WHITE PAPER

Demistificare la sicurezza per i data center iperscalabili

L’hyperscale computing è insufficiente a causa delle limitazioni a livello di rete e sicurezza

Sintesi preliminare

Molti settori hanno introdotto l'hyperscale computing per affrontare tutta una serie di casi d'uso diversi. Ad esempio, le grandi aziende che costruiscono architetture IT ibride per lanciare rapidamente le applicazioni richiedono una comunicazione superveloce tra risorse distribuite in domini fisici e virtuali. Gli istituti di ricerca avanzati, come quelli che svolgono attività di ricerca genomica o aerospaziale, hanno bisogno di poter trasferire enormi set di dati in rete. Le aziende di e-commerce, tra cui l'e-retail ad alta velocità, utilizzano architetture iperscalabili per gestire i picchi di connessioni durante i grandi eventi promozionali come il Cyber Monday e affrontare le impennate dell'attività derivanti dalla pandemia COVID-19.

Un problema dei data center iperscalabili, tuttavia, è la mancanza di un'adeguata sicurezza. Perché? Perché l'attivazione della sicurezza spesso crea un collo di bottiglia nella rete se le soluzioni di sicurezza faticano a tenere il passo o se l'infrastruttura IT non è adeguatamente segmentata. Con firewall di rete non sufficientemente qualificati che creano colli di bottiglia, molti responsabili della gestione tecnica e operativa della rete bypassano del tutto la sicurezza. Questo lascia scoperta l'organizzazione di fronte a vari tipi di attacchi che possono influenzare il modo in cui vengono svolte anche le funzioni di base.

La sicurezza dei data center iperscalabili richiede una strategia aggiornata e una tecnologia all'altezza.

Introduzione

Le iniziative di innovazione digitale (DI) e le esigenze dell'attività hanno cambiato il modo in cui i data center aziendali vengono utilizzati e i parametri prestazionali che devono rispettare. L'adattamento alla domanda di nuove capacità di rete ha guidato l'evoluzione dei data center iperscalabili.

Un data center iperscalabile è un data center in grado di adattarsi in modo efficiente e comportarsi in modo dinamico per soddisfare le esigenze aziendali variabili. Le architetture iperscalabili sono progettate per soddisfare esigenze senza precedenti di enormi capacità e prestazioni astronomiche, esigenze che possono variare da settore a settore. Tra gli esempi di attività che hanno bisogno di un'architettura iperscalabile potremmo citare:

- **Grandi aziende, compresi provider di servizi cloud.** Alle organizzazioni che utilizzano la virtualizzazione per creare reti virtuali molto scalabili servono segmentazioni di rete su larga scala basate su VXLAN (Virtual Extensible Local-Area Network) e comunicazioni veloci tra servizi co-ospitati su piattaforme fisiche e virtuali.
- **E-commerce dinamico, compreso l'e-retail ad alta velocità.** I picchi di connessioni in caso di eventi promozionali di vendita come il Black Friday, la presentazione online delle dichiarazioni fiscali o delle richieste di sussidio di disoccupazione, richiedono la capacità di gestire un enorme numero di connessioni di utenti al secondo.[1]
- **Ricerca avanzata nel settore farmaceutico, petrolchimico e aerospaziale.** L'uso di big data e algoritmi di apprendimento automatico (ML) per la ricerca avanzata richiede la capacità di trasmettere flussi di lunga durata da 40 Gbps e 100 Gbps.[2]
- **Borse.** Per l'infrastruttura di trading elettronico è necessario che i dati di mercato siano ricevuti con la minore latenza possibile.[3]
- **Hyperscaler (grandi aziende tecnologiche globali).** Le interconnessioni ad alta velocità tra i data center in cloud per replicare i dati nei siti di disaster recovery (DR) richiedono interfacce ad alta velocità e capacità di tunneling IPsec ad alto throughput per garantire la privacy e la riservatezza dei dati.[4]

