White paper

Breve storia dello zero trust: le tappe principali del percorso verso il ripensamento della sicurezza aziendale

Perché raccontare la storia dello zero trust?

Nel settore della sicurezza informatica, in molti ritengono lo zero trust una svolta, un rinnovamento radicale della sicurezza aziendale e della protezione delle reti e delle risorse che ospitano le nostre idee migliori, connettono i nostri brillanti talenti e garantiscono l'accesso a strumenti di produttività trasformativi.

Per comprendere la portata rivoluzionaria del modello zero trust per la sicurezza informatica, è necessario conoscere il modo in cui l'idea di un'architettura zero trust si è evoluta fino a diventare una tecnologia che ha portato al ripensamento di decenni di vecchie convinzioni.

Reti 2D e sicurezza a castello e fossato

"Hub-and-spoke" e "castello e fossato" sono le due principali espressioni utilizzate rispettivamente per descrivere l'architettura di rete legacy e la sicurezza della rete. Le idee alla base di queste due espressioni e metafore sono molto antiche, e questo non sorprende.

L'architettura di rete hub-and-spoke si riferisce al modello in cui vi è un hub centrale intorno a cui sono disposte altre reti. Questo modello prevede che il traffico interno ed esterno, prima di poter procedere verso la sua destinazione, venga instradato verso un data center principale attraverso un set di soluzioni di sicurezza. Sebbene questo approccio abbia funzionato per diverso tempo, è diventato più complicato e costoso a causa dell'adozione del cloud, della forza lavoro sempre più distribuita e della crescente importanza della mobilità.

Con il termine "sicurezza a castello e fossato" ci si riferisce invece alle reti indipendenti progettate per ammettere il traffico non dannoso e mantenere i nemici al di fuori del perimetro. Gli apparecchi di sicurezza che si trovano in sede agiscono come guardie del perimetro, e hanno l'obiettivo di far entrare ciò che non presenta rischi e tenere al di fuori potenziali nemici. L'enorme transizione delle applicazioni verso il cloud, insieme allo spostamento dei lavoratori al di fuori dei perimetri aziendali, ha reso rapidamente obsoleto questo approccio, che è invecchiato più in fretta delle palle di cannone che proteggevano i castelli veri e propri.

Le VPN e il Wi-Fi hanno complicato ulteriormente il problema. La vecchia architettura a castello e fossato non consentiva agli amministratori di collegare gli ospiti a una rete senza lasciare loro libertà di movimento, e l'unico modo per connettere gli endpoint alle reti in modo sicuro era quello di usare una qualche forma di segmentazione della rete.

Avevamo bisogno di una soluzione migliore.

802.1X e i problemi con il NAC

Nel 2001, la IEEE Standards Association ha pubblicato il protocollo 802.1X per il controllo dell'accesso alla rete (NAC, Network Access Control).

> "Un mezzo per autenticare e autorizzare i dispositivi collegati a una porta LAN con caratteristiche di connessione punto a punto e per impedire l'accesso a tale porta nei casi in cui il processo di autenticazione e autorizzazione fallisca."
>
> IEEE sul protocollo 802.1X →

Poco dopo, i dispositivi wireless hanno iniziato a includere un supplicant (o client) 802.1X, che permetteva alle reti di autenticare l'endpoint prima di consentire la connessione. Questo progresso era pensato per offrire la possibilità di bloccare le reti cablate e wireless, in modo che solo i dispositivi gestiti e gli utenti autorizzati potessero connettersi. Il meccanismo del supplicant può essere visto come un buttafuori alla porta di accesso alla rete a cui viene presentato un documento d'identità, che decide chi far entrare e chi lasciare fuori.

Purtroppo, il modello NAC non è stato una panacea, e i problemi sono partiti proprio dalla rete. Le reti interne, infatti, erano state progettate tenendo conto dell'attendibilità implicita, e cercare di aggiungere l'autenticazione/autorizzazione a posteriori era estremamente complesso. Affinché il NAC fosse pienamente efficace, tutte le porte accessibili dovevano essere bloccate, ma non tutti i dispositivi erano compatibili con il protocollo 802.1X. L'aumento dell'adozione di stampanti connesse a Internet, lettori di badge e altri dispositivi connessi alla rete ha generato un problema di sicurezza non indifferente. Torniamo all'esempio di prima: in uno scenario del genere, è come se il nostro buttafuori continuasse a presidiare un'unica porta di accesso, ma nel frattempo si stessero aprendo molte altre entrate alternative.

