Technical Guide

Understanding the Next Generation Firewall and its Architecture

Introduction

Allied Telesis Next-Generation Firewalls (NGFWs) are the ideal choice for high-speed Enterprise gateway applications. A typical solution could involve providing secure Internet access for several hundred hosts.

The NGFWs run the advanced AlliedWare Plus fully featured OS, and the CLI can be accessed locally, or remotely via telnet or SSH. The NGFW GUI allows graphical setup and monitoring, and the firewall can also be part of an AMF network, to ensure backup and recovery is automated to maximize the availability of online services.

This document discusses the architecture and performance of the NGFW, to help the reader understand the balance between security and performance when using multiple traffic control and threat protection features together.

Related documents

You may also find the following AlliedWare Plus documents useful:

- UTM Firewall Overview
- Getting Started with the Device GUI on UTM Firewalls
- Advanced Network Protection Feature Overview and Configuration Guide

Which products and software version does it apply to?

This Technical Guide applies to the following Allied Telesis NGFWs:

- AT-AR3050S
- AT-AR4050S

It requires AlliedWare Plus software version **5.4.5** or later.

Contents

Related documents....1
Which products and software version does it apply to? 1
Firewall architecture 3
Packet flow....3
Balancing network security with performance 4
Proxy versus stream....5
Proxy-based threat scanning....6
Stream-based threat scanning 7
Performance considerations....7
Options available on the Allied Telesis NGFW....8
URL Filtering versus web control 9
Performance optimizing architectures....11
Software architecture - multiple parallel processing paths....12
Hardware architecture 13
Platform selection....14
Business requirements 14
Network considerations....14
Hardware and software tables 15
Frequently asked questions 16

Firewall architecture

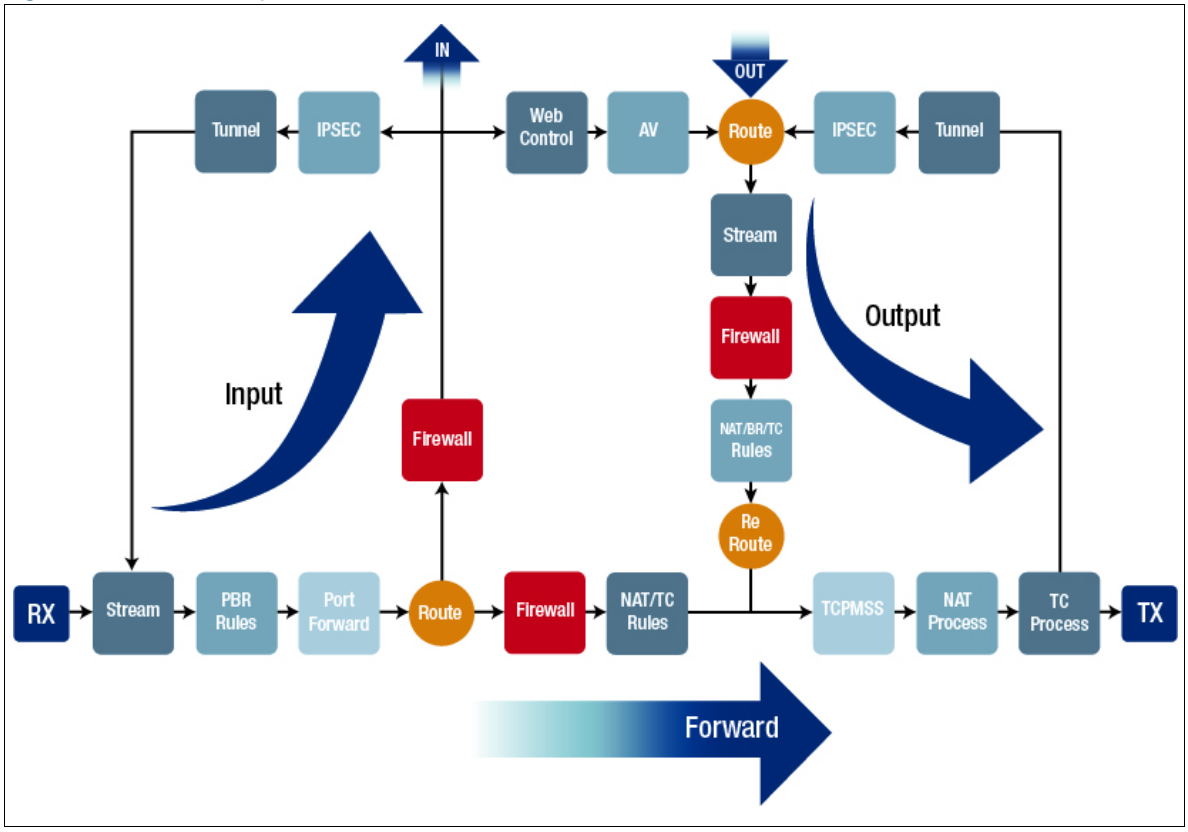
The Allied Telesis Next Generation Firewall is built and configured around an application and protocol decoding engine that performs Deep Packet Inspection (DPI). Firewall and NAT rules are defined to allow or deny IPv4 and/or IPv6 application traffic between network entities, such as individual hosts, servers, subnets, and networks.

Typically, a network administrator might use an NGFW to provide security zones based on business functions such as admin, sales, IT, and R&D staff etc. Alternatively, they might use an NGFW to implement security based on the traditional three zone approach - public zone, private zone and de-militarized zone (DMZ). Typical configuration could involve many network entity definitions (often involving several networks per zone), including several hundred rules to control access between hosts, networks, zones, and the Internet.

Packet flow

When protecting a private LAN from a public WAN, packets are most commonly forwarded by the firewall. Network packets pass through the firewall using a fixed set of processing steps as illustrated in the diagram below:

Figure 1: Packet flow process



Before you configure different firewall features, it is important to understand the packet processing order.