In molti casi, le organizzazioni che operano in questi settori hanno investito nell'infrastruttura di rete di cui hanno bisogno. Tuttavia, l'approvvigionamento di soluzioni di sicurezza in grado di soddisfare tali esigenze rappresenta una sfida in quanto gli attuali firewall di nuova generazione (NGFW) non sono in grado di soddisfare le enormi esigenze di scalabilità e prestazioni delle architetture iperscalabili. I NGFW esistenti si trovano in difficoltà se le organizzazioni vogliono eseguire il controllo di ammissione su decine di milioni di connessioni di utenti al secondo, o introdurre protocolli di protezione dagli attacchi DDoS (Denial-of-Service) insieme a firewall essenziali di livello 4. Il conseguente degrado delle prestazioni fa sì che molte organizzazioni si limitino a disattivare le funzionalità di sicurezza, temendo che ciò rallenti l'attività e impedisca di ottimizzare il throughput e la latenza dell'infrastruttura di rete. Ma questo è un compromesso pericoloso: la scelta dell'organizzazione di soddisfare le crescenti esigenze senza adeguati controlli di sicurezza è come sfidare la fortuna sperando di non essere attaccata.

Sfide delle architetture iperscalabili

Ogni ambiente di distribuzione iperscalabile pone sfide per la sicurezza.

Difficoltà nell'eseguire servizi virtualizzati molto scalabili

Le aziende devono essere in grado di lanciare i servizi nel modo più agile possibile per aumentare la produttività e i ricavi. Per massimizzare i benefici del ritorno sull'investimento (ROI),i servizi devono interagire tra le risorse fisiche e virtuali.

Sfruttando tecnologie molto scalabili come la VXLAN, i clienti possono segmentare tutti i servizi virtualizzati per realizzare una scalabilità che non è raggiungibile con la VLAN. I servizi virtualizzati possono essere scalati in entrambe le direzioni e spostarsi senza costi di esercizio significativi. Questi servizi sono spesso necessari per comunicare con altri servizi che si trovano nell'infrastruttura fisica esistente. La maggior parte delle soluzioni attuali, tuttavia, ha basse prestazioni e alta latenza, oltre a non disporre della sicurezza essenziale di livello 4 per monitorare lo stato della sessione e i controlli di ammissione che stabiliscono chi è autorizzato e chi non lo è. Inoltre, non offrono la sicurezza avanzata di livello 7 che rileverebbe ulteriormente le minacce e farebbe rispettare la policy per ottenere la compliance e gestire i rischi.

I picchi di connessioni in caso di particolari eventi travolgono la sicurezza anelastica

In altri settori, il volume di ogni singola connessione non è tanto significativo quanto il numero totale di connessioni che l'organizzazione deve essere in grado di elaborare in tempi molto brevi. Durante i grandi eventi promozionali come il Black Friday o il Cyber Monday, i siti di e-commerce registrano volumi estremamente elevati di traffico di consumatori nell'arco delle 24 ore, fino a 1,5 volte in più rispetto ai giorni dell'anno in cui si registra il maggior numero di acquisti.[5]

Analoghi picchi di connessioni si verificano durante i periodi di presentazione delle dichiarazioni fiscali, all'apertura della vendita di biglietti per grandi eventi, durante festività come il capodanno cinese e negli ambienti di gaming online, in particolare nelle partite multigiocatore a cui possono partecipare contemporaneamente centinaia di giocatori con picchi di 30 minuti.

In risposta, le architetture iperscalabili consentono ai contribuenti, ai retailer e ai servizi di game-hosting di accettare ed elaborare in modo efficiente milioni di connessioni in entrata al secondo. La motivazione commerciale per investire nell'iperscalabilità è semplice: la perdita di connessioni o la lentezza nella risposta può causare la perdita di vendite e danni all'immagine. Ad esempio, in media, da uno a tre secondi di ritardo nel tempo di caricamento delle pagine comporta un aumento del 32% dei clienti che lasciano il sito.[6]

Grandi flussi di rete vulnerabili agli attacchi

Le applicazioni di intelligenza artificiale (AI) e apprendimento automatico (ML) richiedono enormi set di dati, che spesso raggiungono più terabyte[7] per l'apprendimento e il test degli algoritmi. Le organizzazioni del settore farmaceutico, biotecnologico, genomico e petrolchimico hanno bisogno tutte di questi grandi set di dati per la ricerca. Per elaborare e analizzare i dati, gli istituti di ricerca devono poter trasmettere enormi set di dati in modo efficiente in rete. Ma una trasmissione efficiente ha bisogno di una larghezza di banda che può raggiungere fino a 100 Gbps, o quello che è noto come "elephant flow", ossia flusso di lunga durata.

