White paper

Uma breve história da zero trust: marcos importantes na reformulação da segurança corporativa

Por que contar a história da zero trust?

Em segurança de TI, muitos acreditam que zero trust é uma reformulação fundamental e revolucionária sobre a segurança corporativa e a proteção de redes e recursos que abrigam nossas melhores ideias, conectam nossos talentos mais brilhantes e concedem acesso a ferramentas de produtividade transformadoras.

Mas para entender como a estratégia zero trust é realmente revolucionária na segurança cibernética, é necessário entender as fraquezas da abordagem de segurança de redes legadas e como a ideia da arquitetura zero trust evoluiu para se tornar algo que remodela o pensamento de décadas de forma fundamental.

Redes 2D e segurança tipo castelo e fosso

'Em estrela' e 'castelo e fosso' são as duas metáforas principais usadas para descrever a arquitetura da rede legada e a segurança da rede, respectivamente. Adequadamente, as imagens usadas em ambas já existem há algum tempo.

A arquitetura de rede em estrela refere-se a redes satélites dispostas ao redor de um núcleo central. Esse modelo envolve rotear o tráfego interno e externo de volta por uma pilha de segurança em um data center principal antes de continuar para seu destino. Embora esta abordagem tenha funcionado por um tempo, ela se tornou mais complicada e cara devido à adoção da nuvem, à distribuição das equipes e à importância cada vez maior da mobilidade nos negócios.

Por outro lado, a segurança tipo castelo e fosso refere-se a redes autônomas projetadas para admitir tráfego amigável e ao mesmo tempo manter os inimigos firmemente fora de seus muros. Assim como um guarda no portão, os dispositivos de segurança internos servem para deixar entrar as pessoas certas e afastar os malfeitores. A enorme transição de aplicativos para a nuvem, juntamente com a migração de trabalhadores para fora dos perímetros corporativos, tornou essa abordagem obsoleta com mais rapidez do que balas de canhões para castelos reais.

As VPNs e o Wi-Fi complicaram ainda mais o problema. A antiga arquitetura castelo e fosso não dava aos administradores nenhuma maneira de conectar convidados a uma rede sem permitir que eles tivessem rédeas soltas enquanto estivessem por lá. Em última análise, não havia uma boa maneira de conectar terminais às redes sem alguma forma de segmentação para manter as redes seguras.

Precisávamos de algo melhor.

802.1X e os problemas com o NAC

Em 2001, a IEEE Standards Association publicou seu padrão de protocolo 802.1X para controle de acesso à rede (NAC).

> “Um meio de autenticar e autorizar dispositivos ligados a uma porta LAN com características de conexão ponto a ponto, e de impedir o acesso a essa porta nos casos em que o processo de autenticação e autorização falhar.”
>
> IEEE em 802.1X →

Logo depois, dispositivos sem fio começaram a incluir um cliente ou suplicante de 802.1X, que permitiu que as redes autenticassem o terminal antes de permitir a conexão. Esse avanço pretendia oferecer a capacidade de bloquear redes cabeadas e sem fio, para que somente dispositivos gerenciados e usuários autorizados pudessem se conectar. Pense no suplicante como fornecendo o ID para o segurança na porta da rede, que decide quem deixa entrar e quem fica de fora ao relento.

Infelizmente, o modelo NAC não foi uma solução milagrosa, e os problemas começaram com o fato de que redes internas com NAC eram projetadas com a confiança implícita em mente, e tentar fechar a autenticação/autorização após o fato causava um enorme esforço. Para que o NAC fosse totalmente eficaz, todas as portas acessíveis precisavam ser bloqueadas, mas nem todos os dispositivos eram compatíveis com 802.1X. A adoção cada vez maior de impressoras conectadas à Internet, leitores de crachás e outros dispositivos habilitados para redes deixou um grande buraco na segurança. Agora, imagine que nosso segurança ainda estava operando apenas uma porta de rede quando múltiplas (ou mesmo dezenas de) entradas alternativas estavam disponíveis.

