Guía de diseño de dispositivos de seguridad web

Contenido

Introducción

Este documento describe cómo diseñar Cisco Web Security Appliance (WSA) y los componentes asociados para obtener un rendimiento óptimo.

Antecedentes

Al diseñar una solución para WSA, es necesario tener en cuenta con cuidado, no sólo en lo que respecta a la configuración del dispositivo en sí, sino también a los dispositivos de red asociados y sus funciones. Cada red es una colaboración de varios dispositivos, y si uno de ellos no participa correctamente en la red, las experiencias de los usuarios podrían disminuir.

Hay dos componentes principales que deben tenerse en cuenta al configurar el WSA: el hardware y el software. El hardware se presenta en dos tipos diferentes. El primero es el tipo físico de hardware, como los modelos de las series S170, S380 y S680, así como otros modelos de fin de vida útil (EoL), como los modelos de las series S160, S360, S660, S370 y S670. El otro tipo de hardware es virtual, como los modelos de las series S000v, S100v y S300v. El sistema operativo (SO) que se ejecuta en este hardware se denomina *AsyncOS para Web*, que se basa en FreeBSD en su núcleo.

WSA ofrece servicio proxy y también analiza, inspecciona y clasifica todo el tráfico (HTTP, HTTPS y protocolo de transferencia de archivos (FTP)). Todos estos protocolos se ejecutan en la parte superior de TCP y dependen en gran medida del sistema de nombres de dominio (DNS) para un funcionamiento adecuado. Por estas razones, el estado de la red es vital para el correcto funcionamiento del dispositivo y su comunicación con diversas partes de la red, tanto dentro como fuera del control empresarial.

Diseño

Utilice la información que se describe en esta sección para diseñar el WSA y los componentes

relacionados para obtener un rendimiento óptimo.

Red

Una red rápida y sin errores es vital para el correcto funcionamiento del WSA. Si la red es inestable, la experiencia del usuario podría disminuir. Los problemas de red se detectan generalmente cuando las páginas web tardan más en llegar o son inalcanzables. La tendencia inicial es culpar al dispositivo, pero normalmente es la red la que se comporta mal. Por lo tanto, se debe realizar un examen y una auditoría cuidadosos para garantizar que la red ofrece el mejor servicio para los protocolos de aplicaciones de alto nivel como HTTP, HTTPS, FTP y DNS.

Consideraciones generales

A continuación se indican algunas consideraciones generales que puede implementar para garantizar el mejor comportamiento de la red:

- Asegúrese de que la red de capa 2 (L2) es estable, de que el funcionamiento del árbol de extensión es correcto y de que no hay frecuentes cálculos del árbol de extensión ni cambios de topología.

- El protocolo de ruteo que se utiliza también debe proporcionar una rápida convergencia y estabilidad. Los temporizadores rápidos Open Shortest Path First (OSPF) o Enhanced Interior Gateway Routing Protocol (EIGRP) son buenas opciones para dicha red.

- Utilice siempre al menos dos interfaces de datos en el WSA: uno que se dirige a los equipos del usuario final y otro para el funcionamiento saliente (conectado al proxy ascendente o a Internet). Esto se hace para eliminar las posibles restricciones de recursos, como cuando se agota el número de puertos TCP o cuando las memorias intermedias de red se llenan (con el uso de una única interfaz tanto para el interior como para el exterior especialmente).

- Dedique la interfaz de administración al tráfico solo de administración para aumentar la seguridad. Para lograr esto a través de la GUI, navegue hasta **Red > Interfaces** y marque la **casilla de verificación Ruteo separado (puerto M1 restringido a servicios de administración de dispositivos solamente)**.

- Utilice servidores DNS rápidos. Cualquier transacción a través de WSA requiere al menos una búsqueda de DNS (si no en la caché). Un servidor DNS que se comporta de forma lenta o incorrecta afecta a cualquier transacción y se observa como conectividad a Internet lenta o retardada.

- Cuando se utilizan tablas de ruteo separadas, se aplican estas reglas:

  Todas las interfaces se incluyen en la tabla de routing *Management* predeterminada (M1, P1, P2).

  Sólo las interfaces de datos se incluyen en la tabla de ruteo *de datos*.

  Nota: La separación de las tablas de ruteo no es por interfaz, sino por servicio. Por ejemplo, el tráfico entre el WSA y el controlador de dominio de Microsoft Active Directory (AD)

siempre obedece a las rutas especificadas en la tabla de routing de administración, y es posible configurar rutas que se dirijan desde la interfaz P1/P2 en esta tabla. No es posible incluir rutas en la tabla de ruteo de datos que utilizan las interfaces de administración.