Teoricamente, gli istituti di ricerca dovrebbero essere in grado di sfruttare architetture di rete iperscalabili costruite su router e switch per ottenere questa larghezza di banda. Questi dispositivi, tuttavia, non monitorano lo stato della sessione e offrono una sicurezza essenziale di livello 4. Poiché gli attacchi DDoS diventano sempre più frequenti, questi dispositivi sono inoltre vulnerabili agli attacchi.

Peraltro, i dati trasmessi su queste connessioni sono spesso sensibili e coperti da leggi sulla protezione dei dati come il regolamento generale sulla protezione dei dati dell'Unione europea (GDPR) o l'Health Insurance Portability and Accountability Act (HIPAA). Tali normative impongono vari controlli di accesso, il che significa che il traffico di rete deve essere instradato attraverso tecnologie di sicurezza come i firewall e il flusso dei messaggi va crittografato. La maggior parte dei NGFW, però, non può gestire larghezze di banda di connessione superiori a 10 Gbps. E questo non solo rallenta la ricerca di un margine significativo, ma impedisce anche alle organizzazioni di massimizzare il ROI sui collegamenti WAN esistenti realizzati per trasmettere dati a 40 Gbps e 100 Gbps poiché un singolo flusso attraverso gli attuali NGFW può supportare solo 10 Gbps senza bloccarsi.

La latenza del firewall può portare a perdite milionarie

Per il trading sul mercato azionario, il gaming competitivo e settori analoghi, la bassa latenza della rete è estremamente importante. Anche lievi ritardi nel tempo di andata e ritorno (RTT) del traffico di rete possono avere un impatto significativo sulla redditività o sulle prestazioni.

Di conseguenza, le organizzazioni del settore finanziario generalmente investono in infrastrutture di rete che forniscono una latenza estremamente bassa per i loro data center. L'infrastruttura di trading elettronico non tollera più di 5 µs di latenza.[8] In contesti sensibili alla latenza, molte organizzazioni configurano i loro NGFW in modalità di monitoraggio, sacrificando la sicurezza per il throughput di rete.[9]

Le interconnessioni ad alta velocità tra i data center hanno bisogno una connettività IPsec ad alto throughput

Per i provider di servizi cloud e le organizzazioni che gestiscono reti di distribuzione di contenuti (CDN), la capacità di replicare i dati su più siti regionali è essenziale. La ragione principale per cui le organizzazioni utilizzano i siti regionali per ospitare copie complete dei dati memorizzati è che aumentano la resilienza, diminuiscono la latenza delle risposte alle richieste dei clienti e diminuiscono il carico sul data center principale.

Affinché ciò sia possibile, alle organizzazioni servono interconnessioni tra i data center, ossia collegamenti a banda larga tra i siti regionali con il compito di supportare la sincronizzazione della rete.[10] Poiché i provider di servizi cloud e le CDN trasmettono spesso dati sensibili o proprietari, questi collegamenti sono spesso realizzati come tunnel IPsec. Tuttavia, allo stesso tempo, la sicurezza della rete di livello 4 richiede che i NGFW siano in grado di elaborare il traffico IPsec con lo stesso throughput dei collegamenti di rete. Ma, poiché la maggior parte dei NGFW esistenti non può raggiungere un throughput IPsec superiore a 10 Gbps, i NGFW che proteggono questi collegamenti possono rallentare il trasferimento complessivo di grandi quantità di dati attraverso i data center.

Conclusioni

Gli sforzi di innovazione digitale, volti a migliorare l'efficienza e l'esperienza del cliente, richiedono l'evoluzione dell'infrastruttura di rete. I data center iperscalabili sono progettati per supportare flussi di rete massicci, picchi di connessioni e diversi altri casi d'uso.

Mentre molte organizzazioni hanno distribuito un'architettura di rete iperscalabile, l'ottenimento di una sicurezza iperscalabile è una grossissima sfida. Disabilitare i NGFW o metterli in modalità di monitoraggio, per eliminare i colli di bottiglia nella rete, lascia scoperta un'organizzazione di fronte agli attacchi e potenzialmente non conforme alle normative sulla protezione dei dati. Non segmentare le applicazioni e l'infrastruttura IT permette a un intruso di entrare e raggiungere il nucleo della rete una volta superato il perimetro. E questi esiti pericolosi si aggravano ancora di più se gli attacchi provengono da utenti interni e fidati.