Derrubando as mulharas de Jericó e reformulando o papel do perímetro na segurança

Em 2003, ficou claro que o uso de dispositivos pessoais continuaria a proliferar, e as empresas precisavam começar a pensar em como proteger máquinas que não estavam trancadas por trás das muralhas do castelo. Além disso, o uso crescente da criptografia estava reduzindo a eficácia de firewalls perimetrais, forçando uma escolha entre aumentar o dimensionamento para enfrentar os desafios de capacidade impostos por descriptografar e inspecionar, ou permitir que o tráfego criptografado passasse sem ser desafiado.

Naquele ano, um grupo multinacional de líderes tecnológicos europeus se reuniu para tratar de temas como autenticação de usuários, criptografia, gerenciamento de identidades e aplicação de políticas. Após se estabelecer formalmente em 2004, o Fórum Jericó apresentou ao mundo a noção de "desperimetrização".

Com um nome lembrando a história bíblica dos israelitas derrubando as muralhas da antiga cidade de Jericó, o fórum se concentrou em resolver o problema de como "garantir fluxos de informações seguros e sem limites entre as empresas".

Além da metáfora apropriada, o grupo deixou de herança Os Mandamentos do Fórum Jericó, o mais próximo que chegamos até agora das verdades do alto sobre como reger redes sem perímetro. Infelizmente, o conjunto de controles e mitigações prescritas estava além da capacidade da maioria das empresas de implantar ou administrar naquele momento.

O termo "zero trust" estreia no léxico de TI

Em 2010, o analista da Forrester, John Kindervag, publicou um artigo intitulado "No More Chewy

Centers: Apresentando o Modelo Zero Trust de Segurança da Informação" e, rapidamente, tínhamos uma nova palavra da moda representando uma nova forma de pensar sobre segurança de redes. Uma afirmação fundamental no artigo foi que a mera presença em uma rede não era suficiente para conceder confiança.

"Foi aí que começamos a ouvir coisas como, 'a identidade é o novo perímetro'," afirma Lisa Lorenzin, Zscaler Field CTO e veterana da zero trust. "Autenticávamos um usuário e usávamos sua identidade para determinar o que ele podia fazer. Talvez, com sorte, conseguíamos reunir algum contexto, como, por exemplo, se tínhamos um dispositivo gerenciado ou não gerenciado, e tomávamos decisões sobre acesso com base nesse entendimento rudimentar."

Progresso. Mas isso deixou a segurança corporativa presa na proteção das próprias redes. Ainda não estávamos prontos para abandonar isso totalmente. Estávamos aquém de uma abordagem transformacional, por isso a adoção desses princípios voltou a fracassar. Por exemplo, ainda dependíamos do mesmo conjunto de ferramentas focado na rede: 802.1X e RADIUS na Camada 2, firewalls com acesso por identidade na Camada 3, etc.

O novo caminho era apenas NAC com um nome chamativo.

BeyondCorp (no perímetro)

Enquanto isso, hackers ligados ao Exército de Libertação Popular da China (PLA) estavam fazendo com que os melhores especialistas do setor de tecnologia reconsiderassem totalmente a questão da confiança. Em 2010, o Google divulgou uma operação de 2009 que tinha como alvo essa e várias outras empresas de alta tecnologia, incluindo Akamai, Adobe e Juniper Networks. A campanha foi chamada de "Operação Aurora" pelos pesquisadores de segurança da McAfee.

Ao dar um pontapé no ninho de talentos da elite da engenharia de TI, os hackers chineses involuntariamente aceleraram o desenvolvimento da arquitetura zero trust nos principais laboratórios de tecnologia do país. O Google desenvolveu o BeyondCorp em resposta à Operação Aurora, que se concentrou em **"deslocar controles de acesso do perímetro da rede para usuários individuais...[permitindo] o trabalho seguro de praticamente qualquer local sem a necessidade de uma VPN tradicional."**

Porém, "o Google é uma empresa dirigida por engenheiros, para engenheiros, com um orçamento realmente infinito, e comparativamente pouca infraestrutura legada em comparação com muitas empresas", diz Lorenzin. "E, mesmo assim, ainda foram necessários sete anos e seis white papers de design e implementação."