This is the packet processing order:

1. Packets are received on one of the firewall's physical interfaces (LAN or WAN) and follow the **Forward** path shown in Figure 1 on page 3.
2. Stream processing always happens straight after the packet is received on the physical interface. Stream processing is DPI, IPS, IP Reputation, Malware Protection, and URL Filtering.
3. Policy Based Routing and Port Forwarding is applied before packets are routed to their destination.
4. After routing, firewall and NAT rules are matched.
5. Finally, NAT processing and traffic control is applied before packet transmission.

The other two important processing packet paths in the firewall are **Input** and **Output**.

- The **Input** path is applied to packets that are destined to the firewall itself:

  Along with management and network protocol packets received by the firewall, this also includes tunnel packets for which the firewall is the tunnel decapsulation endpoint and proxied packets where the firewall terminates both sides of HTTP(S) connections (web-control and antivirus).

- The **Output** path is applied to packets generated by the firewall itself:

  Along with management and network protocol packets generated by the firewall, this also includes tunnel packets encapsulated by the firewall and proxied packets where the firewall terminates both sides of HTTP(S) connections.

  Note that because packets generated by the firewall were never received on a physical interface, stream processing is also applied first in the **Output** path.

Be aware that when tunnel packets are encapsulated or decapsulated they skip stream processing the second time around unless **tunnel security-reprocessing** is enabled in a configuration.Pacekt file.

Balancing network security with performance

Network administrators often have different concerns when it comes to network security coverage. Some require maximum boundary protection for their business network, while others need a good level of protection whilst also having minimal latency and maximum throughput.

An administrator needs to find the security/performance balance that best suits the requirements of their network.

Proxy versus stream

The elements that compose a threat management system fall into two general classes: proxy-based processes and stream-based processes.

Both types of processes focus on delivering secure and robust network protection via application-level inspection and scanning. However, each works in a different way with a distinctly different impact upon network latency and performance.

- **Proxy**-based processes are those in which the security device acts as a proxy for the data's destination. The security device will receive and reconstruct a whole file, and examine it for threats, before passing it on to the eventual destination.
- **Stream**-based processes are those in which packets are examined as they pass through in a stream.

The NGFWs have two subscription licensing options for firewall and threat protection features. The following table shows the features included in those licenses, and whether they are proxy or stream-based processes:

Table 1: License type and features

LICENSE TYPE	INCLUDED FEATURE
Base	Intrusion Prevention System (IPS) Stream
Next-Gen Firewall (NGFW)	Application Control—Stream Web Control—Proxy
Advanced Threat Protection (ATP)	IP Reputation—Stream Malware Protection—Stream Antivirus—Proxy URL Filtering—Stream

Proxy-based threat scanning

Proxy-based threat scanning uses a proxy antivirus engine to extract the stored object data, and match that data against various known threat signatures. Threat signatures are contained in regularly updated threat signature database files.

Downloading, scanning, re-ordering, re-assembling, and re-transferring object data, uses large amounts of memory and system CPU resource. In addition, proxying the TCP session reduces the overall data throughput.

Figure 2: Antivirus file scanning

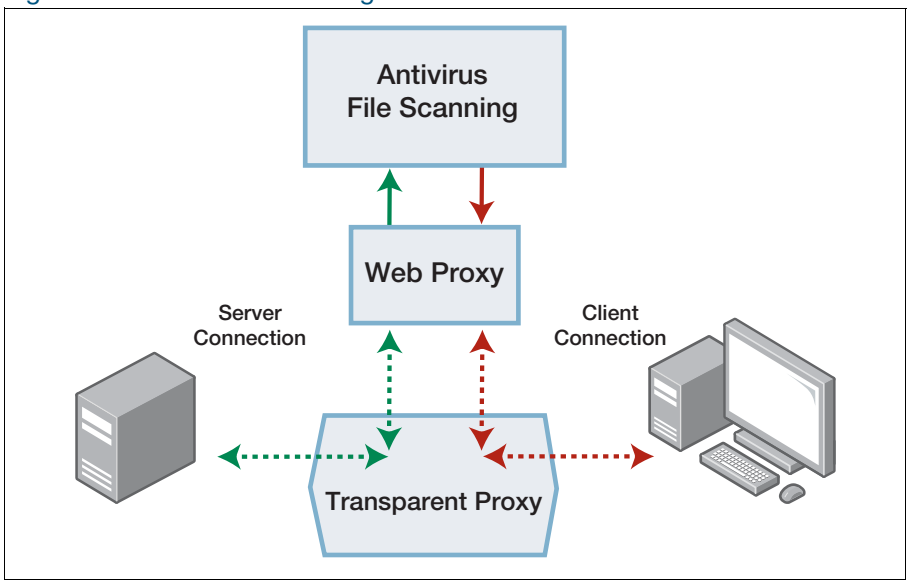
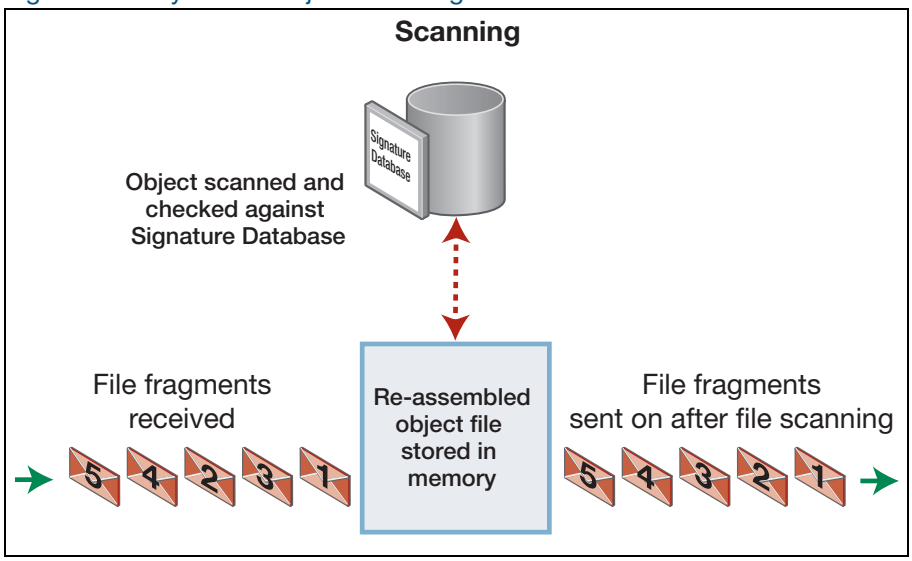