I data center iperscalabili hanno bisogno di un approccio completamente diverso alle soluzioni di sicurezza, un approccio che permetta loro di adattarsi alle crescenti esigenze aziendali. In assenza di una soluzione di sicurezza iperscalabile che sia in grado di gestire i picchi di connessioni di utenti possibili, elaborare decine di milioni di connessioni al secondo, supportare flussi di lunga durata a 100 Gbps, segmentare in modo efficiente gli ambienti virtuali massicci, proteggere il perimetro aziendale con una sicurezza essenziale di livello 4 ad alte prestazioni e prevenire gli attacchi DDoS, ogni scommessa è persa a tutto vantaggio di qualsiasi malintenzionato, il cui unico scopo è sferrare attacchi informatici che interferiscano con l'attività di un'azienda, scatenino una cattiva pubblicità e ne decretino infine la morte.

[1] Marisa Sanfilippo, "The Best Days for Holiday Sales: A Guide for Businesses," Business News Daily, 2 dicembre 2019.

[2] Rajiv Kohli and Nigel P. Melville, "Digital innovation: A review and synthesis," Information Systems Journal, 29 gennaio 2018.

[3] "Deterministic Communications for Secure High-speed Performance: Fortinet Protects Connections to Electronic Trading Platforms with the Industry's Lowest Latency and Jitter Rates," Fortinet, 23 settembre 2019.

[4] "What is DCI?" Ciena, 16 maggio 2019.

[5] Marisa Sanfilippo, "The Best Days for Holiday Sales: A Guide for Businesses," Business News Daily, 2 dicembre 2019.

[6] "Find Out How You Stack Up to New Industry Benchmarks for Mobile Page Speed," Google, marzo 2017.

[7] Mohammad Shaikh and Harsha Gurukar, "Machine Learning and HPC in Pharma Research and Development," Super Computing 2019, novembre 2019.

[8] "Deterministic Communications for Secure High-speed Performance: Fortinet Protects Connections to Electronic Trading Platforms with the Industry's Lowest Latency and Jitter Rates," Fortinet, 23 settembre 2019.

[9] Jason Pappalexis, "The NGFW Today: A Staple of Network Security in Spite of Challenges," NSS Labs, 11 marzo 2019.

[10] "What is DCI?" Ciena, 16 maggio 2019.

Copyright © 2020 Fortinet, Inc. Tutti i diritti riservati. Fortinet®, FortiGate®, FortiCare®, FortiGuard® e altri marchi sono marchi registrati di Fortinet, Inc. Anche altri nomi Fortinet qui citati possono essere marchi registrati e/o marchi di diritto comune di Fortinet. Tutti gli altri nomi di prodotti o società possono essere marchi registrati dei rispettivi proprietari. I dati riportati relativi a prestazioni e altre caratteristiche sono stati ottenuti con prove interne di laboratorio in condizioni ideali e, pertanto, le prestazioni effettive e altri risultati possono variare. Elementi variabili della rete, diversi ambienti di rete e altre condizioni possono influenzare i risultati delle prestazioni. Nulla di quanto qui contenuto rappresenta un impegno vincolante per Fortinet, e Fortinet esclude qualsiasi garanzia, esplicita o implicita, eccetto quelle previste da un contratto scritto, firmato da un rappresentante legale di Fortinet, che garantisca esplicitamente all'acquirente che le prestazioni del prodotto indicato saranno conformi a determinati dati esplicitamente indicati. In tal caso, solo gli specifici dati delle prestazioni esplicitamente identificati in tale contratto scritto saranno vincolanti per Fortinet. Per chiarezza, qualsiasi garanzia è limitata alle prestazioni ottenute nelle stesse condizioni ideali delle prove interne di laboratorio di Fortinet. Fortinet esclude in toto qualsiasi convenzione, rappresentanza e garanzia, esplicita o implicita, sulla base del presente documento. Fortinet si riserva il diritto di cambiare, modificare, trasferire o comunque revisionare questa pubblicazione senza alcun preavviso. La versione applicabile della presente pubblicazione è quella più recente.

novembre 19, 2020 11:49 AM

744553-A-1-IT