Mesmo com o exemplo bem documentado do Google, a verdadeira arquitetura zero trust ainda estava fora do alcance da maioria das empresas. Apesar de tentar "preparar o caminho para que outras empresas fizessem sua própria implementação de uma rede Zero Trust", o futuro que o Google imaginava ainda estava bem longe.

Enquanto isso, para os usuários, a popularidade da nuvem e a ênfase contínua na mobilidade significava que mais dados estavam disponíveis e eram acessados de fora do perímetro da rede, em vez de dentro dele. A necessidade de uma abordagem generalizada à confiança era maior do que nunca.

O Gartner e a chegada definitiva do acesso à rede zero trust

A empresa de pesquisas tecnológicas Gartner foi responsável pelos próximos avanços significativos da zero trust como uma estrutura amplamente adaptável. Embora já existisse, o termo "zero trust" não estava no topo das atenções em 2010, quando a empresa lançou sua Avaliação Contínua de Risco Adaptativo e de Confiança (CARTA, na sigla em inglês).

O documento descrevia a necessidade de entender quem está solicitando acesso e conceder esse acesso com base em uma avaliação dinâmica do ambiente, do contexto disponível e das justificativas quanto às responsabilidades do usuário.

Lorenzin descreve a CARTA como "um grande modelo que nunca teve a atenção que merecia".

No Gartner, a CARTA eventualmente se transformou em "Acesso à Rede Zero Trust" (ZTNA) após a estrutura original não ter conseguido ganhar notoriedade entre os profissionais de tecnologia (observe o foco persistente nas redes como alvo do acesso!). Mas, fundamentalmente, a CARTA continua sendo importante para a história da zero trust, pois os princípios que ela estabeleceu vivem sob a forma do ZTNA.

A próxima contribuição significativa do Gartner para essa discussão veio com o reconhecimento de que estava havendo uma convergência entre os campos de rede e segurança. Em 2019, esse casamento foi manifestado com o Secure Access Service Edge (SASE). Porém, foi uma união de curta duração, e, em 2021, as categorias se dividiram novamente, introduzindo a categoria de mercado Secure Service Edge (SSE): SASE sem WAN.

Não importa o nome, o Gartner já tinha se estabelecido como um árbitro importante quanto ao significado de zero trust. Agora, os fornecedores estavam se mexendo para se adequarem devidamente a uma de suas novas categorias de mercado.

"O Homem" entra no bate-papo: NIST, OMB e o aval do governo ao ZTA

Em 2020, o National Institute for Standards and Technology (NIST) reestruturou o debate com a norma NIST 800-207 para arquitetura zero trust. Esse novo paradigma de segurança cibernética focava na proteção de recursos e na premissa de que a confiança nunca deve ser concedida de forma implícita, mas deve ser continuamente avaliada.

Com esse documento, as correntes do perímetro e da rede privada virtual foram finalmente descartadas. O foco mudou de proteger a rede para proteger os usuários, os dados e os aplicativos que interagem pela rede. Zero trust agora significava basicamente o acesso baseado no contexto e de privilégio mínimo, aplicável em uma variedade muito maior de casos de uso e fluxos de tráfego.

A norma 800-207 estipula os principais princípios e suposições para a zero trust. Três dos pontos mais cruciais (de uma lista muito mais longa) são:

1. Nenhum recurso é inerentemente confiável.
2. Toda comunicação é protegida, não importa a localização da rede. Encerrar e inspecionar a solicitação; observar todo o conteúdo disponível associado ao usuário e à solicitação.
3. O processo de autenticação e autorização de recursos é dinâmico e rigorosamente aplicado antes de permitir o acesso.

