different network elements and map the users to ip-addreses, which we can use in security rules and policies.
In this case the Agents serve as syslog servers, where a listener is brought up on a dedicated port, to collect the log messages that can be parsed and the ip-user-mappings can be collected. The communication can be in clear text and can be encrypted if needed as well.

The use of Syslog Listener is recommended for the following scenarios:
- Non-domain or guest users that need to be mapped

Revision History

Date	Revision	Comment
9/11/13	.9	Draft
9/13/13	.91	Redone with more coverage
10/1/13	1	First Published Edition
6/30/15	2	Redone with more coverage