Web Security Appliance Design Guide

Contents

Introduction

This document describes how to design the Cisco Web Security Appliance (WSA) and associated components for optimal performance.

Background Information

When you design a solution for the WSA, it requires careful consideration, not only in regards to the configuration of the appliance itself, but also the associated network devices and their features. Every network is a collaboration of multiple devices, and if one of them does not participate correctly in the network, then user experiences might decline.

There are two main components that must be considered when you configure the WSA: the hardware and the software. The hardware comes in two different types. The first is the physical type of hardware, such as the S170, S380, and S680 Series models, as well as other End of Life (EoL) models, such as the S160, S360, S660, S370, and S670 Series models. The other hardware type is virtual, such as the S000v, S100v, and S300v Series models. The Operating System (OS) that runs on this hardware is called *AsyncOS for Web*, which is based on FreeBSD at its core.

The WSA offers proxy service and also scans, inspects, and categorizes all traffic (HTTP, HTTPS, and File Transfer Protocol (FTP)). All of these protocols run on top of TCP and heavily rely on Domain Name System (DNS) for proper operation. For these reasons, the network health is vital for proper operation of the appliance and its communication with various parts of the network, both inside and outside of the enterprise control.

Design

Use the information that is described in this section in order to design the WSA and related components for optimal performance.

Network

An error free, fast network is vital for the proper operation of the WSA. If the network is unstable, user experience might decline. Network problems are usually detected when web pages take longer to reach or are unreachable. The initial inclination is blame the appliance, but it is usually the network that misbehaves. Thus, careful consideration and audit should be made in order to ensure that the network offers the best service for high-level application protocols such as HTTP, HTTPS, FTP, and DNS.

General Considerations

Here are some general considerations that you can implement in order to ensure the best network behavior:

- Ensure that the Layer 2 (L2) network is stable, that the spanning-tree operation is correct, and that there are not frequent spanning-tree computations and topology changes.

- The routing protocol that is used should also provide fast convergence and stability. The Open Shortest Path First (OSPF) fast timers or the Enhanced Interior Gateway Routing Protocol (EIGRP) are good choices for such a network.

- Always use at least two data interfaces on the WSA: one that faces the end-user computers, and another one for outbound operation (connected to the upstream proxy or Internet). This is done in order to eliminate possible resource constrains, such as when the number of TCP ports are exhausted or when network buffers become full (with the use of a single interfaces for both inside and outside especially).

- Dedicate the Management Interface for management-only traffic in order to increase security. In order to achieve this via the GUI, navigate to **Network > Interfaces** and check the **Separate routing (M1 port restricted to appliance management services only)** check box.

- Use fast DNS servers. Any transaction via the WSA requires at least one DNS lookup (if not in the cache). A DNS server that is slow or misbehaves affects any transaction and is observed as delayed or slow internet connectivity.

- When separate routing tables are used, these rules apply:

  All interfaces are included in the default *Management* routing table (M1, P1, P2).

  Only Data interfaces are included in the *Data* routing table.

  Note: The separation of routing tables is not per interface, but rather per service. For example, traffic between the WSA and the Microsoft Active Directory (AD) domain controller always obey the routes that are specified in the Management routing table, and it is possible to configure routes that point out of the P1/P2 interface in this table. It is not possible to include routes in the Data routing table that use the Management interfaces.

Load-Balancing

Here are some load-balancing considerations that you can implement in order to ensure the best network behavior:

- DNS rotation – This is the term used when a single hostname is used as a proxy, but it has multiple A records on the DNS server. Each client resolves this to a different IP address and uses different proxies. A limitation is that changes of DNS records are reflected on clients upon reboot (local DNS caching), so it offers a low level of robustness if a change must be made. However, this is transparent to the end-users.

- Proxy Address Control (PAC) files – These are proxy-automatic scripting files that determine how each URL should be handled on a browser based on the written functions within it. It has the feature to forward the same URL always directly or to the same proxy.

- Auto discovery – This describes the use of DNS/DHCP methods in order to obtain PAC files (described in the previous consideration). Usually, these first three considerations are combined into one solution. However, this can be complicated and many user-agents, such as Microsoft Office, Adobe Downloader, Javascripts, and Flash, cannot read PAC files at all.

- Web Cache Control Protocol (WCCP) – This protocol (especially WCCP Version 2) provides a robust and very powerful way to create load-balancing between several WSAs and also incorporate high availability.

- Separate load-balancing appliance(s) – Cisco recommends that you use load-balancers as dedicated machines.

Firewalls

Here are some Firewall considerations that you can implement in order to ensure the best network behavior:

- Ensure that Internet Control Message Protocol (ICMP) is allowed throughout the network from each source. This is vital, as the WSA depends on the path Maximum Transition Unit (MTU) discovery mechanism, as described in RFC 1191, which depends on ICMP Echo requests (type and Echo replies (type 0), and ICMP unreachable-fragmentation is required (type 3, code 4). If you disable path MTU discovery on the WSA with the **pathmtudiscovery** CLI command, then the WSA uses the default MTU of 576 bytes, as per RFC 879. This impacts performance due to increased overhead and a reassembly of packets.