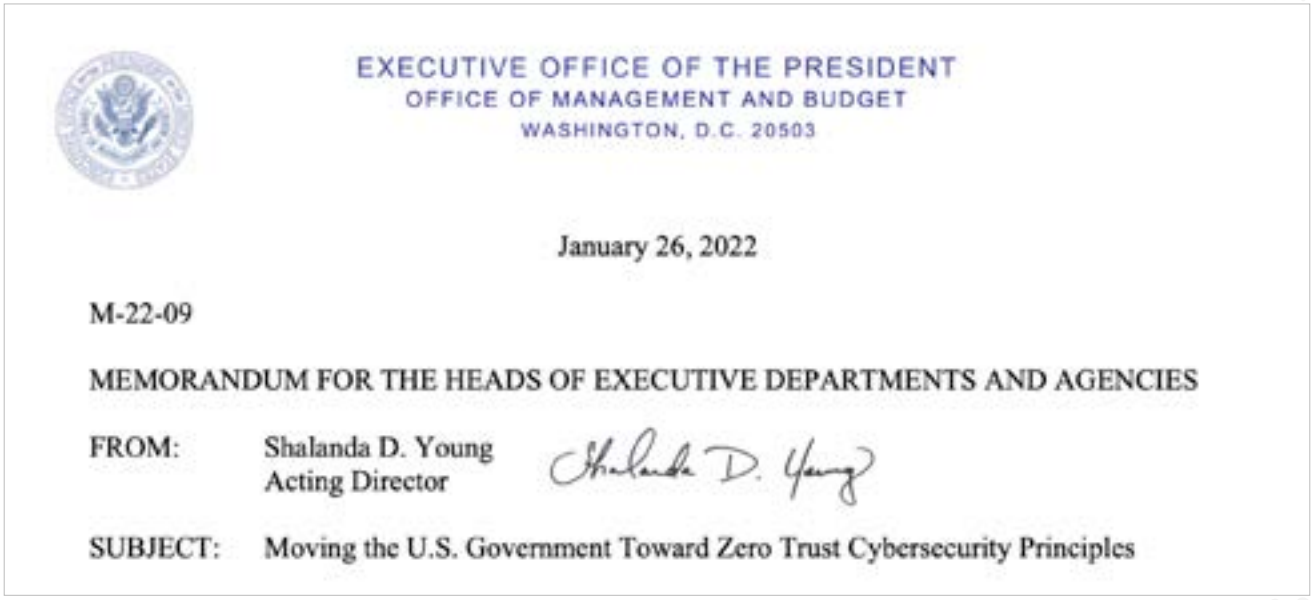
Mas o verdadeiro ponto sem volta para a promoção de princípios da zero trust veio do topo, pelo menos nos Estados Unidos. O U.S. Office of Management and Budget, o escritório responsável pela implementação das políticas presidenciais, divulgou sua diretiva M-22-09 em 2022, afirmando que todos os escritórios do governo federal devem adotar princípios de arquitetura zero trust até 2024 e delinear marcos claros e datas previstas ao longo do caminho.

"Até agora, tivemos documentos de orientação. Tivemos modelos de administradores. Mas esse foi o ponto em que a coisa ficou séria com a estratégia zero trust federal", segundo Lorenzin.

O ataque à cadeia de fornecimento contra a plataforma de gerenciamento de TI Solar Winds — divulgado em 2021 e responsável pelo comprometimento de pelo menos nove agências federais, incluindo Estado, Tesouro, Segurança Nacional, Comércio e Energia — foi talvez o ataque mais descarado e prejudicial patrocinado pelo Estado desde a Operação Aurora. Em resposta, o governo federal apostou suas fichas na zero trust, adotando essa abordagem como estrela-guia da segurança cibernética para os anos a seguir.

Como implementar a zero trust

A abordagem da Zscaler quanto à arquitetura zero trust está bem alinhada à estrutura ZTA do NIST e a definição do Gartner para SSE. Mas ela vai além de qualquer padrão, com seu compromisso com três avanços fundamentais no pensamento zero trust. Juntos, esses três princípios avançados ajudam a levar a algumas conclusões lógicas quanto à aplicação da zero trust.