Il crollo delle mura di Gerico e il ripensamento del perimetro nella sicurezza.

Nel 2003 era ormai chiaro che l'uso dei dispositivi personali avrebbe continuato a diffondersi e che le organizzazioni avrebbero dovuto iniziare a pensare a come proteggere i computer non protetti dalle mura del castello. Inoltre, l'uso crescente della crittografia riduceva l'efficacia dei firewall perimetrali, e costringeva a una scelta: adattare le prestazioni, per affrontare le sfide di decrittazione e ispezione, o consentire al traffico criptato di passare inosservato?

Durante lo stesso anno, un gruppo di leader europei attivi nel settore della tecnologia si è riunito per affrontare i temi dell'autenticazione degli utenti, della crittografia, della gestione delle identità e dell'applicazione delle policy. Dopo essersi formalmente istituito nel 2004, il forum Jericho introdusse il concetto di "deperimetrazione".

Con un nome che ricorda la storia biblica degli israeliti che abbatterono le mura dell'antica città di Gerico, questo forum si è dedicato a risolvere il problema di come "abilitare flussi di informazioni sicuri e senza confini tra le aziende".

Oltre all'azzeccata metafora, il gruppo dettò i Comandamenti del forum Jericho, quanto di più vicino ci fosse fino a quel momento a delle direttive fornite dall'alto per la gestione delle reti senza perimetro. Sfortunatamente, la serie di controlli e mitigazioni prescritte andava oltre la capacità di distribuzione e amministrazione della maggior parte delle aziende in quel momento storico.

Lo "zero trust" entra per la prima volta a far parte del lessico informatico

Nel 2010, l'analista di Forrester John Kindervag pubblicò un documento intitolato “No More

Chewy Centers: Introducing The Zero Trust Model of Information Security", e presto si ebbe una nuova parola chiave per descrivere un nuovo modo di concepire la sicurezza della rete. Questo documento conteneva un'affermazione fondamentale: la semplice presenza su una rete non era sufficiente a garantire l'attendibilità.

"Da quel momento abbiamo iniziato a sentire parole come: l'identità è il nuovo perimetro", afferma Lisa Lorenzin, CTO di Zscaler Field e veterana dello zero trust. "Autenticavamo un utente e usavamo quell'identità per determinare ciò che potesse fare. Se eravamo fortunati, potevamo raccogliere del contesto, ad esempio se si trattava di un dispositivo gestito o non gestito, e prendere decisioni sull'accesso basate su tale comprensione rudimentale".

Era un progresso; tuttavia, vincolava la sicurezza aziendale alla protezione delle reti stesse, e non eravamo ancora pronti ad abbandonare completamente questo approccio. Non riuscendo ancora a implementare una strategia trasformativa, l'applicazione di questi principi si è arenata. Il problema era che si continuava a fare affidamento sullo stesso set di strumenti incentrati sulla rete: 802.1X e RADIUS al Livello 2, firewall che prendevano in considerazione l'identità al Livello 3, ecc.

Si trattava del solito NAC con un nome più accattivante.

Oltre (il perimetro) dell'azienda

Nel frattempo, l'azione sofisticata e brillante di hacker legati all'Esercito Popolare di Liberazione (PLA) cinese stava portando il settore a riconsiderare completamente il concetto di attendibilità. Nel 2010, Google rivelò un'operazione del 2009 che aveva preso di mira l'azienda e diverse altre società tecnologiche di alto profilo, tra cui Akamai, Adobe e Juniper Networks. I ricercatori di sicurezza di McAfee battezzarono quella campagna "Operazione Aurora".

Colpendo direttamente l'élite dell'ingegneria informatica, gli hacker cinesi hanno accelerato involontariamente il lavoro sull'architettura zero trust svolto dai migliori laboratori tecnologici della nazione. A seguito dell'Operazione Aurora, Google ha sviluppato BeyondCorp, che si concentrava sullo **"spostamento dei controlli di accesso dal perimetro della rete ai singoli utenti...[consentendo di] lavorare in sicurezza praticamente da qualsiasi luogo, senza la necessità di una VPN tradizionale".**

Ma "Google è un'azienda gestita da ingegneri, per ingegneri, con un budget praticamente infinito e un'infrastruttura legacy relativamente ridotta rispetto a quella di molte aziende", afferma Lorenzin. "E ci sono voluti comunque sette anni e sei white paper per la progettazione e l'implementazione".

