zscaler™

Documento técnico

Un breve historial de la confianza cero: los principales hitos a la hora de replantearse la seguridad empresarial

¿Por qué contar la historia de la confianza cero?

Muchos de quienes trabajan en seguridad informática creen que la confianza cero es un punto de inflexión, un replanteamiento esencial de la seguridad empresarial y la protección de las redes y los recursos que albergan nuestras mejores ideas, conectan nuestro talento más brillante y dan acceso a herramientas de productividad transformadoras.

Pero para entender lo verdaderamente revolucionario que es el modelo de confianza cero en la ciberseguridad, es necesario comprender los puntos débiles del enfoque de seguridad de red heredado y cómo la idea de la arquitectura de confianza cero evolucionó hasta convertirse en una idea que transforma fundamentalmente el pensamiento reinante durante décadas.

Redes 2D y seguridad de castillo y foso

La arquitectura radial y el castillo y foso son las dos metáforas principales utilizadas para describir la arquitectura de red heredada y la seguridad de la red, respectivamente. Acertadamente, las imágenes utilizadas en ambos llevan en uso bastante tiempo.

Al hablar de arquitectura radial, nos referimos a las redes de satélites dispuestas alrededor de un hub central. Este modelo implica enrutar el tráfico interno y externo a través de una pila de seguridad en un centro de datos principal antes de que llegue a su destino. Aunque este enfoque ha funcionado durante un tiempo, se ha vuelto más complicado y costoso como consecuencia de la adopción de la nube, la distribución de la fuerza laboral y la creciente importancia de la movilidad en el negocio.

Por otro lado, la seguridad del modelo de castillo y foso se refiere a redes autónomas diseñadas para admitir el tráfico amistoso y mantener a los enemigos firmemente fuera de sus muros. Igual que un guardia en la puerta, los dispositivos de seguridad internos están pensados para permitir que las personas adecuadas entren y para alejar a los forajidos. La transición masiva de las aplicaciones a la nube, junto con la migración de los trabajadores fuera de los perímetros corporativos, hizo que este enfoque quedara obsoleto más rápidamente de lo que las balas de cañón destruían los castillos reales.

Las VPN y el wifi complican aún más el problema. La arquitectura antigua de castillo y foso no daba a los administradores ninguna forma de conectar a los huéspedes a una red sin permitirles moverse libremente mientras estaban allí. En última instancia, no había una buena manera de conectar los puntos finales a las redes sin alguna forma de segmentación para mantener estas últimas seguras.

Necesitábamos algo mejor.

802.1X y los problemas con NAC

En 2001, la Asociación de Estándares IEEE publicó su estándar de protocolo 802.1X para el control de acceso a la red (NAC).

> "Un medio para autenticar y autorizar dispositivos conectados a un puerto LAN que tiene características de conexión punto a punto y para evitar el acceso a ese puerto en los casos en los que falle el proceso de autenticación y autorización."
>
> IEEE en 802.1X arrow-right

Poco después, los dispositivos inalámbricos empezaron a incluir un suplicante (o cliente) 802.1X, que permitía a las redes autenticar el terminal antes de permitir una conexión. Este avance pretendía ofrecer la posibilidad de bloquear las redes por cable e inalámbricas, de modo que solo pudieran conectarse los dispositivos gestionados y los usuarios autorizados. Piense en el suplicante como si proporcionara una identificación al portero en la puerta de la red que decide a quién se deja entrar y a quién se deja fuera.

Desafortunadamente, el modelo NAC no era la panacea y los problemas comenzaron al observar que las redes internas N. estaban diseñadas teniendo en cuenta la confianza implícita; tratar de aplicar la autenticación/autorización después del hecho suponía un gran esfuerzo. Para que el NAC fuera totalmente eficaz, era necesario bloquear todos los puertos accesibles, pero no todos los dispositivos eran compatibles con 802.1X. La creciente adopción de impresoras conectadas a Internet, lectores de tarjetas de identificación y otros dispositivos conectados a la red era un problema de seguridad flagrante. Ahora bien, imaginemos que nuestro portero sigue vigilando solo esa puerta de la red cuando hay múltiples (o incluso docenas) de entradas alternativas disponibles.

Derribar el Muro de Jericó y replantearse el papel del perímetro en la seguridad

En 2003, estaba claro que el uso de dispositivos personales seguiría en aumento y las organizaciones necesitaban empezar a pensar en cómo proteger las máquinas que no estaban guardadas tras los muros del castillo. Además, el creciente uso del cifrado estaba reduciendo la eficacia de los cortafuegos perimetrales, lo que obligaba a elegir entre aumentar la escala para hacer frente a los retos de capacidad impuestos por el descifrado y la inspección, o permitir que el tráfico encriptado pasara sin problemas.