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

January 26, 2022

M-22-09

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Shalanda D. Young
Acting Director

SUBJECT: Moving the U.S. Government Toward Zero Trust Cybersecurity Principles

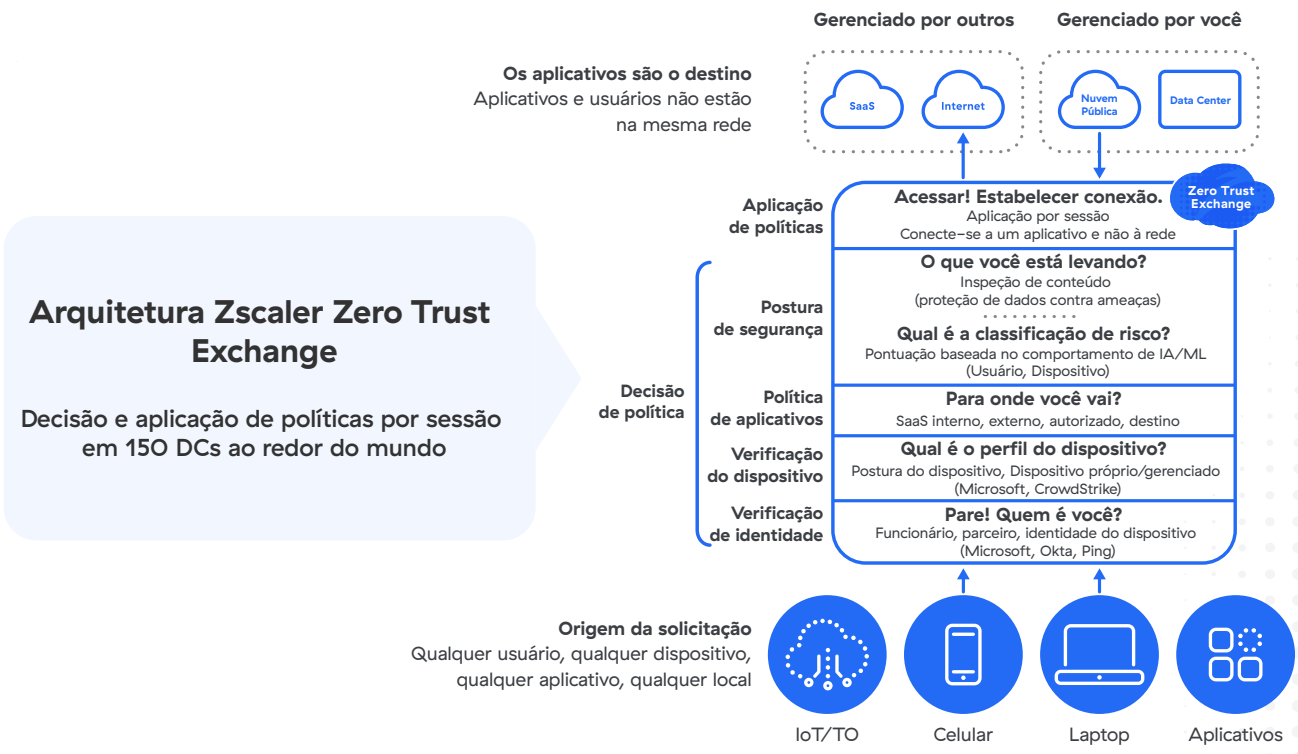
Todo o tráfego é tráfego zero trust

A estratégia zero trust começou como uma nova maneira de proteger as redes. Eventualmente, se ampliou além das redes locais, mas ainda estava focada principalmente no tráfego de aplicativos privados. Por muito tempo,

o tráfego foi considerado conforme sua relação com uma rede, em vez de afastar completamente a rede.

Mas agora sabemos que os princípios da zero trust podem ser aplicados para proteger aplicativos SaaS, tráfego de e para nuvens públicas e até mesmo usuários no acesso à Internet pública. E os originadores desse tráfego podem ser cargas de trabalho e também usuários. O acesso pode ser feito independentemente do transporte, com o tráfego fluindo por qualquer roteador e passando por qualquer rede, com ou sem fio, 4G ou 5G, e assim por diante.