Anche con l'esempio ben documentato di Google, l'architettura zero trust era ancora fuori dalla portata dalla maggior parte delle aziende. Nonostante stesse provando a facilitare il percorso per le altre organizzazioni, affinché potessero realizzare la propria implementazione di una rete zero trust, il futuro immaginato da Google era ancora lontano.

Nel frattempo, la popolarità del cloud e la continua enfasi sulla mobilità hanno fatto sì che un numero maggiore di dati fosse disponibile e accessibile dall'esterno del perimetro della rete, e un approccio diffuso per valutare l'attendibilità risultava ormai essenziale.

Gartner e l'arrivo dello ZTNA (Zero Trust Network Access)

La società di ricerca tecnologica Gartner è stata responsabile dei successivi progressi dello zero trust, che si è evoluto in un'architettura largamente adattabile. Sebbene fosse già in circolazione, il concetto di "zero trust" non era una priorità nel 2010, quando l'azienda ha pubblicato il documento CARTA (Continuous Adaptive Risk and Trust Assessment).

Questo documento descriveva la necessità di capire chi richiedesse l'accesso e che quest'ultimo fosse concesso in base a una valutazione dinamica dell'ambiente, del contesto disponibile e delle responsabilità dell'utente.

Lorenzin descrive CARTA come "un ottimo modello che non ha mai ottenuto la diffusione che meritava".

Con il lavoro di Gartner, il CARTA alla fine si è trasformato nel modello "Zero Trust Network Access" (ZTNA), dato che il framework originale non è riuscito ad acquisire il supporto dei professionisti della tecnologia (il focus dell'accesso era ancora sulle reti). Tuttavia, CARTA rimane importante nella storia dello zero trust, perché i principi enunciati nel documento continuano a vivere nell'approccio ZTNA.

Il successivo importante contributo di Gartner a questa discussione è stato dato dal riconoscere la convergenza di rete e sicurezza. Nel 2019, questa convergenza si è tradotta nel concetto di Secure Access Service Edge (SASE). L'unione, però, è stata di breve durata, e nel 2021 le categorie sono state nuovamente divise con l'introduzione del Secure Service Edge (SSE): il SASE senza WAN.

Gartner si era ormai affermata come un arbitro determinante di ciò che sarebbe stato o meno il futuro dello zero trust, e i provider di servizi si affannavano per inserirsi in una delle nuove categorie di mercato.

Le autorità entrano nella discussione: NIST, OMB e il supporto dello ZTA da parte del governo

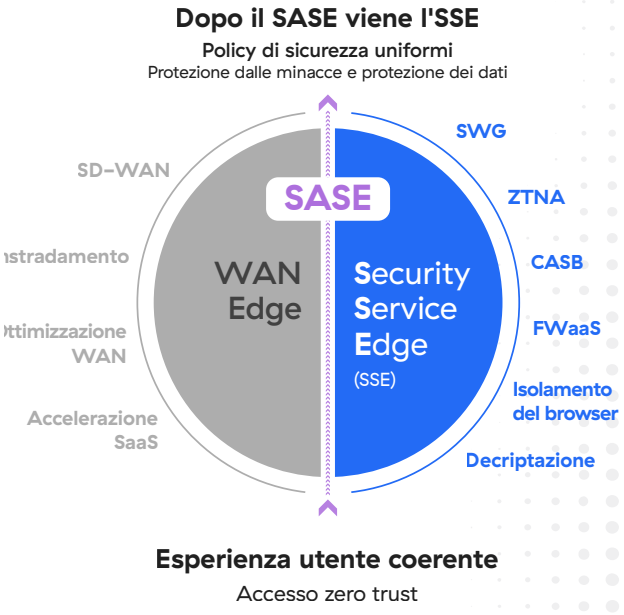
Nel 2020, il National Institute for Standards and Technology (NIST) ha introdotto lo standard NIST 800-207 per l'architettura zero trust. Questo nuovo paradigma della sicurezza informatica era incentrato sulla protezione delle risorse e sulla premessa che l'attendibilità non deve mai essere concessa automaticamente, ma costantemente rivalutata.

Questo white paper ha portato all'abbandono dei vincoli del perimetro e del concetto di VPN. L'attenzione si è spostata dalla protezione della rete alla protezione degli utenti, dei dati e delle applicazioni che interagivano attraverso la rete.

Il concetto di zero trust, da quel momento, ha iniziato a indicare semplicemente l'accesso basato sul contesto e a privilegi minimi, applicabile a una varietà molto più ampia di casi d'uso e flussi di traffico.