Figure 3: Proxy-based object scanning



Proxy-based scanning provides the best detection by its very nature of operation, however it is also more resource intensive and inherently slower than stream-based scanning. Proxy-based engines (by their very nature) must act as an intermediary and terminate each and every session from a client, establish an associated session to the target server, and monitor the associated session state in a transparent manner. A single user connection to a single website can potentially involve managing multiple simultaneous sessions.

Stream-based threat scanning

In contrast, stream-based scanning processes data in order of arrival.

Stream-based engines are designed for maximum throughput with minimum latency. They don't inherently suffer from having to proxy connections, and do not have to wait to receive, store, and scan entire object data transfers prior to forwarding across a security boundary.

Data is scanned on a layer by layer approach as it arrives. The more data (for a given data stream) passing through the device, the more deeply it is scanned against various threat signatures in real time. The threat signature checks include:

- Source/destination IP against an IP Reputation list (if IP Reputation is configured),
- Layer 7 application data information (such as a HTTP/1.1 Get requests embedded in HTTP packets),
- Embedded user data within the stream, such as a Torrent or Skype and so on.

There is inherently slightly less protection using this approach compared to proxy-based protection, as data is allowed to pass through the security boundary up until the point that a threat is detected, at which point it is blocked.

Performance considerations

Consider a network of hundreds of users accessing websites predominantly containing many images. The images need to be downloaded into system memory, scanned, and forwarded in order to present the entire contents of a web page.

Proxy-based engines

Proxy-based engines scanning requires the proxy engine to manage (proxy) the TCP connection state for all of these individual sessions simultaneously. Large amounts of system resources such as system memory and CPU cycles can be spent performing these actions at the expense of other processes.

This can increase the latencies involved for client to server traffic, but conversely this allows the device to fully download, store, and deeply scan an entire object file transfer for malicious threats and embedded viruses against a threat signature database. Proxy-based engines by their very nature of operation offer the greatest level of protection against threat vectors.

Stream-based engines

Stream-based engines consume noticeably less system memory and CPU processing power compared to proxy-based engines. This is because entire files traversing the security device don't need to be downloaded. Also, file fragments don't need to be re-assembled prior to scanning and subsequent fragmentation and forwarding.

Options available on the Allied Telesis NGFW

As shown in the table in Figure 1 on page 5, Allied Telesis NGFWs combine the benefit of both proxy and stream-based protection options.

You can configure a mix of:

- Proxy-based antivirus scanning of HTTP file transfers for various file object types (for example, zip and image files associated with a website).
- Proxy-based web-control to categorize and filter URL lookups, to help prevent access to known malicious and phishing websites.
- A variety of stream-based threat protection measures, such as IP Reputation, intrusion detection and prevention, malware protection, and blacklist/whitelists URL filtering.

Allied Telesis Next Generation Firewalls control the system resources devoted to proxy-based scanning. Currently, for proxy-based antivirus:

- objects up to 10MB per file can be individually scanned
- up to 100MB of objects can be concurrently scanned

Proxy-based antivirus can extract nested object files to a maximum depth of three, for scanning within an embedded data flow. For extracted decompressed files, up to 10MB size can be scanned.

You can control what alternative actions to take, such as log or allow, if an object file fails to be scanned for whatever reason (for example if it is too large). The default action for failed scans is to deny.

URL Filtering versus web control

Some threat protection jobs can be done in either a proxy-based or a stream-based manner. An example of this is the process of controlling which websites users are allowed to access.

This can be done in a stream-based manner, whereby lists of allowed/blocked websites are stored in the NGFW, and the lists are consulted whenever the NGFW processes packets that are attempting to access a web service.

Alternatively, it can be done in a proxy-based manner, whereby the NGFW extracts the details from such packets, and sends them off to an external service, which responds with a verdict about the suitability of the website the user wants to access.

The Allied Telesis NGFWs implement both methods:

- The stream-based method is called URL Filtering
- The proxy-based method is called Web Control

There are pros and cons to both methods:

URL Filtering

URL Filtering is a **stream-based** service. where URLs are filtered using either a:

- user-defined list (in which up to a thousand blacklist/whitelist URL entries can be configured),
- downloadable list (consisting of many thousands of known malicious website URLs) that can be **frequently updated**.

URLs are extracted from GET, HEAD, POST, PUT, and DELETE HTTP requests for matching against white and black lists in real time. URL filtering might be used within an organization wanting to prevent access to a specific (or user-defined) list of URLs via a low latency stream-based service. Network administrators are allowed to statically configure their own black-listed and white-listed URLs without impeding performance.

Web Control

Web Control is a **proxy-based** web categorization service. This feature utilizes an external categorization service to provide **real-time protection**. The list of malicious and phishing websites is constantly being updated in real-time by the categorization service provider.

The NGFW caches the categorization responses from the external categorization service provider. This avoids unnecessary and repeated external lookups to URLs and improves performance.