Balanceo de Carga

A continuación se indican algunas consideraciones de equilibrio de carga que puede implementar para garantizar el mejor comportamiento de la red:

- Rotación de DNS - Este es el término utilizado cuando se utiliza un solo nombre de host como proxy, pero tiene varios registros A en el servidor DNS. Cada cliente resuelve esto en una dirección IP diferente y utiliza proxies diferentes. Una limitación es que los cambios de los registros DNS se reflejan en los clientes al reiniciar (almacenamiento en caché de DNS local), por lo que ofrece un bajo nivel de solidez si se debe realizar un cambio. Sin embargo, esto es transparente para los usuarios finales.

- Archivos de control de direcciones proxy (PAC): son archivos de secuencias de comandos automáticos de proxy que determinan cómo se debe gestionar cada URL en un explorador en función de las funciones escritas que contiene. Tiene la función de reenviar la misma URL siempre directamente o al mismo proxy.

- Detección automática: describe el uso de métodos DNS/DHCP para obtener archivos PAC (descritos en la consideración anterior). Por lo general, estas tres primeras consideraciones se combinan en una sola solución. Sin embargo, esto puede ser complicado y muchos agentes de usuario, como Microsoft Office, Adobe Downloader, Javascripts y Flash, no pueden leer archivos PAC en absoluto.

- Protocolo de control de caché web (WCCP): este protocolo (especialmente la versión 2 de WCCP) proporciona una forma robusta y muy potente de crear equilibrio de carga entre varios WSA y también incorpora alta disponibilidad.

- Dispositivos de equilibrio de carga independientes: Cisco recomienda utilizar equilibradores de carga como máquinas dedicadas.

Firewalls

A continuación se indican algunas consideraciones de firewall que puede implementar para garantizar el mejor comportamiento de la red:

- Asegúrese de que se permite el protocolo de mensajes de control de Internet (ICMP) en toda la red desde cada origen. Esto es vital, ya que WSA depende del mecanismo de detección de la unidad máxima de transición (MTU) de la ruta, como se describe en RFC 1191, que depende de las solicitudes de eco ICMP (tipo ﷺ y respuestas de eco (tipo 0), y se requiere fragmentación inalcanzable ICMP (tipo 3, código 4). Si inhabilita la detección de MTU de trayectoria en el WSA con el comando CLI **pathmtudiscovery**, entonces el WSA utiliza la MTU predeterminada de 576 bytes, según RFC 879. Esto afecta el rendimiento debido al aumento de la sobrecarga y al reensamblado de paquetes.

- Asegúrese de que no haya un ruteo asimétrico dentro de la red. Aunque esto no es un problema en el WSA, cualquier firewall que se encuentre a lo largo de la trayectoria descarta los paquetes porque no ha recibido ambos lados de la comunicación.

- Con los firewalls, es muy importante excluir las direcciones IP de WSA de las amenazas como estaciones informáticas finales normales. El firewall podría bloquear
- las direcciones IP de WSA debido a demasiadas conexiones (según el conocimiento general del firewall).

- Si se utiliza la traducción de direcciones de red (NAT) para cualquier dirección IP de WSA en el dispositivo de las instalaciones del cliente, asegúrese de que cada WSA utiliza una dirección global externa independiente en la NAT. Si utiliza NAT para varios WSA que tienen una única dirección global externa, puede que se encuentre con estos problemas:

  Todas las conexiones de todos los dispositivos WSA al mundo exterior utilizan una única dirección global externa, y el firewall se queda rápidamente sin recursos.

  Si hay un pico de tráfico hacia ese único destino, el servidor de destino podría bloquearlo y cortar el acceso de toda la empresa a este recurso. Este puede ser un recurso valioso como el almacenamiento en nube de la empresa, las conexiones de Office Cloud o las actualizaciones de software antivirus por ordenador.

Identidades

Recuerde que el principio *lógico AND* se aplica en todos los componentes de la identidad. Por ejemplo, si configura tanto el usuario-agente como la dirección IP, significa el usuario-agente *de* esta dirección IP. No significa el usuario-agente *o* esta dirección IP.

Utilice una identidad para la autenticación del mismo tipo de sustituto (o sin sustituto) y/o agente de usuario.

Es importante asegurarse de que cada identidad que requiere autenticación incluye las cadenas user-agent para exploradores/agentes de usuario conocidos que admiten autenticación proxy, como Internet Explorer, Mozilla Firefox y Google Chrome. Hay algunas aplicaciones que requieren acceso a Internet pero que no admiten autenticación de proxy/WWW.

Las identidades coinciden de arriba a abajo con la búsqueda de coincidencias que finaliza en la primera entrada coincidente. Por esta razón, si tiene *Identidad 1* e *Identidad 2* configuradas, y una transacción coincide con Identidad 1, no se verifica con Identidad 2.