Lo standard 800-207 sancisce i principi e i presupposti fondamentali dello zero trust. Tre dei punti più critici (da un elenco molto più lungo) sono:

1. Nessuna risorsa è intrinsecamente attendibile.
2. Tutte le comunicazioni sono protette, indipendentemente dalla posizione in rete. La richiesta va interrotta e ispezionata; va esaminato tutto il contesto disponibile associato all'utente e alla richiesta.
3. L'autenticazione e l'autorizzazione delle risorse sono dinamiche e vengono rigorosamente applicate prima che venga consentito l'accesso.



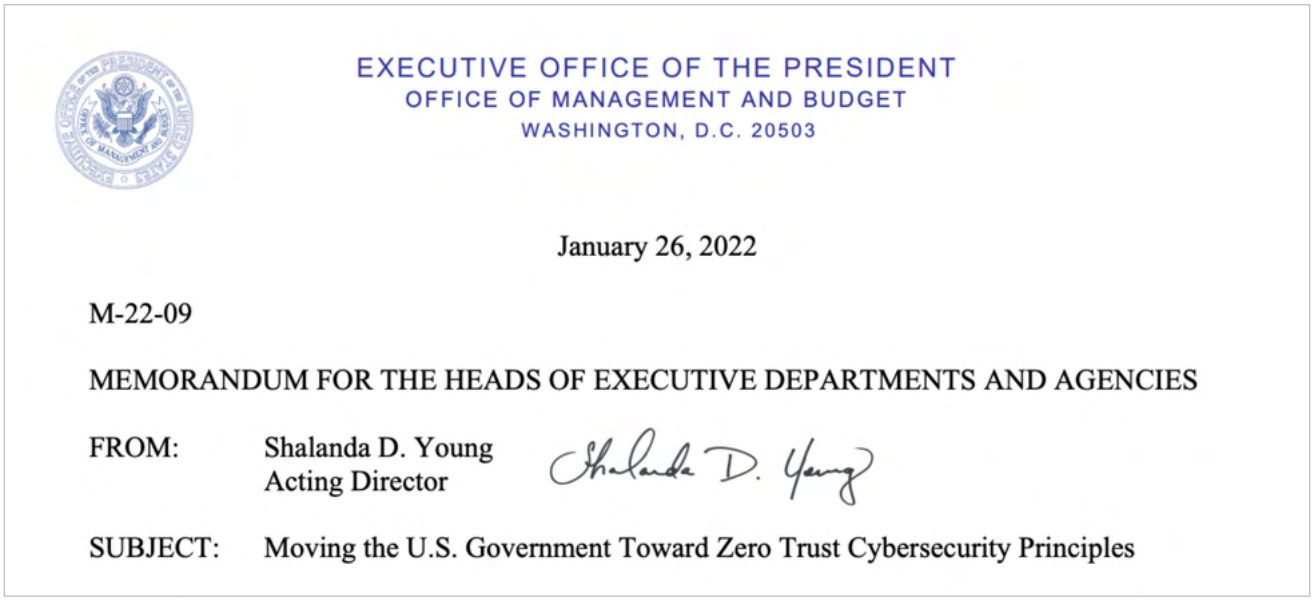
Ma il vero punto di svolta nella promozione dei principi dello zero trust è arrivato dall'alto, almeno negli Stati Uniti. L'Office of Management and Budget degli Stati Uniti, l'ufficio responsabile dell'implementazione delle politiche presidenziali, con la sua direttiva M-22-09 emessa nel 2022, ha affermato che entro il 2024 tutti gli uffici del governo federale dovranno adottare i principi dell'architettura zero trust, e delinea in modo preciso le tappe principali e le date target di questo percorso.

"Finora abbiamo avuto dei documenti di orientamento e modelli per gli amministratori. Ma la strategia federale di adozione dello zero trust è il primo vero punto di svolta", continua Lorenzin.

L'attacco alla catena di approvvigionamento contro la piattaforma di gestione IT Solar Winds, reso noto nel 2021 e responsabile della compromissione di almeno nove agenzie federali, tra cui Stato, Tesoro, Sicurezza interna, Commercio ed Energia, è stato forse l'attacco partito da uno Stato più audace e dannoso dai tempi dell'Operazione Aurora. In risposta, il governo federale ha puntato sullo zero trust, adottando questo approccio come punto di riferimento per la sicurezza informatica per gli anni a venire.