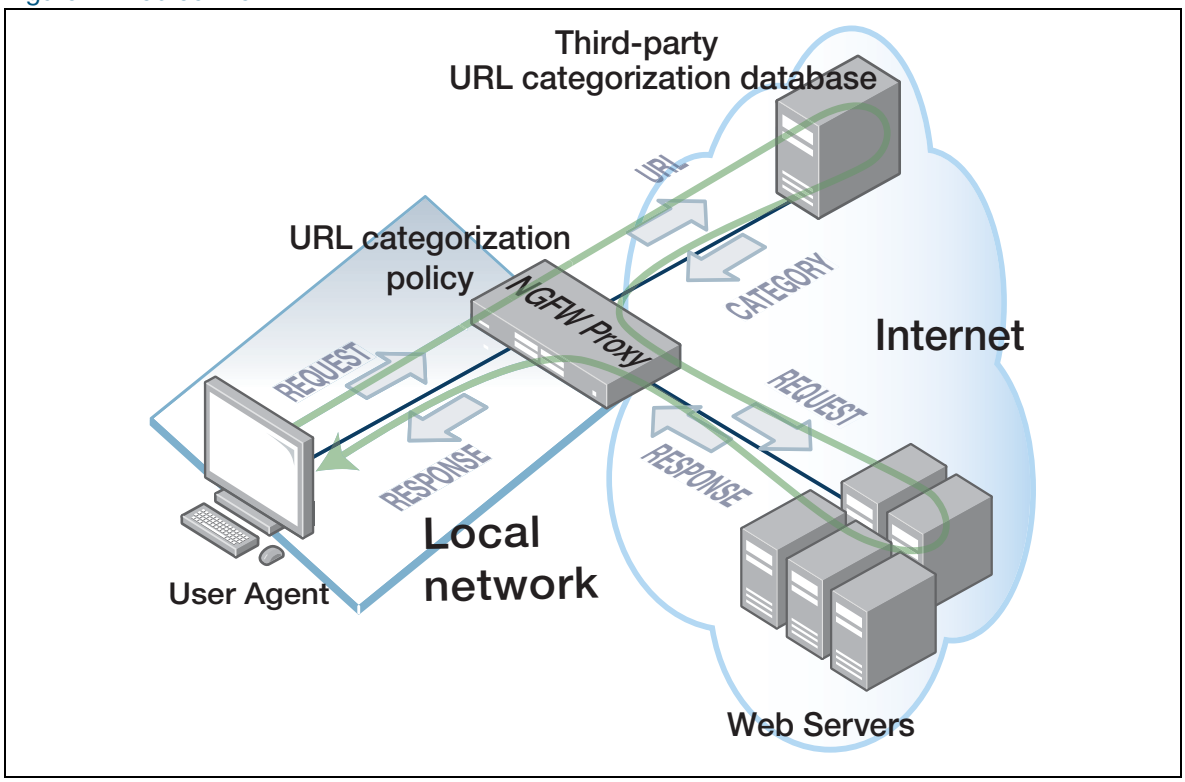
You can define up to 50 URL category match criteria. This enables organizations to provide access to URLs that are relevant to their business, but might otherwise be blocked by the external categorization service.

Summary

By its very nature, **Web Control** therefore provides **maximum protection** against malicious and phishing websites that are dynamically and constantly changing. However, this does come at the expense of the latencies involved with a proxied service. URL filtering has less latency, but can introduce a small risk of exposure to threats between updates.

If both features (Web Control and URL Filtering) are simultaneously enabled, then URLs will be checked first via URL Filtering lists then subsequently be categorized via Web Control. Either feature can block a connection. If a connection is blocked by one feature, the decision cannot be over-ruled by the other feature.

Figure 4: Web control



Performance optimizing architectures

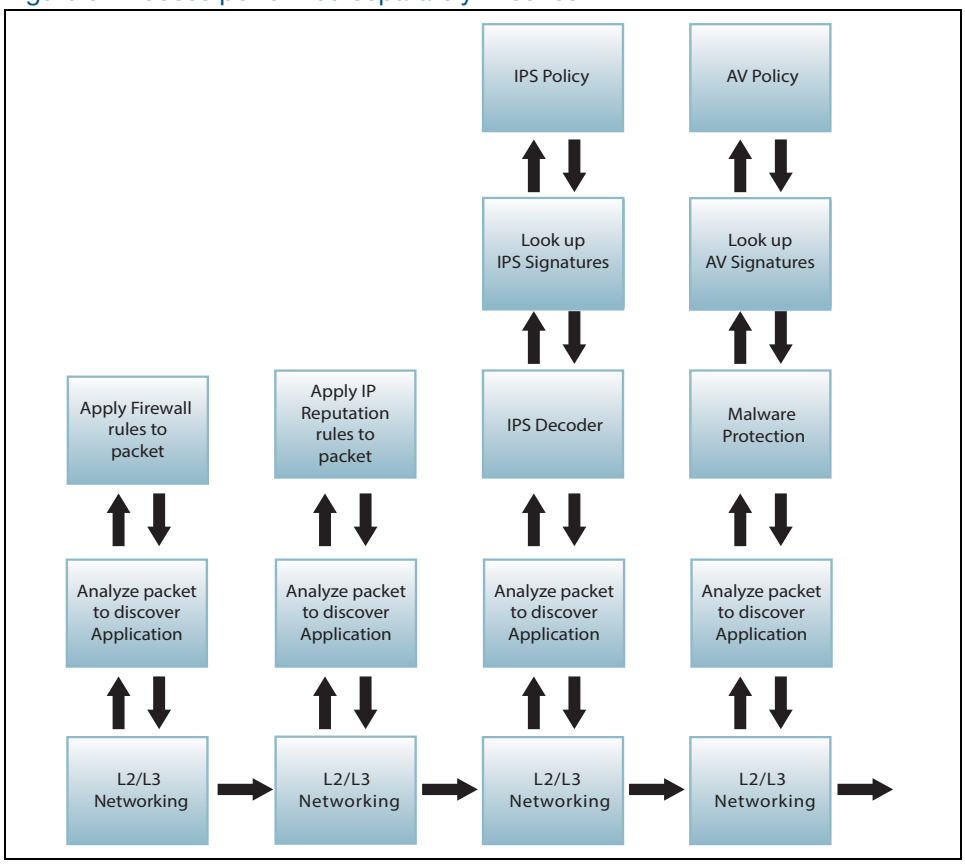
Allied Telesis products integrate a variety of NGFW and Unified Threat Management (UTM) features, that are traditionally provided by a range of devices, into a single security appliance. This allows the network administrator to replace multiple devices with a single appliance, thereby reducing the total cost of ownership.