Políticas de acceso/descifrado/routing/malware saliente

Estas políticas se aplican a diferentes tipos de tráfico:

- Las políticas de acceso se aplican a conexiones HTTP o FTP simples. Determinan si la transacción debe aceptarse o cancelarse.

- Las políticas de descifrado determinan si las transacciones HTTPS se deben descifrar, descartar o pasar. Si se descifra la transacción, la parte consecutiva puede verse como una

solicitud HTTP simple y se compara con las políticas de acceso. Si debe descartar una solicitud HTTPS, suéltela en las políticas de descifrado, no en las políticas de acceso. De lo contrario, consume más CPU y memoria para que una transacción descartada se descifre primero y luego se descarte.

- Las políticas de routing determinan la dirección ascendente de una transacción una vez que se permite a través del WSA. Esto se aplica si hay proxies ascendentes o si el WSA está en modo *Conector* y envía tráfico a la torre Cloud Web Security.

- Las políticas de malware saliente se aplican a las cargas HTTP o FTP de usuarios finales hacia servidores web. Esto suele verse como una solicitud de envío HTTP.

Para cada tipo de política, es importante recordar que se aplica el principio *lógico OR*. Si se hace referencia a varias identidades, la transacción debe coincidir con cualquiera de las identidades configuradas.

Para un control más granular, utilice estas políticas. Las identidades mal configuradas por política pueden crear problemas, donde es más beneficioso utilizar varias identidades a las que se hace referencia en una política. Recuerde que las identidades no afectan al tráfico, simplemente identifican los tipos de tráfico para las coincidencias posteriores en una política.

A menudo, las políticas de descifrado utilizan identidades con autenticación. Aunque esto no es incorrecto y a veces se necesita, el uso de una identidad con autenticación referenciada en la política de descifrado significa que todas las transacciones que coincidan con la política de descifrado se descifran para que se realice la autenticación. La acción de descifrado puede ser descartada o transferida, pero dado que hay una identidad con autenticación, el descifrado se realiza para luego descartar o pasar a través del tráfico. Esto es caro y debe evitarse.

Se han observado algunas configuraciones que contienen 30 o más identidades y 30 o más políticas de acceso, donde todas las políticas de acceso incluyen todas las identidades. En este caso, no es necesario utilizar estas muchas identidades si coinciden con todas las políticas de acceso. Aunque esto no perjudica la operación del dispositivo, crea confusión con los intentos de solucionar problemas y es caro en lo que respecta al rendimiento.

Categorías de URL personalizadas

El uso de categorías de URL personalizadas es una potente herramienta en el WSA que normalmente se malinterpreta y se utiliza incorrectamente. Por ejemplo, hay configuraciones que contienen todos los sitios de vídeo para coincidencias en la identidad. WSA cuenta con una herramienta integrada que se actualiza automáticamente cuando los sitios de vídeo cambian de URL, lo que ocurre con frecuencia. Por lo tanto, tiene sentido permitir que el WSA administre automáticamente las categorías de URL y utilice las categorías de URL personalizadas para los sitios especiales, que aún no están categorizados.

Tenga mucho cuidado con las expresiones regulares. Si se utilizan coincidencias especiales de caracteres como punto (.) y estrella (*), puede que resulten muy extensas en la CPU y en la memoria. WSA expande cualquier expresión regular para que coincida con cada transacción. Por ejemplo, aquí hay una expresión regular:

```
example.*
```

Esta expresión coincidirá con cualquier URL que contenga la palabra *ejemplo*, no sólo el dominio

example.com. Evite el uso de *punto* y *estrella* en expresiones regulares y utilícelas sólo como último recurso.

Aquí hay otro ejemplo de una expresión regular que podría crear problemas:

```
www.example.com
```

Si utiliza este ejemplo en el campo Expresiones regulares, no sólo coincidirá con *www.example.com*, sino también con *www.www3example2com.com*, ya que el punto aquí significa *cualquier carácter*. Si desea coincidir sólo con *www.example.com*, escape el punto:

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

En este caso, no hay razón para utilizar la función de expresiones regulares cuando se puede incluir esto dentro del dominio de categoría de URL personalizado con este formato:

```
www.example.com
```

Anti-Malware y reputación

Si hay más de un motor de escaneo activado, considere la opción de habilitar también el escaneo adaptable. El escaneo adaptable es un motor potente pero pequeño en el WSA que escanea previamente cada solicitud y determina el motor completo que se debe utilizar para escanear las solicitudes. Esto aumenta ligeramente el rendimiento del WSA.