Implementazione dello zero trust

L'approccio di Zscaler all'architettura zero trust si allinea strettamente all'architettura ZTA del NIST e alla definizione di Gartner per l'SSE, ma si spinge oltre questi standard, perché si impegna a implementare tre miglioramenti fondamentali nel concetto di zero trust. Questi principi avanzati aiutano a realizzare pienamente il concetto di 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

Lo zero trust va applicato a tutto il traffico

Lo zero trust è un approccio nato come un nuovo modo per proteggere le reti. Successivamente, si è esteso oltre le reti locali, ma ha continuato a focalizzarsi principalmente sul traffico delle applicazioni private. Per troppo

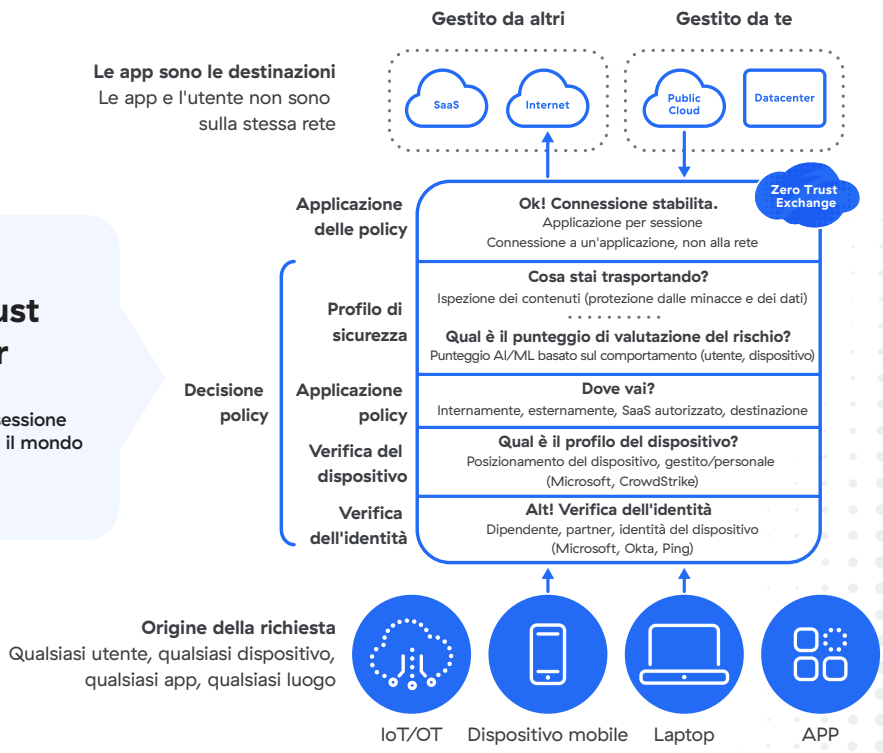
tempo il traffico è stato valutato solo sulla base della sua relazione con una rete, e non ci si è completamente allontanati dal concetto stesso di rete.

Tuttavia, ora sappiamo che i principi dello zero trust possono essere applicati alla protezione delle applicazioni SaaS, del traffico da e verso i cloud pubblici, e persino degli utenti che accedono a Internet, e che le fonti che danno origine a questo traffico possono essere sia i carichi di lavoro che gli utenti. L'accesso può essere reso indipendente dal mezzo, e il traffico può passare attraverso qualsiasi router e attraverso qualsiasi rete, cablata o wireless, 4G o 5G, e così via.

È giunto il momento di applicare i principi dello zero trust a tutto il traffico, indipendentemente dall'origine e dalla destinazione. Abbiamo già superato il concetto della distinzione tra entità attendibili e non attendibili e il concetto del posizionamento all'interno o al di fuori della rete. Non dobbiamo più pensare alla connessione fra entità e reti, ma dobbiamo utilizzare lo zero trust per connettere tutte le entità in modo diretto utilizzando le policy aziendali. Internet è la nuova rete aziendale, e tutto il traffico è un potenziale bersaglio.

L'architettura Zero Trust Exchange di Zscaler

La decisione delle policy avviene per ogni sessione e l'applicazione si verifica su 150 DC in tutto il mondo



1 L'identità e il contesto vengono sempre prima della connettività

La verifica dell'identità è il cuore dello zero trust. In passato, però, abbiamo confuso l'identità con la connettività, e questo ci ha portato ad adottare modelli inappropriati. Indirizzi IP, indirizzi MAC e porte e protocolli non fanno parte dell'identità.