However, the network administrator may inquire about the effects of configuring all or combinations of these individual protection services, as they are enabled within a single security appliance. A common pitfall with some earlier integrated security appliance implementations was a lack of consistent and predictable performance as each security service was utilized in turn.

Stream-based features are capable of performing high throughput/low latency threat protection. However, as discussed earlier, when proxy-based features are also enabled, performance can decrease and latency will increase as the proxy connections are formed, and data is processed and completely re-scanned through each security feature in turn. This can lead to valid concerns around effects on performance, connections per second, latency, and so on as each security feature is utilized.

This is particularly the case if a traditional processing architecture is used for implementing the security features in the firewall device. In such an architecture, each security process is operated in isolation. Therefore, activities such as identifying the application within a packet would be carried out multiple times on the same packet.

Figure 5: Process performed separately in series

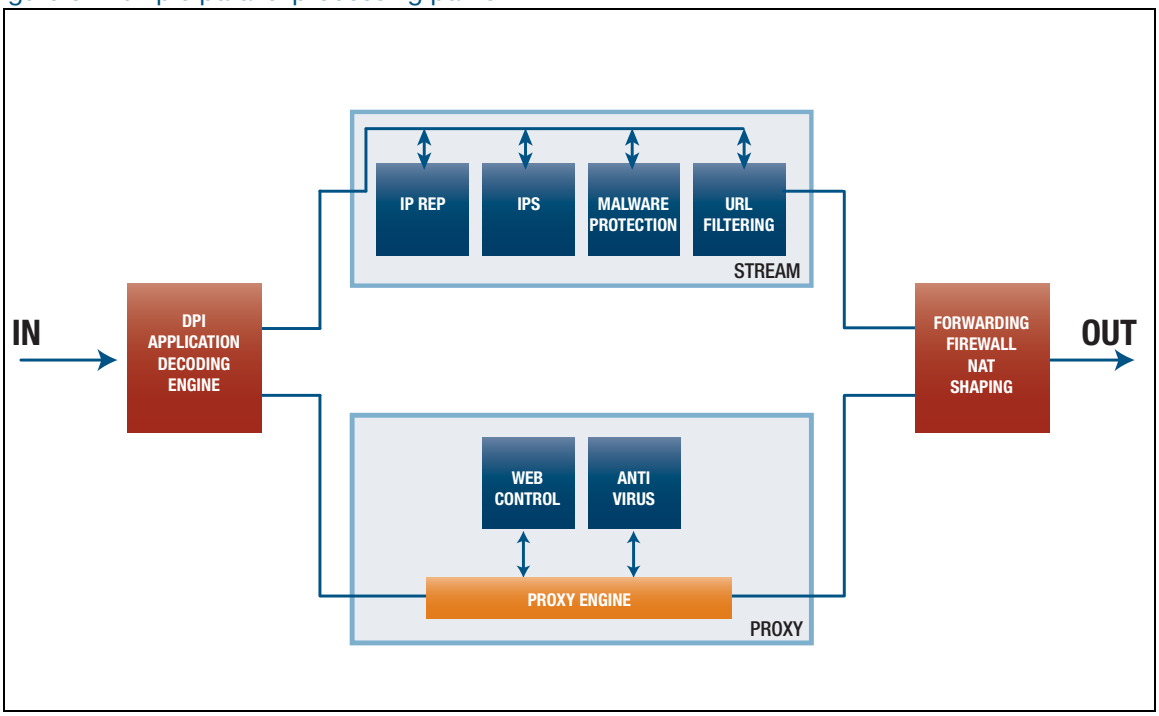


The Allied Telesis NGFW architecture is designed to alleviate these problems as much as possible by utilizing a more efficient architecture model based on multiple parallel processing paths, in conjunction with a multi-core CPU.

Software architecture - multiple parallel processing paths

All data is first identified by its application, protocol, and content within the application decoding engine. So, the process of analyzing the packet, and identifying its application, and other characteristics, is done just once per packet.

Figure 6: Multiple parallel processing paths



Content is processed only via the **appropriate** software data processing path associated with the security feature.

For example, what happens if a proxy-based service (such as HTTP based antivirus) is utilized, and a stream-based feature (such as IP-reputation) is also enabled, and the application data is a UDP based Skype call?

The Skype call will not be processed via the proxy service (as the antivirus service scans HTTP data streams, not UDP data streams).

This means Skype performance won't be affected - even when the proxy service is utilized.

