Guia de design do Web Security Appliance

Contents

Introduction

Este documento descreve como projetar o Cisco Web Security Appliance (WSA) e os componentes associados para um desempenho ideal.

Informações de Apoio

Quando você projeta uma solução para o WSA, ela requer uma consideração cuidadosa, não apenas em relação à configuração do próprio dispositivo, mas também dos dispositivos de rede associados e seus recursos. Cada rede é uma colaboração de vários dispositivos e, se um deles não participar corretamente na rede, as experiências dos usuários podem cair.

Há dois componentes principais que devem ser considerados ao configurar o WSA: o hardware e o software. O hardware vem em dois tipos diferentes. O primeiro é o tipo físico de hardware, como os modelos S170, S380 e S680 Series, assim como outros modelos de fim da vida útil (EoL), como os modelos S160, S360, S660, S370 e S670 Series. O outro tipo de hardware é virtual, como os modelos S000v, S100v e S300v Series. O sistema operacional (SO) executado neste hardware é chamado de *AsyncOS para Web*, baseado no FreeBSD em seu núcleo.

O WSA oferece serviço proxy e também verifica, inspeciona e categoriza todo o tráfego (HTTP, HTTPS e FTP). Todos esses protocolos são executados sobre o TCP e dependem muito do Domain Name System (DNS) para uma operação adequada. Por esses motivos, a integridade da rede é vital para a operação adequada do dispositivo e sua comunicação com várias partes da rede, dentro e fora do controle da empresa.

Projeto

Use as informações descritas nesta seção para projetar o WSA e os componentes relacionados para um desempenho ideal.

Rede

Uma rede rápida e sem erros é vital para a operação adequada do WSA. Se a rede estiver instável, a experiência do usuário pode cair. Os problemas de rede geralmente são detectados quando as páginas da Web demoram mais para serem acessadas ou estão inacessíveis. A inclinação inicial é a culpa do dispositivo, mas geralmente é a rede que se comporta mal. Assim, deve-se considerar e auditar cuidadosamente para garantir que a rede ofereça o melhor serviço para protocolos de aplicativos de alto nível, como HTTP, HTTPS, FTP e DNS.

Considerações gerais

Aqui estão algumas considerações gerais que você pode implementar para garantir o melhor comportamento de rede:

- Verifique se a rede da camada 2 (L2) está estável, se a operação do spanning tree está correta e se não há cálculos frequentes do spanning tree e alterações de topologia.

- O protocolo de roteamento usado também deve fornecer convergência e estabilidade rápidas. Os temporizadores rápidos do OSPF (Open Shortest Path First) ou do EIGRP (Enhanced Interior Gateway Routing Protocol) são boas opções para tal rede.

- Sempre use pelo menos duas interfaces de dados no WSA: um que enfrenta os computadores do usuário final e outro para operação de saída (conectado ao proxy de upstream ou à Internet). Isso é feito para eliminar possíveis restrições de recursos, como quando o número de portas TCP está esgotado ou quando os buffers de rede ficam cheios (com o uso de uma única interface, especialmente para dentro e para fora).

- Dedicar a Interface de Gerenciamento para tráfego somente de gerenciamento para aumentar a segurança. Para conseguir isso por meio da GUI, navegue até **Rede > Interfaces** e marque a caixa de seleção **Separate routing (porta M1 restrita somente a serviços de gerenciamento de dispositivos)**.

- Use servidores DNS rápidos. Qualquer transação via WSA requer pelo menos uma pesquisa de DNS (se não estiver no cache). Um servidor DNS que é lento ou mal comportado afeta qualquer transação e é observado como conectividade de Internet lenta ou atrasada.

- Quando tabelas de roteamento separadas são usadas, essas regras se aplicam:

  Todas as interfaces estão incluídas na tabela de roteamento de *gerenciamento* padrão (M1, P1, P2).

  Somente as interfaces de dados são incluídas na tabela de roteamento *de dados*.

  Note: A separação das tabelas de roteamento não é por interface, mas por serviço. Por exemplo, o tráfego entre o WSA e o controlador de domínio do Microsoft Ative Diretory (AD) obedece sempre às rotas especificadas na tabela de roteamento de Gerenciamento, e é possível configurar rotas que apontam para fora da interface P1/P2 nesta tabela. Não é possível incluir rotas na tabela de roteamento de dados que usam as interfaces de gerenciamento.