I dispositivi OT possono connettersi alle reti dalle fabbriche e gli utenti possono accedere dai bar, ma questo non ci dice nulla sulla loro identità. Dobbiamo quindi partire dagli attributi che compongono l'identità reale e dal contesto; solo a partire da queste informazioni possiamo autorizzare la connettività.

Quando un utente richiede l'accesso a una risorsa, dobbiamo prima considerare chi è ed esaminare le altre informazioni che abbiamo sulla sua identità, come il ruolo o reparto e il dispositivo che sta utilizzando; a questo punto, passiamo quindi alle policy di sicurezza. Cosa sta cercando di fare? Dove sta andando? Quali elementi dell'ambiente potrebbero contribuire alla nostra decisione di consentire o negare l'azione?

Il contesto va oltre l'identità, e viene valutato costantemente. Gli altri fattori che possono essere verificati per individuare eventuali anomalie includono la geolocalizzazione, l'indirizzo IP, il profilo del dispositivo e l'ora del giorno. Inoltre, una soluzione zero trust dovrebbe essere in grado di decriptare il traffico per ispezionare le minacce e i rischi di esfiltrazione dei dati, inline e in modo scalabile.

Nel caso di Zero Trust Exchange, correliamo anche l'intelligence sulle minacce proveniente dal nostro cloud globale e da partner terzi, come ad esempio i provider di servizi di sicurezza e di verifica dell'identità, per determinare il rischio e prendere decisioni sulle policy e sull'accesso.

2 Le applicazioni e gli ambienti delle app dovrebbero rimanere invisibili agli utenti non autorizzati

Ora che abbiamo risolto il problema di come conoscere l'identità dell'utente prima di concedere l'accesso, possiamo affrontare la prossima sfida: come lo colleghiamo alle risorse autorizzate, riducendo al contempo i rischi e riducendo al minimo la possibilità di compromissione? Dopo che il contesto che circonda un utente, un dispositivo, una policy e un ambiente è stato raccolto e analizzato, possiamo proseguire con i passaggi successivi.

Eliminando l'ascoltatore in entrata dalle connessioni da remoto, si elimina anche la superficie di attacco esterna. In caso contrario, sarebbe davvero semplice per gli aggressori individuare dei gateway VPN vulnerabili o delle applicazioni esposte per compromettere gli obiettivi. Le VPN in attesa di connessioni in entrata sono delle facili prede, e gli aggressori lo sanno bene. Si tratta di un problema che non dipende dal provider di servizi, e può essere risolto solo cambiando il modello architetturale.

Zero Trust Exchange di Zscaler risponde instaurando connessioni solo in uscita, sia dall'utente che dall'ambiente applicativo verso il nostro security cloud, utilizzando dei microtunnel criptati per mediare le connessioni tra le richieste e le relative destinazioni.

Questo "terzo luogo" online fa da cuscinetto tra gli utenti verificati e le risorse a cui sono autorizzati ad accedere. Una volta che un utente è connesso alla risorsa richiesta, le policy granulari assicurano che non vi sia alcuna possibilità che si avventuri verso altre destinazioni, e il movimento laterale diventa sostanzialmente impossibile.

3 Il capitolo finale?

I principi discussi in precedenza ci permettono di superare definitivamente la concezione tradizionale in cui vi erano perimetri di rete protetti da firewall ed endpoint in remoto collegati tramite VPN. Questi principi non si limitano a replicare i controlli di sicurezza esistenti in un'istanza virtuale ospitata sul cloud, né si basano su una comprensione artificiale di ciò che è presente sulla rete e di ciò che non lo è.

Un'architettura completa, progettata per garantire la massima sicurezza per utenti, carichi di lavoro, applicazioni, dispositivi OT e IoT e molto altro, riduce i rischi, migliora la protezione, semplifica l'esperienza utente e rappresenta un radicale miglioramento della sicurezza aziendale.

zscaler™ | Experience your world, secured.™

Informazioni su Zscaler

Zscaler (NASDAQ: ZS) accelera la trasformazione digitale, in modo che i clienti possano essere più agili, efficienti, resilienti e sicuri. Zscaler Zero Trust Exchange protegge migliaia di clienti dagli attacchi informatici e dalla perdita dei dati, grazie alla connessione sicura di utenti, dispositivi e applicazioni in qualsiasi luogo. Distribuita in più di 150 data center a livello globale, Zero Trust Exchange, basata su SASE, è la più grande piattaforma di cloud security in linea del mondo. Scopri di più su **zscaler.it** o seguici su **Twitter @zscaler**.

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