- Ensure that there is no asymmetrical routing inside of the network. While this is not a problem on the WSA, any Firewall that is encountered along the path drops the packets because it has not received both sides of the communication.

- With Firewalls, it is very important to exclude the WSA IP addresses from threats as regular end computer stations. The Firewall might block
- the WSA IP addresses due to too many connections (as per general Firewall knowledge).

- If Network Address Translation (NAT) is employed for any WSA IP address on the customer premises device, ensure that each WSA uses a separate outside global address in the NAT. If

you use NAT for multiple WSAs that have a single outside global address, you might encounter these issues:

All of the connections from all of the WSAs to the outside world use a single outside global address, and the Firewall quickly runs out of resources.

If there is a spike of traffic towards that single destination, the destination server might block it and cut off the entire enterprise from access to this resource. This might be a valuable resource as the company Cloud storage, the Office Cloud connections, or the per-computer antivirus software updates.

Identities

Remember that the *logical AND* principle applies in all components of the identity. For example, if you configure both the user-agent and IP address, it means the user-agent *from* this IP address. It does not mean the user-agent *or* this IP address.

Use one identity for authentication of the same surrogate type (or no surrogate) and/or user-agent.

It is important to ensure each identity that requires authentication includes the user-agent strings for known browsers/user-agents that support proxy authentication, such as Internet Explorer, Mozilla Firefox, and Google Chrome. There are some applications that require Internet access but do not support proxy/WWW authentication.

Identities are matched top to bottom with the search for matches that ends on the first matched entry. For this reason, if you have *Identity 1* and *Identity 2* configured, and a transaction matches Identity 1, it is not checked against Identity 2.

Access/Decryption/Routing/Outbound Malware Policies

These policies are applied against different types of traffic:

- Access policies are applied against plain HTTP or FTP connections. They determine whether the transaction should be accepted or dropped.

- Decryption policies determine whether HTTPS transactions should be decrypted, dropped, or passed through. If the transaction is decrypted, then the consecutive part of it can be seen as a plain HTTP request and is matched against Access policies. If you must drop an HTTPS request, drop it in the Decryption policies, not in the Access policies. Otherwise, it consumes more CPU and memory for a dropped transaction first to be decrypted and then to be dropped.

- Routing policies determine the upstream direction of a transaction once it his allowed through the WSA. This applies if there are upstream proxies or if the WSA is in *Connector* mode and sends traffic to the Cloud Web Security tower.

- Outbound malware policies are applied against HTTP or FTP uploads from end-users towards web servers. This is usually seen is an HTTP Post request.

For each type of policy, it is important to remember that the *logical OR* principle applies. If you

have multiple identities referred, then the transaction should match any of the identities that are configured.

For more granular control, use these policies. Wrongly configured identities per policy can create issues, where it is more beneficial to use several identities referenced in a policy. Remember that identities do not impact the traffic, they just identify the types of traffic for later matches in a policy.

Often times, Decryption policies use identities with authentication. While this is not wrong and is sometimes needed, the use of an identity with authentication referenced in the Decryption policy means that all transactions that match the Decryption policy are decrypted in order for authentication to take place. The decryption action might be dropped or passed through, but since there is an identity with authentication, the decryption takes place in order to later drop or pass through the traffic. This is expensive and should be avoided.

Some configurations have been observed that contain 30 or more identities and 30 or more Access policies, where all of the Access policies include all of the identities. In this case, there is no need to use this many identities if they are matched in all of the Access policies. While this does not harm the appliance operation, it creates confusion with attempts to troubleshoot and is expensive in regards to performance.

Custom URL Categories

The use of custom URL categories is a powerful tool on the WSA that is usually misunderstood and misused. For example, there are configurations that contain all video sites for matches in the identity. The WSA has a built-in tool that automatically updates when video sites change URLs, which occurs frequently. Thus, it makes sense to allow the WSA to manage the URL categories automatically, and use the custom URL categories for special, not yet categorized sites.

Be very careful with regular expressions. If special character matches such as dot (.) and star(*) are used, they might prove to be very CPU and memory extensive. The WSA expands any regular expression to match it against each transaction. For example, here is a regular expression:

```
example.*
```

This expression will match any URL that contains the word *example*, not only the *example.com* domain. Avoid the use of *dot* and *star* in regular expressions and use them only as a last resort.

Here is another example of a regular expression that might create issues:

```
www.example.com
```

If you use this example in the Regular Expressions filed, it will not only match *www.example.com*, but also *www.www3example2com.com*, as the dot here means a*ny character*. If you desire to match only *www.example.com*, escape the dot:

```
www\.example\.com
```

In this case, there is no reason to use the Regular Expressions feature when you can include this inside the custom URL category domain with this format:

```
www.example.com
```

Anti-Malware and Reputation

If more than one scanning engine is enabled, consider the option to enable adaptive scanning also. Adaptive scanning is a powerful but small engine on the WSA that pre-scans each request and determines the comprehensive engine that should be used in order to scan requests. This slightly increases performance on the WSA.