Ese año, un grupo multinacional de líderes tecnológicos europeos se reunió para abordar temas como la autenticación de usuarios, el cifrado, la gestión de identidades y la aplicación de políticas. Tras establecerse formalmente en 2004, el Foro de Jericó presentó la noción de "desperimetrización" al mundo.

Con un nombre que recuerda a la historia bíblica de los israelitas derribando los muros de la antigua ciudad de Jericó, el foro se propuso resolver el problema de cómo "permitir flujos de información seguros y sin fronteras entre empresas".

Además de la acertada metáfora, el grupo creó Los mandamientos del Foro de Jericó, lo más cercano que habíamos llegado hasta ese momento a las verdades acerca del gobierno de las redes sin perímetro. Lamentablemente, el conjunto de controles y mitigaciones prescritas estaba más allá de las capacidades de despliegue o administración de la mayoría de las empresas en aquel momento.

La "confianza cero" entra por primera vez en el léxico informático

En 2010, el analista de Forrester John Kindervag publicó un artículo titulado "No más centros

problemáticos: presentamos el modelo de confianza cero de seguridad de la información" y enseguida tuvimos una nueva palabra de moda que representaba una nueva forma de pensar en la seguridad de la red. Una afirmación clave del documento era que la mera presencia en una red no era suficiente para otorgar confianza.

"Ahí es cuando empezamos a escuchar cosas como 'la identidad es el nuevo perímetro'", dice la directora de tecnología de campo de Zscaler y veterana en el campo de la confianza cero Lisa Lorenzin. "Autenticábamos a un usuario y utilizábamos esa identidad para determinar lo que podía hacer. Tal vez, si teníamos suerte, podíamos reunir algo de contexto, como si era un dispositivo gestionado o no gestionado, y así tomábamos decisiones sobre el acceso basadas en esa comprensión rudimentaria".

Progreso. Pero esto dejó la seguridad empresarial atascada en proteger las propias redes. Todavía no estaba preparada para abandonarlas del todo. Seguíamos sin llegar a un enfoque transformador, por lo que la adopción de estos principios volvió a fracasar. Por un lado, aún dependía del mismo conjunto de herramientas centradas en la red: 802.1X y RADIUS en la capa 2, cortafuegos sensibles a la identidad en la capa 3, etc.

La nueva forma era realmente lo mismo que NAC pero con un nombre pegadizo.

Más allá (del perímetro) Corp

Mientras tanto, los piratas informáticos vinculados al Ejército Popular de Liberación (EPL) de China estaban provocando que las mejores y más brillantes mentes del sector tecnológico se replantearan por completo la cuestión de la confianza. En 2010, Google desveló una operación de 2009 que tenía como objetivo al propio Google y a otras empresas tecnológicas de alto nivel, como Akamai, Adobe y Juniper Networks. Los investigadores de seguridad de McAfee bautizaron la campaña como "Operación Aurora".

Al revolver el avispero del talento de ingeniería de TI de élite, los hackers chinos aceleraron involuntariamente el trabajo en una arquitectura de confianza cero en los laboratorios de tecnología más importantes del país. Google desarrolló BeyondCorp en respuesta a la Operación Aurora, que se centró en **"cambiar los controles de acceso del perímetro de la red a usuarios individuales... [habilitando] el trabajo seguro desde prácticamente cualquier lugar sin la necesidad de una VPN tradicional".**

Pero "Google es una empresa dirigida por ingenieros, para ingenieros, con, efectivamente, un presupuesto infinito y una infraestructura heredada comparativamente pequeña en contraposición con muchas empresas", afirma Lorenzin. "Y aun así les llevó siete años y seis documentos técnicos de diseño e implementación".

Incluso con el ejemplo bien documentado de Google, la auténtica arquitectura de confianza cero aún estaba fuera del alcance de la mayoría de las empresas. A pesar de tratar de "allanar el camino para que otras organizaciones pusieran en marcha su propia implementación de una red de confianza cero", el futuro imaginado por Google estaba aún a gran distancia.

Mientras tanto, para los usuarios, la popularidad de la nube y el continuo énfasis en la movilidad significaron que había más datos disponibles y a los que se accedía desde fuera del perímetro de la red que dentro de ella. La necesidad de un enfoque generalizado de la confianza era mayor que nunca.