Já passou da hora de aplicar os princípios zero trust a todo o tráfego, não importa a origem, não importa o destino. Já eliminamos as distinções entre confiável e não confiável, dentro da rede ou fora dela. Agora, é hora de parar de pensar sobre qual entidade está se conectando a qual rede, e usar a zero trust para conectar todas as entidades de forma direta usando políticas corporativas. A Internet é a nova rede corporativa e todo o tráfego é um alvo fácil.



1 Identidade e contexto sempre vêm antes de conectividade

A verificação da identidade está na essência da zero trust. Mas, no passado, confundimos identidade com conectividade, e isso nos levou a modelos rompidos. Endereços IP, endereços MAC, e porta e protocolo não são identidade.

Dispositivos TO podem se conectar às redes a partir de fábricas. Os usuários podem se conectar a partir de cafeterias. Mas isso não significa que sabemos algo sobre eles. Então, precisamos começar com identidade e contexto. Só a partir daí podemos autorizar a conexão.

Quando um usuário solicita acesso a um recurso, devemos primeiro considerar quem ele é, outras informações sobre ele, como função ou departamento, o dispositivo sendo usado e, em seguida, as políticas de segurança. O que o usuário está tentando fazer? Para onde ele vai? O que no ambiente pode contribuir para nossa decisão de permitir ou negar a ação?

O contexto vai além da identidade e é avaliado de forma contínua. Outros fatores que podem ser cruzados para verificar anomalias incluem geolocalização, endereço IP, postura do dispositivo e hora do dia. E uma solução zero trust deve ser capaz de descriptografar o tráfego, inspecionar ameaças e riscos de exfiltração de dados em linha e em escala.

No caso da Zero Trust Exchange, também correlacionamos a inteligência de ameaças — de toda nossa nuvem global, bem como de parceiros tecnológicos terceirizados, como fornecedores de segurança e verificação de identidade — para determinar riscos e tomar decisões sobre políticas e acesso.

2 Aplicativos – e até mesmo ambientes de aplicativos – devem permanecer invisíveis para usuários não autorizados

Agora que resolvemos o problema de saber quem você é antes de conceder o acesso, podemos enfrentar o próximo desafio: como conectamos você aos seus recursos autorizados, reduzindo ao mesmo tempo o risco e o potencial de comprometimento? Assim que reunimos e analisamos o contexto que envolve um usuário, dispositivo, política e ambiente, podemos dar os próximos passos nessa direção.

Ao eliminar o interlocutor de entrada para conexões remotas, eliminamos a superfície de ataque externa. Caso contrário, fica simplesmente fácil demais para invasores localizarem gateways de VPN vulneráveis ou aplicativos expostos para comprometer alvos. VPNs à espera de conexões de entrada são presas fáceis, e os agentes das ameaças percebem isso. Esse é um problema independente do fornecedor e que pode ser resolvido mudando o modelo da arquitetura.

A Zscaler Zero Trust Exchange faz isso formando conexões somente de saída, tanto do usuário como do ambiente do aplicativo para nossa nuvem de segurança, usando microtúneis criptografados para conexões de agentes entre as solicitações e seus destinos.

Esse "terceiro lugar" online fornece um buffer entre os usuários verificados e qualquer recurso que tenham autorização de acessar. Assim que o usuário é conectado ao ativo solicitado, políticas granulares garantem que não exista a opção de se aventurar além dele. O movimento lateral se torna essencialmente impossível.

3 O capítulo final?

Os princípios discutidos acima nos permitem finalmente superar de forma real o entendimento antigo sobre perímetros de rede protegidos por firewalls e terminais remotos conectados por redes privadas virtuais. Eles não apenas replicam os controles de segurança existentes em uma instância virtual hospedada na nuvem, ou confiam em algum entendimento artificial sobre o que está na rede contra o que não está.

Uma arquitetura abrangente criada para oferecer segurança zero trust — para usuários, cargas de trabalho, aplicativos, dispositivos TO e IoT, e muito mais — reduz o risco, melhora a proteção, simplifica a experiência do usuário e representa uma melhoria fundamental na maneira como formulamos a segurança corporativa.

+1 408.533.0288 Zscaler, Inc. (Sede) • 120 Holger Way • San Jose, CA 95134 zscaler.com