Balanceamento de carga

Aqui estão algumas considerações de balanceamento de carga que você pode implementar para garantir o melhor comportamento da rede:

- Rotação de DNS - Este é o termo usado quando um único nome de host é usado como proxy, mas tem vários registros A no servidor DNS. Cada cliente resolve isso em um endereço IP diferente e usa proxies diferentes. Uma limitação é que as alterações nos registros DNS são refletidas nos clientes na reinicialização (cache DNS local), portanto, ele oferece um baixo nível de robustez se uma alteração tiver que ser feita. No entanto, isso é transparente para os usuários finais.

- Arquivos de Controle de Endereço de Proxy (PAC - Proxy Address Control) - são arquivos de script automáticos por proxy que determinam como cada URL deve ser tratado em um navegador com base nas funções gravadas dentro dele. Ele tem o recurso para encaminhar a mesma URL sempre diretamente ou para o mesmo proxy.

- Autodescoberta - descreve o uso de métodos DNS/DHCP para obter arquivos PAC (descritos na consideração anterior). Normalmente, essas três primeiras considerações são combinadas em uma solução. No entanto, isso pode ser complicado e muitos agentes de usuário, como Microsoft Office, Adobe Downloader, Javascripts e Flash, não podem ler arquivos PAC de forma alguma.

- Protocolo de Controle de Cache da Web (WCCP - Web Cache Control Protocol) - Esse protocolo (especialmente o WCCP versão 2) fornece uma maneira robusta e muito poderosa de criar balanceamento de carga entre vários WSAs e também incorporar alta disponibilidade.

- Dispositivo(s) de balanceamento de carga separado - A Cisco recomenda que você use balanceadores de carga como máquinas dedicadas.

Firewalls

Aqui estão algumas considerações de firewall que você pode implementar para garantir o melhor comportamento de rede:

- Certifique-se de que o Internet Control Message Protocol (ICMP) seja permitido em toda a rede a partir de cada origem. Isso é vital, pois o WSA depende do mecanismo de descoberta MTU (Maximum Transition Unit, Unidade Máxima de Transição) do caminho, conforme descrito no RFC 1191, que depende das solicitações de eco ICMP (tipo ⓿⓿ e respostas de eco (tipo 0), e a fragmentação inalcançável ICMP é necessária (tipo 3, código 4). Se você desabilitar a descoberta de MTU do caminho no WSA com o comando CLI **pathtudiscovery**, o WSA usará o MTU padrão de 576 bytes, conforme RFC 879. Isso afeta o desempenho devido ao aumento da sobrecarga e à remontagem de pacotes.

- Verifique se não há roteamento assimétrico dentro da rede. Embora isso não seja um problema no WSA, qualquer Firewall encontrado ao longo do caminho descarta os pacotes

porque não recebeu os dois lados da comunicação.

- Com os Firewalls, é muito importante excluir os endereços IP WSA das ameaças como estações de computadores finais regulares. O firewall pode bloquear
- os endereços IP do WSA devido a muitas conexões (de acordo com o conhecimento geral do Firewall).

- Se a Network Address Translation (NAT) for empregada para qualquer endereço IP WSA no dispositivo local do cliente, certifique-se de que cada WSA use um endereço global externo separado no NAT. Se você usar NAT para vários WSAs que têm um único endereço global externo, poderá encontrar estes problemas:

  Todas as conexões de todos os WSAs para o mundo externo usam um único endereço global externo, e o Firewall rapidamente fica sem recursos.

  Se houver um pico de tráfego em direção a esse único destino, o servidor de destino poderá bloqueá-lo e cortar todo o acesso da empresa a esse recurso. Esse pode ser um recurso valioso como o armazenamento em nuvem da empresa, as conexões em nuvem do Office ou as atualizações de software antivírus por computador.

Identidades

Lembre-se de que o princípio *lógico AND* se aplica a todos os componentes da identidade. Por exemplo, se você configurar o agente de usuário e o endereço IP, significa o agente de usuário *desse* endereço IP. Isso não significa o usuário-agente *ou* este endereço IP.

Use uma identidade para autenticação do mesmo tipo de substituto (ou sem substituto) e/ou agente de usuário.

Éimportante garantir que cada identidade que exige autenticação inclua as strings de usuário-agente para navegadores/agentes de usuário conhecidos que suportam autenticação de proxy, como Internet Explorer, Mozilla Firefox e Google Chrome. Existem alguns aplicativos que exigem acesso à Internet, mas que não suportam autenticação proxy/WWW.

As identidades são combinadas de cima para baixo com a pesquisa de correspondências que termina na primeira entrada correspondente. Por esse motivo, se você tiver a *Identidade 1* e *Identidade 2* configuradas e uma transação corresponder à Identidade 1, ela não será verificada em relação à Identidade 2.