Gartner y la llegada final del acceso a la red de confianza cero

La empresa de investigación tecnológica Gartner fue la responsable de los siguientes avances significativos de la confianza cero como marco ampliamente adaptable. A pesar de que ya existía, el término "confianza cero" no se convirtió en término prioritario hasta 2010, cuando la empresa lanzó su evaluación adaptativa continua de riesgos y confianza (CARTA).

El documento describía la necesidad de comprender quién solicita acceso y conceder ese acceso en función de una evaluación dinámica del entorno, el contexto disponible y las responsabilidades de un usuario.

Lorenzin describe CARTA como "un gran modelo que nunca tuvo la tracción que merecía".

En Gartner, CARTA finalmente se transformó en “Zero Trust Network Access" (ZTNA) después de que el marco original no lograra ganar conciencia entre los profesionales de la tecnología (obsérvese el enfoque persistente en las redes como objetivo de acceso). Sin embargo, CARTA sigue siendo importante para la historia de la confianza cero porque los principios que estableció siguen vivos en forma de ZTNA.

La siguiente contribución significativa de Gartner a este debate vino con el reconocimiento de que los campos de las redes y la seguridad estaban convergiendo. En 2019, puso de manifiesto este vínculo presentando Secure Access Service Edge (SASE). No obstante, fue una unión efímera y en 2021 volvió a dividir las categorías al introducir la categoría de mercado Secure Service Edge (SSE): SASE sin WAN.

Independientemente de cómo se denominara, para ese momento Gartner se había establecido como árbitro significativo de lo que constituía o no constituía la confianza cero. Los vendedores ya estaban luchando por encajar en una de sus nuevas categorías de mercado.

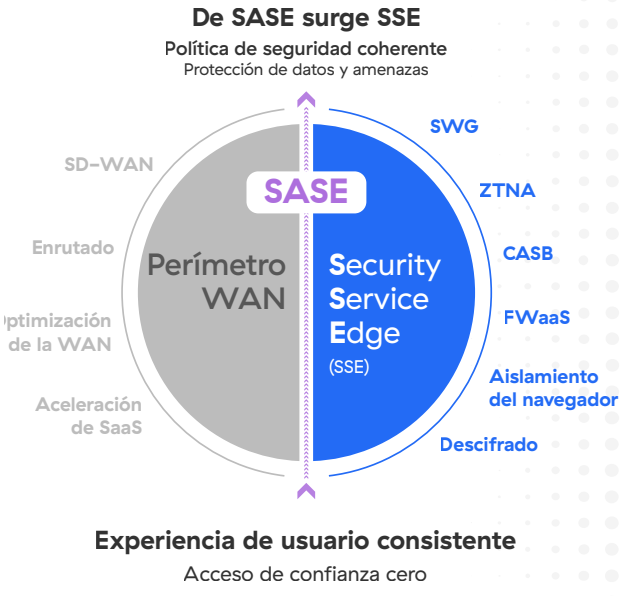
"El hombre" entra en el chat: NIST, OMB y el respaldo del gobierno a la confianza cero

En 2020, el Instituto Nacional de Estándares y Tecnología (NIST) restableció los parámetros con su norma NIST 800-207 para la arquitectura de confianza cero. Este nuevo paradigma de ciberseguridad se centraba en la protección de los recursos y en la premisa de que la confianza nunca debe concederse de forma implícita, sino que debe evaluarse continuamente.

Con este documento, los límites del perímetro y de la red privada virtual fueron finalmente eliminados. El enfoque pasó de proteger la red a proteger a los usuarios, los datos y las aplicaciones que interactúan a través de la red. Ahora, confianza cero significaba un acceso simplemente basado en contexto y con menos privilegios, aplicable a una variedad mucho más amplia de casos de uso y flujos de tráfico.

La norma 800-207 estipula los principios y supuestos clave para la confianza cero. Tres de los puntos más importantes (de una lista mucho más larga) son:

1. Ningún recurso es inherentemente confiable.
2. Todas las comunicaciones están protegidas independientemente de la ubicación de la red. Cancelar e inspeccionar la solicitud; mirar todo el contexto disponible asociado con el usuario y la solicitud.
3. Todas las autenticaciones y autorizaciones de recursos on dinámicas y se aplican estrictamente antes de permitir el acceso.