Hardware architecture

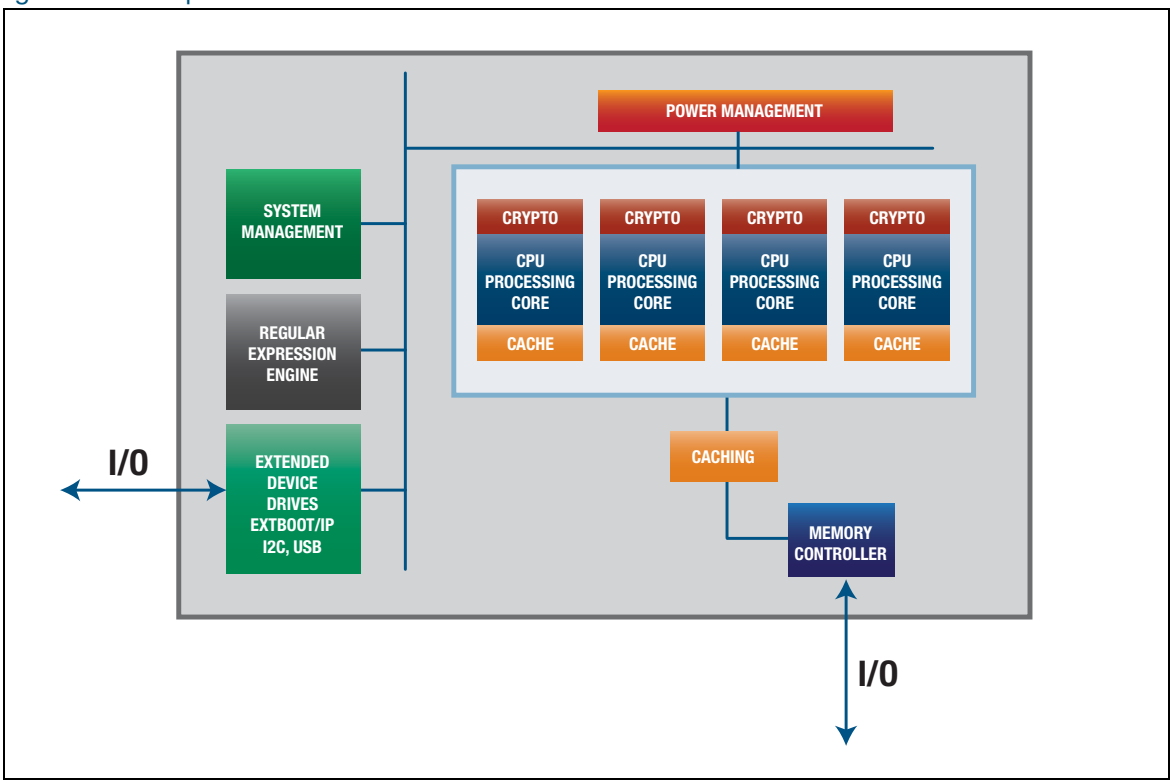
To increase performance, a purpose built, dedicated multi-core Network Services Processing (NSP) CPU is used. The CPU uses a core balancing algorithm that load balances data to be processed by each CPU core, based on a variety of information, including protocol, port numbers, and IPv4/IPv6 source address.

For example, a Youtube video will be processed via one CPU core, and throughput for that application data remains unaffected by unrelated processing for other application data that is being processed on other cores.

Performance throughput is therefore typically measured to match real world usage, with enough flows and enough variance between each flow to ensure all available CPU processing cores are evenly load balanced.

Signature files are processed by the regular expression engine of the CPU. The CPU therefore provides hardware-based processing for signature-based file scanning. Additionally, the CPU also provides on-chip hardware (HW) acceleration for IPsec VPN encryption services. This improves security throughput without the need to encrypt data streams via software, and avoids the need for external off-chip co-processing.

Figure 7: On-chip hardware acceleration



Platform selection

Matching the appropriate NGFW appliance platform to the customer business security requirements is an important consideration.

Business requirements

It is important to understand the business requirements that the appliance is being integrated into. Understanding the customer application data and which applications and user groups are allowed to traverse security boundaries are important considerations, since the device can also be configured to match and enforce customer business rules (as opposed to the traditional public/private/DMZ approach to security).

When selecting and matching the most appropriate NGFW platform for a solution, network administrators need to consider and understand the business size, the number of clients to be protected, and the security requirements for their business application data.

Maximum security (via proxied services) comes with increased client latencies, and potentially increased memory usage as the number of clients and associated sessions increases. Maximum performance via stream-based services could mean a slight reduction in security. Or perhaps a combination of security and performance is required.

Network considerations

It is also important to consider the location of the device within the business infrastructure.

For example, the device might well be suited at the security boundary of a school connection to the Internet, protecting Internet access for hundreds of students (using a traditional public/private DMZ zone structure).

Or it might be suitable in the core of a small business consisting of dozens of employees, providing separation between internal zones and implementing application access control - based on the company business rules.

For example, network administrators may be permitted to access all applications, support staff may be allowed to browse the Internet to search for solutions, while other staff may be limited to access only certain applications.

Hardware and software tables

The NGFW supports all routing, ARP table storage, and bridging between interfaces via software. The NGFW also supports IPv4 and IPv6 routing protocols, including RIP/RIPng, OSPF, OSPFv3, BGP, BGP4+. Multicast support includes IGMP/MLD Querier functionality, and PIM routing.

All of this unicast and multicast routing and bridging activity is performed in software, Load balancing across the multiple CPU cores provides a good level of forwarding performance.

There is limited hardware traffic forwarding performed in the NGFW. Layer 2 traffic is hardware switched between Gigabit switch ports. The Layer 2 switch chip used in the NGFW includes a shared MAC/multicasting hardware table supporting up to 1024 FDB entries, and switch ports can be 802.1q trunked members of up to a recommended configuration limit of 256 VLANs. So, the NGFW provides a useful platform for wire-speed Layer 2 switching within a small LAN.

Frequently asked questions

1. As Deep Packet Inspection (DPI), Antivirus, Malware protection, and IPS security features are activated on an NGFW, what will be the effect on throughput as each feature is enabled, or if all features are enabled at once?

Answer

The throughput values for each stream-based security feature are listed in each product datasheet, and the quoted values apply, regardless of whether the other features are simultaneously enabled and in use, or not.

If proxied services are enabled, and the data is to be processed via the proxy-based path, then the overall throughput (and new connections per second) for the **proxied data** could be **lower** compared to **stream-based** service.

Throughput via **proxied** services will depend on variable and external factors such as:

- Time taken to proxy individual TCP sessions.
- The number of sessions involved.
- Whether the object file data to be scanned is associated with an existing TCP flow, in which case, the rest of the data fragments associated with the object (which might be a picture from a web site) first need to be downloaded, re-assembled, and scanned, re-fragmented, and retransmitted.
- The size and number of objects to be scanned.
- Available memory and system resources.
- The HTTP connection - is it to a previously unseen URL that needs to be sent off to the external web control service for categorization, or is there already a locally stored cached URL response?