Políticas de acesso/descriptografia/roteamento/malware de saída

Essas políticas são aplicadas a diferentes tipos de tráfego:

- As políticas de acesso são aplicadas em conexões HTTP ou FTP simples. Eles determinam se a transação deve ser aceita ou descartada.

- As políticas de descriptografia determinam se as transações HTTPS devem ser descriptografadas, descartadas ou passadas. Se a transação for descriptografada, a parte consecutiva poderá ser vista como uma solicitação HTTP simples e será comparada às

políticas de acesso. Se você precisar descartar uma solicitação HTTPS, solte-a nas políticas de descriptografia, não nas políticas de acesso. Caso contrário, ele consome mais CPU e memória para que uma transação perdida seja descriptografada e, em seguida, removida.

- As políticas de roteamento determinam a direção de upstream de uma transação quando ela é permitida pelo WSA. Isso se aplica se houver proxies de upstream ou se o WSA estiver no modo *Connector* e enviar tráfego para a torre do Cloud Web Security.

- As políticas de malware de saída são aplicadas a carregamentos HTTP ou FTP de usuários finais em direção a servidores Web. Isso é normalmente visto como uma solicitação HTTP Post.

Para cada tipo de política, é importante lembrar que se aplica o princípio *lógico OU*. Se você tiver várias identidades mencionadas, a transação deverá corresponder a qualquer uma das identidades configuradas.

Para um controle mais granular, use essas políticas. Identidades configuradas incorretamente por política podem criar problemas, onde é mais benéfico usar várias identidades referenciadas em uma política. Lembre-se de que as identidades não afetam o tráfego, elas apenas identificam os tipos de tráfego para correspondências posteriores em uma política.

Frequentemente, as políticas de descriptografia usam identidades com autenticação. Embora isso não esteja errado e às vezes seja necessário, o uso de uma identidade com autenticação referenciada na política de descriptografia significa que todas as transações que correspondem à política de descriptografia são descriptografadas para que a autenticação ocorra. A ação de descriptografia pode ser ignorada ou passada, mas como há uma identidade com autenticação, a descriptografia ocorre para posteriormente descartar ou passar pelo tráfego. Isso é caro e deve ser evitado.

Foram observadas algumas configurações que contêm 30 ou mais identidades e 30 ou mais políticas de acesso, nas quais todas as políticas de acesso incluem todas as identidades. Nesse caso, não há necessidade de usar essas várias identidades se elas forem combinadas em todas as políticas de acesso. Embora isso não prejudique a operação do dispositivo, ele cria confusão com tentativas de solução de problemas e é caro em relação ao desempenho.

Categorias de URL personalizadas

O uso de categorias de URL personalizadas é uma ferramenta poderosa no WSA que geralmente é mal compreendida e mal usada. Por exemplo, há configurações que contêm todos os sites de vídeo para correspondências na identidade. O WSA tem uma ferramenta incorporada que é atualizada automaticamente quando sites de vídeo mudam URLs, o que ocorre com frequência. Assim, faz sentido permitir que o WSA gerencie automaticamente as categorias de URL e use as categorias de URL personalizadas para sites especiais, ainda não categorizados.

Tenha muito cuidado com expressões regulares. Se forem usadas correspondências de caracteres especiais como ponto (.) e estrela(*), elas podem ser muito extensas para CPU e memória. O WSA expande qualquer expressão regular para corresponder a cada transação. Por exemplo, aqui está uma expressão regular:

```
example.*
```

Esta expressão corresponderá a qualquer URL que contenha a palavra *exemplo*, não apenas o

domínio *example.com*. Evite o uso de *ponto* e *estrela* em expressões regulares e use-os somente como último recurso.

Aqui está outro exemplo de uma expressão regular que pode criar problemas:

```
www.example.com
```

Se você usar este exemplo no campo Expressões regulares, ele não só corresponderá a *www.example.com*, como também a *www.www3example2com.com*, pois o ponto aqui significa *qualquer caractere*. Se você deseja corresponder apenas *www.example.com*, saia do ponto:

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

Nesse caso, não há motivo para usar o recurso Expressões regulares quando você pode incluir isso no domínio de categoria de URL personalizada com este formato:

```
www.example.com
```

Anti-malware e reputação

Se mais de um mecanismo de varredura estiver ativado, considere a opção de ativar também a varredura adaptável. A varredura adaptável é um mecanismo poderoso, mas pequeno, no WSA que faz a pré-varredura de cada solicitação e determina o mecanismo abrangente que deve ser usado para digitalizar solicitações. Isso aumenta um pouco o desempenho no WSA.