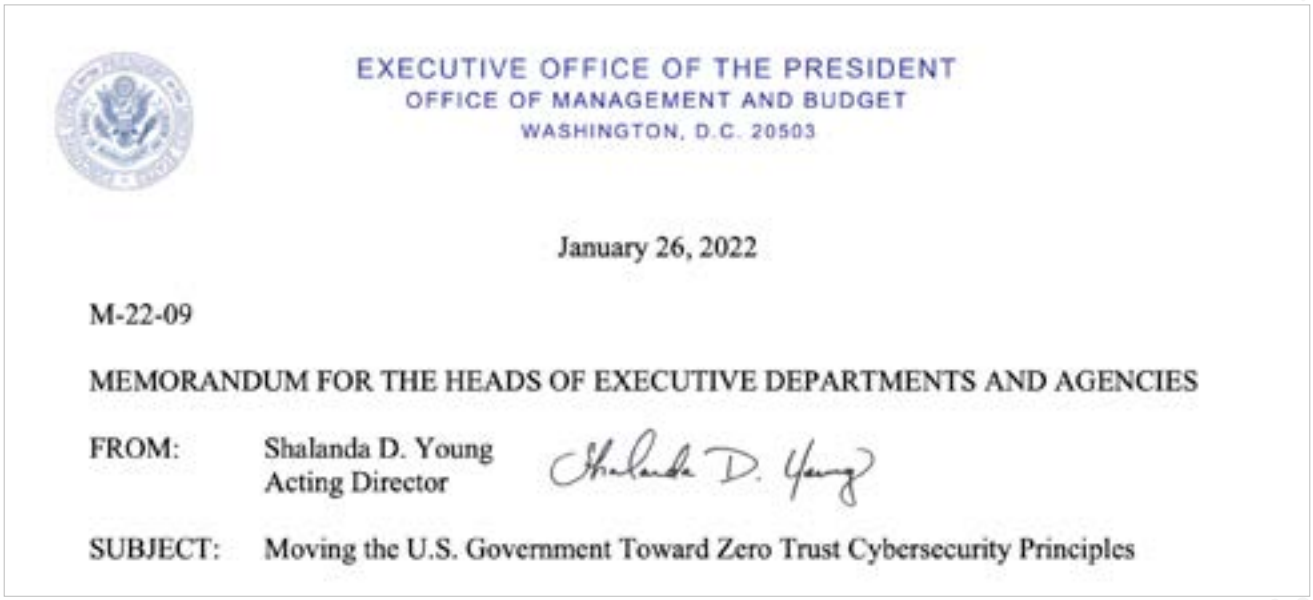
Pero el verdadero punto de no retorno para la promoción de los principios de confianza cero vino de lo más alto, al menos en Estados Unidos. La Oficina de Gestión y Presupuesto de Estados Unidos (OMB, por sus siglas en inglés), responsable de la aplicación de políticas presidenciales, publicó su directiva M-22-09 en 2022, en la que se establece que todas las oficinas del gobierno federal deben adoptar los principios de la arquitectura de confianza cero antes de 2024 y se esbozan hitos claros y fechas límite en el camino.

"Hasta ahora, hemos tenido documentos de guía. Hemos tenido modelos de administrador. Pero este es el primer punto de auténtica puesta en marcha, con la estrategia de confianza cero federal", según Lorenzin.

El ataque a la cadena de suministro contra la plataforma de gestión de TI Solar Winds (divulgado en 2021 y responsable de comprometer a al menos nueve agencias federales, incluidas Estado, Tesoro, Seguridad Nacional, Comercio y Energía) fue quizás el ataque más osado y dañino dirigido al Estado desde la Operación Aurora. En respuesta, el gobierno federal se ha decantado claramente por la confianza cero, adoptando ese enfoque como su lema de ciberseguridad para los próximos años.

Implantar la confianza cero

El enfoque de Zscaler para la arquitectura de confianza cero se alinea estrechamente con el marco ZTA del NIST y la definición de Gartner para SSE. Pero va más allá de cualquier norma gracias a su compromiso con tres avances fundamentales en el pensamiento de confianza cero. En conjunto, estos principios avanzados ayudan a llevar la aplicación de la confianza cero a algunas conclusiones lógicas.



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 el tráfico es de confianza cero

La confianza cero comenzó como una forma novedosa de proteger las redes. Con el tiempo, se expandió más allá de las redes locales, pero seguía centrándose principalmente en el tráfico de aplicaciones privadas.

Durante demasiado tiempo se consideró el tráfico en función de su relación con una red, en lugar de eliminar la red por completo.

Pero ahora sabemos que se pueden aplicar principios de confianza cero para proteger las aplicaciones SaaS, el tráfico hacia y desde las nubes públicas, e incluso a los usuarios cuando acceden a la Internet pública. Y el origen de ese tráfico pueden ser cargas de trabajo, así como usuarios. Se puede acceder independientemente del transporte, con tráfico fluyendo a través de cualquier enrutador y llegando a través de cualquier red, por cable o inalámbrica, 4G o 5G, etc.