However, each application data flow will utilize a **different CPU processing core**, and each application will only be processed via the specific security feature data path. Therefore, switching on proxied services may have little effect on throughput of other unrelated data flows if they are **stream-based**.

For example, if the traffic is **not** an HTTP1.1 Get Request as part of a connection to a website, or is **not** an HTTP object file in the process of being downloaded for re-assembly and scanning, then the traffic won't be processed by the proxy-based Web Control feature processing path (if that proxy feature is enabled). But it will be processed via the stream features code path, and so the effect on throughput of the non-HTTP data will be negligible - regardless of whether proxy features are enabled or in use.

And, for example, if throughput for a couple of data streams are measured, and all stream security features are enabled, and both data streams include an HTTP/1.1 Get request, then each data stream will be processed simultaneously via a different CPU core. So again the effect on each individual stream throughput is negligible, as is the effect on overall traffic throughput for the device.

Most client connections involve varying traffic mixes such as differing application protocol, port, source/destination IP, and the number of sessions involved. So generally, multiple security features can be simultaneously in use. Traffic flows processed via one security feature do not necessarily directly translate into a performance degradation as other security features are utilized.

If multiple stream-based security features are simultaneously enabled and utilized, you can still expect the maximum throughput for each feature as stated in the datasheet, depending on the nature of the traffic mixes involved.

If a single application flow is established through the device that must be processed serially via all stream-based security features, then the effect on the throughput for that individual flow will be negligible. However, there will be some initial latencies involved during the establishment of the flow as security checks are performed by each stream feature in turn.

If multiple flows are established, the overall throughput documented in the datasheet for each individual stream security service still applies. This is due to the **multiple processing paths** in the software architecture, and **multi-core CPU** architecture.

2. What is the effect on the number of concurrent sessions, with antivirus, IPS, DPI, features working/activated simultaneously? If stream-based security features are configured, how much does measured 'sessions per second' degrade if proxied services (web-control and antivirus) are also then enabled?

Answer

It again depends on the nature of customer application data and traffic mix involved. If proxied services are enabled, then the concurrent session per seconds' limit on an existing stream-based security feature may not be affected at all.

This is because the proxied traffic may be processed via a different CPU core, and will take the proxied services processing path, which will be a different processing path to the stream services.

However, (for example) if all of the traffic traversing the device is HTTP 1.1 Get requests, and proxy-based web control is enabled, then TCP connections need to be formed and proxied for each connection request, and the connection requests will be accumulated into bulk categorization requests, and then sent off to the external URL categorization service.

Therefore, various external factors, such as latency of the Internet and response time of the categorization service servers on the Internet, and processing of responses, will slow down the overall connections per second for traffic processed via proxied services.

The connections per second rate via the proxied service may be able to regularly burst up to the maximum (specified in the product datasheet) and within a second or so briefly drop again (potentially in a saw-tooth like fashion) - also depending on the external categorization server responsiveness and Internet latency, as each bulk request is made. The connections per second that can be established for other applications (such as traffic to be processed via stream-based security services) may continue to remain completely unaffected, as those new connections may be processed via a different CPU core, and via a different parallel software processing path.

3. What happens if IP data streams are sent to **different** Internet addresses, and stream-based IP Reputation only is enabled - how will it affect throughput and connections per second?

Answer

IP Reputation will check the source and destination IP addresses of each data stream against the downloaded threat database file in real-time. There will be some initial sub-second latency as each new data stream is established and checked, but once a data stream is established, the effect on overall throughput should be negligible, because the IP Reputation of subsequent packets matching each established flow won't need to be 'rechecked'.

The number of concurrent connections per second will still be able to burst up to the maximum for various other application data streams, such as new data streams between the same source and destination address.

4. What happens if there are multiple data streams to and from the **same** source and destination Internet addresses, and stream-based IP Reputation only is enabled - how will it affect throughput and connections per second?

Answer

As you might have guessed, it also depends on the nature of the customer application data flows.

If **all** traffic is TCP, but each session uses the same source IP and same destination IP (but each session uses different TCP port numbers), then the connections per second rate would not be affected when IP reputation is enabled.

This is because the IP Reputation check only needs to be performed once for each source IP and destination IP. Subsequent new TCP sessions (with the same source and destination IP) do not need to be rechecked. This also means the overall throughput remains unaffected.

5. What is the effect on the throughput and connections per second, when Deep Packet Inspection (DPI) is turned on, as compared with only having a basic Intrusion Detection System (IDS) / Intrusion Protection System (IPS) rule set enabled?

Answer

There is no difference in throughput with DPI turned on compared to only having a basic IDS/IPS rule set enabled.

Enabling DPI enables awareness of, and support for, a higher number of Internet applications to be scanned against the DPI engine, signature database, and flow-based connection tracker. The NGFW becomes **aware** of the appropriate behavior of a far greater number of applications when DPI is enabled.

In other words, enabling DPI means a much higher number of applications are able to be automatically recognized, and to be deeply probed by the DPI engine. All this extra application recognition and probing is offloaded to the DPI engine, so there is no effect on overall throughput, nor connections per second that can be established and tracked via the internal software connection tracker table.

Without DPI being enabled, 'custom applications' would have to be manually defined and manually configured to allow less-common applications to be recognized by the firewall security processing - if required by the business. Creating custom applications is somewhat equivalent to legacy firewall **allow** rules (pinholes) based on protocol and port numbers, so, this provides basic recognition of those applications, and the provision of allow/deny rules for those applications.

However, the DPI functionality goes much further than simply recognizing applications - it can check what actions the users are performing within the applications. Therefore, the DPI functionality enables blocking of particular actions within the applications.

6. What is the effect on connections per second or throughput as the number of firewall and/or NAT rules increases? And are there any guidelines for numbers of rules, and entity definitions, such as zones, networks, and hosts that can be realistically configured?

Answer

As you configure additional NAT/firewall rules (in the order of tens or hundreds of rules), there will be additional sub-second latencies to check each new session flow against the entity rule-sets, but once a flow is established in the internal connection tracking table, the addition of more rules has negligible effect on latency and throughput for the established flows.

There is no enforced software limit to the number of configurable rules, and entities. However, as a guideline, a couple of thousand rules to allow application data to flow between various firewall entity definitions would **not** be recommended. A more realistic maximum number of rules would be in the order of several hundred.

It would be reasonable to configure a few zones, based on business functions, or alternatively configure the device to provide boundary security via the traditional three zone approach, i.e. DMZ, private and public zones. Configuring up to tens of network entity definitions would be reasonable (maybe several networks per zone), and up to a couple of hundred host-specific entity definitions in most installations.

7. So what are the limits for connections per second?

Answer

Individual limits for each product are stated in each product datasheet. For example, for the AR4050S—300,000 simultaneous firewall sessions is the real-world (measured) maximum and the limit is enforced in software.

Connections above that limit will need to retry, to eventually become connected when an existing session is terminated.

And (for example) the AR4050S can happily establish new firewall flows at an average measured rate of 12,000 new sessions per second. So, the device can comfortably support a burst of thousands of simultaneous connection requests.

However, the IDS/IPS may limit the number of sessions a single client host can initiate, (which typically occurs for traffic initiated from an infected host whose behavior is abnormal). For example, an infected host sending an excessive rate of TCP SYNs, or sending inappropriate application protocol messages, will be throttled.

Additionally, (supported in software version 5.4.6 onwards), limits can be configured on a 'per-entity' basis. These user-defined limits enforce the maximum number of simultaneous connections each device located within an entity (such as a zone or network) can establish via the NGFW, up to a maximum of 4096 new TCP connection requests per client.

Lastly, if (stream-based) URL filtering is enabled and the traffic is mostly new connections to web sites (HTTP 1.1 get requests), then the maximum rate of new connections per second reduces as the number of entries in the black/white lists increases.

For example, a filter list consisting of 65k URLS could result in a 50% reduction in the maximum rate of new connections per second. A (much smaller, user-defined) filter list containing a few hundred URLs could result in a 30% reduction in the maximum rate of new connections per second.

8. If all stream-based security services are simultaneously utilized, what is the limiting factor for throughput?

Answer

Initially, there would be some higher latency for new flows as each individual stream-based security feature check is performed in turn (DPI, then IP Reputation, IPS, Malware Protection, and URL filtering). The limiting factor/bottleneck will be the IPS throughput value (quoted in the product datasheet) **if** configured, as **all** data will flow through this service.

9. What is the difference between IPS and Application Control?

Answer

IPS is unlicensed. When enabled via the IPS command set, the feature scans all data streams for a wide range of anomalies as each data stream traverses the device. When anomalies are detected, by default an alert log message is generated. The user is able to configure actions (for example, alert, or drop) on a per IPS category basis via the category action command. The documented command **show ips categories** lists the available categories, and anomalies being checked. This includes various checksum checks on a wide variety of traffic types, such as HTTP data, IP data, TCP, and UDP data streams.

By default, the firewall is aware of a small pre-defined set of commonly used applications, listed via the **show application detail** command. Application awareness allows the firewall to understand how each pre-defined application is supposed to operate and behave. For example, many applications might need specific TCP or UDP ports to be dynamically opened/closed or negotiated, or specific information to be contained within headers, or specific handshaking to occur based on a

protocol state machine and so on. Custom applications can also be configured by the user to allow access for specific protocols and port numbers, in much the same way as traditional firewall pin-holes or allow rules. But these lack application awareness.

Application control is licensed and configured via the DPI command set. The feature dynamically and automatically implements full application awareness for a large and ever-changing range of applications being used on the Internet, via regular updates from the external provider.

10. What is the effect on throughput if ''everything" is enabled, all proxy services, all stream services, and the device is very heavily loaded?

Answer

If all of the traffic flowing through the device is basically HTTP traffic to many websites from many clients, then the bottleneck will likely be the proxy-based services. By their very nature, the performance of proxy-based services are also limited by factors external to the device, such as the external categorization service for a lookup.

If hundreds of users are simultaneously browsing hundreds of web sites with lots of images, each website connection from each client browser therefore has many associated sessions with many object image files to download/reassemble/scan/fragment/retransmit via the proxy-based HTTP Antivirus security service checking. In this case, there could be some delay when connecting to some websites for some clients and/or a failure to display all website content for each individual website.

Additionally, (with proxy-based Web Control enabled), users may experience slowness in the initial connection to some websites, because connections are proxied and bulk categorization requests are sent off for external categorization and responses processed.

Once a website URL has been categorized (and locally cached), the latencies involved for subsequent connections to the same websites will be reduced, and delays will be primarily based on the time to proxy the TCP connections via the proxy engine, and for any other higher priority internal system processes to be accomplished.

However, if the CPU becomes completely overwhelmed (for example due to misconfiguration or being placed into a completely inappropriate network environment), with all CPU processing cores therefore fully utilized, with no free CPU cycles to dedicate to various processes, then system (control plane) traffic will take priority over data plane traffic, and packet loss for traffic will occur as transmit and receive buffer queues are exhausted.

But this logic applies to any vendors product if it is placed into an unsuitable environment.

Allied Telesis NGFWs provide comprehensive threat protection in a fully integrated security platform, using specialized multi-core CPUs optimized for single-pass low-latency performance.

North America Headquarters | 19800 North Creek Parkway | Suite 100 | Bothell | WA 98011 | USA | T: +1 800 424 4284 | F: +1 425 481 3895
Asia-Pacific Headquarters | 11 Tai Seng Link | Singapore | 534182 | T: +65 6383 3832 | F: +65 6383 3830
EMEA & CSA Operations | Incheonweg 7 | 1437 EK Rozenburg | The Netherlands | T: +31 20 7950020 | F: +31 20 7950021

alliedtelesis.com