Ya es hora de aplicar los principios de confianza cero a todo el tráfico, independientemente de su origen y su destino. Ya no vamos a aplicar distinciones entre confiable y no confiable, dentro o fuera de la red. Ahora ha llegado el momento de dejar de pensar en qué entidad se está conectando a qué red y, en su lugar, utilizar la confianza cero para conectar a todas las entidades directamente utilizando políticas empresariales. Internet es la nueva red corporativa y todo el tráfico es presa fácil.

1 La identidad y el contexto siempre vienen antes de la conectividad

La verificación de identidad se encuentra en el eje de la confianza cero. Pero en el pasado confundimos la identidad con la conectividad y eso nos ha llevado a modelos rotos. Las direcciones IP, las direcciones MAC y el puerto y protocolo no son identidad.

Los dispositivos OT pueden conectarse a redes de fábricas. Los usuarios pueden iniciar sesión en cafeterías. Pero eso no significa que sepamos cosas sobre ellos. Así que tenemos que empezar con la identidad y el contexto. Solo a partir de ahí podemos autorizar la conectividad.

Cuando un usuario solicita acceso a un recurso, primero debemos considerar quién es, otros datos sobre él, como su rol o el departamento, el dispositivo que está utilizando y, a continuación, las políticas de seguridad. ¿Qué intenta hacer el usuario? ¿A dónde se dirige? ¿Qué elementos del entorno pueden contribuir a nuestra decisión de permitir o denegar la acción?

El contexto va más allá de la identidad y se evalúa continuamente. Otros factores que se pueden verificar para detectar anomalías son la geolocalización, la dirección IP, la postura del dispositivo y la hora del día. Y una solución de confianza cero debe ser capaz de descifrar el tráfico para inspeccionar las amenazas y los riesgos de exfiltración de datos en línea y a escala.

En el caso de Zero Trust Exchange, también correlacionamos la información sobre amenazas (procedente de nuestra nube global, así como de socios tecnológicos de terceros, como proveedores de seguridad y verificación de identidad) para determinar el riesgo y tomar decisiones sobre políticas y acceso.

2 Las aplicaciones, e incluso los entornos de aplicaciones, deben permanecer invisibles para los usuarios no autorizados

Ahora que hemos resuelto el problema de saber quién es usted antes de concederle acceso, podemos abordar el siguiente desafío: ¿cómo lo conectamos con los recursos a los que está autorizado, a la vez que reducimos el riesgo y minimizamos el potencial de riesgo? Una vez que se ha recopilado y analizado el contexto de un usuario, dispositivo, política y entorno, podemos dar los siguientes pasos en esa dirección.

Al eliminar el oyente entrante para las conexiones remotas, eliminamos la superficie de ataque externa. De lo contrario, es demasiado fácil para los atacantes localizar puertas de enlace VPN vulnerables o aplicaciones expuestas para comprometer los objetivos. Las VPN que esperan conexiones entrantes son un blanco fácil y los autores de amenazas se dan cuenta de ello. Se trata de un problema que no tiene que ver con el proveedor, que solo se puede resolver cambiando el modelo arquitectónico.

Zscaler Zero Trust Exchange hace esto formando conexiones solo de salida tanto desde el usuario como desde el entorno de la aplicación hacia nuestra nube de seguridad utilizando microtúneles encriptados para intermediar las conexiones entre las solicitudes y sus destinos.

Este último paso en línea proporciona un búfer entre los usuarios verificados y cualquier recurso al que estén autorizados a acceder. Una vez que un usuario está conectado al activo solicitado, las políticas granulares garantizan que no haya opción para aventurarse más allá de dicho activo. El movimiento lateral se vuelve esencialmente imposible.

3 ¿El capítulo final?

Los principios mencionados anteriormente nos permiten superar verdadera y finalmente el concepto heredado de los perímetros de red protegidos por cortafuegos y puntos finales remotos conectados a través de redes privadas virtuales. No solo replican los controles de seguridad existentes en una instancia virtual alojada en la nube, sino que también dependen de cierta comprensión artificial de lo que hay en la red frente a lo que no.

Una arquitectura integral diseñada para ofrecer seguridad de confianza cero (para usuarios, cargas de trabajo, aplicaciones, dispositivos OT e IoT, etc.) reduce el riesgo, mejora la protección, simplifica la experiencia del usuario y representa una mejora fundamental en la forma en que pensamos en la seguridad empresarial.

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