FORTINET

LIVRE BLANC

Démystifier la sécurité des datacenters hyperscale

Une fragilité due aux limites en termes de réseaux et de sécurité

Synthèse

Nombre de secteurs d'activité ont adopté l'informatique hyperscale (à très grande échelle) et ses différents cas d'utilisation. Ainsi, les grandes entreprises qui bâtissent des architectures informatiques hybrides pour lancer rapidement des applications exigent une communication ultrarapide entre leurs systèmes disséminés dans des environnements physiques et virtualisés. Les institutions de recherche avancée, celles opérant dans la génomique ou l'aérospatiale notamment, doivent pouvoir transférer des volumes importants de données sur leurs réseaux. Les acteurs du e-commerce, et notamment ceux d'un e-retail particulièrement dynamique, font appel à des architectures hyperscale pour gérer les pics de connexion lors de soldes majeures comme le Black Friday, mais aussi pour faire face aux pics d'activité résultant de la pandémie de COVID-19.

Les data centers hyperscale souffrent néanmoins d'une carence : leur sécurité. Pourquoi ? Parce que l'activation de la sécurité crée souvent un goulot d'étranglement dû aux performances atones des outils de sécurité en place et à une segmentation déficiente de l'infrastructure informatique. Avec des pare-feux réseau peu performants, la sécurité devient un frein et non plus une priorité pour les professionnels de l'ingénierie ou des opérations. Et c'est l'entreprise qui devient vulnérable à diverses attaques susceptibles de lourdement impacter ses opérations.

La sécurisation des data centers hyperscale nécessite une stratégie actualisée et basée sur une technologie pertinente.

Introduction

Les innovations digitales et la demande des entreprises ont changé la façon dont les data centers sont utilisés mais aussi les objectifs de performances qu'ils doivent garantir. La réponse à la demande pour de nouvelles capacités réseau a fait émerger les data centers hyperscale.

Un data center devient hyperscale lorsqu'il fait preuve d'une évolutivité optimale et qu'il s'adapte de manière dynamique pour répondre aux exigences toujours plus pointues des entreprises. Les architectures hyperscale sont conçues pour répondre à des exigences sans précédent en matière de capacité et de performances. Ces exigences peuvent varier d'un secteur à l'autre. Voici quelques exemples d'activités qui nécessitent des architectures hyperscale :

- **Les grandes entreprises, et notamment les fournisseurs de services Cloud.** Les entreprises qui misent sur la virtualisation pour définir des réseaux virtuels particulièrement évolutifs doivent segmenter à grande échelle ces réseaux pour constituer des VXLAN et accélérer les communications entre les services qui sont co-hébergés sur des plateformes physiques et virtuelles.
- **E-commerce temps-réel, dont l'e-retail.** Les pics de connexions liés à des événements tels que les soldes, les périodes de déclaration d'impôts en ligne ou les actualisations de situation pour percevoir des allocations chômage, imposent de pouvoir gérer un volume majeur de sessions d'utilisateur par seconde.[1]
- **La recherche de pointe dans les industries pharmaceutiques, du pétrole et gaz et de l'aérospatial.** La R&D de pointe tire parti du big data et d'algorithmes de machine learning et nécessite, à ce titre, la possibilité de faire transiter des flux massifs de données (appelés elephant flows) à des performances de 40 Gbps et 100 Gbps.[2]
- **Les marchés boursiers.** L'infrastructure des salles de marché exige que les données soient reçues avec le plus faible temps de latence possible.[3]
- **Hyperscalers (acteurs technologiques de premier rang).** Les interconnexions haut débit entre les data centers dans le Cloud, pour répliquer les données sur des sites dédiés à la reprise après sinistre (DR), nécessitent des interfaces et des capacités de tunneling IPsec à haut débit pour assurer la confidentialité et la protection des données.[4]

Dans de nombreux cas, les entreprises opérant dans ces industries ont investi dans une infrastructure réseau adaptée. Cependant, la recherche de solutions de sécurité capables de répondre à ces besoins représente un défi car les pare-feux de nouvelle génération (Next-Generation Firewall, NGFW) ne répondent pas aux besoins des architectures hyperscale en termes de taille et de performance. Ces NGFW existants sont à la peine lorsque les entreprises veulent contrôler des dizaines de millions de connexions d'utilisateur par seconde, ou mettre en œuvre des mesures contre les attaques de déni de service (DDoS) associées à une sécurité par pare-feu sur la couche 4. La dégradation des performances qui en résulte amène de nombreuses entreprises à désactiver purement et simplement les fonctionnalités de sécurité, de peur qu'elles ne ralentissent leurs activités et les empêchent d'optimiser le débit et la latence de leur réseau. Mais cet arbitrage s'avère hasardeux : renchérir ses capacités sans contrôle de sécurité adéquat revient à conduire un bolide à tombeau ouvert sans avoir bouclé sa ceinture de sécurité.

Les défis des architectures hyperscale

Chaque environnement présente des défis pour la sécurité de l'hyperscale.

Difficultés à opérer des services virtualisés et ultra-évolutifs

Les entreprises doivent être en mesure de lancer des services le plus aisément possible afin d'accroître leur productivité et leurs revenus. Pour maximiser le retour sur investissement (ROI), les services doivent être interopérables avec des ressources physiques et virtuelles.

Avec des technologies particulièrement évolutives comme le VXLAN, les clients peuvent segmenter tous les services virtualisés et atteindre un niveau d'évolutivité que ne leur offre pas le VLAN. Les services virtualisés peuvent être étendus, réduits et migrés sans coûts d'exploitation majeurs. Ces services sont souvent nécessaires pour communiquer avec d'autres services de l'infrastructure physique existante. Cependant, la plupart des solutions actuelles souffrent de performances atones et d'une latence élevée. Elles ne disposent pas de la sécurité essentielle au niveau de la couche 4 pour suivre les sessions, ni d'un contrôle des accès. Elles n'offrent pas non plus cette sécurité évoluée au niveau de la couche 7 qui permettrait de mieux détecter les menaces et d'appliquer les règles de mise en conformité et de gestion des risques.

Les pics de connexions liées à des événements submergent les infrastructures de sécurité peu "élastiques"

Dans d'autres secteurs, le volume des connexions est inférieur au nombre total des connexions qu'une entreprise doit être capable de traiter dans un délai très court. Lors des dates de grandes soldes, comme le Black Friday et le Cyber Monday, les sites d'e-commerce connaissent un trafic extrêmement élevé sur 24 heures, un trafic jusqu'à 1,5 fois plus important que celui des autres jours de soldes.[5]

Des pics de connexion similaires se produisent pendant la période de déclarations fiscales, à l'ouverture de la billetterie d'un événement majeur, ou lors de fêtes telles que le nouvel an chinois. Mais aussi dans le monde des jeux en ligne, en particulier les jeux multi-joueurs où des centaines de personnes se retrouvent simultanément et occasionnant des pics de connexion.

En réponse, les architectures hyperscale permettent à l'administration fiscale, aux distributeurs et aux jeux en ligne de traiter efficacement des millions de connexions entrantes par seconde. L'argument commercial en faveur d'un investissement dans l'hyperscale est simple : les connexions perdues ou la lenteur des réponses peuvent entraîner une perte de revenus et ternir la réputation de la marque. À titre d'exemple, un retard d'une à trois secondes en moyenne dans le temps de chargement des pages entraîne une augmentation de 32 % du nombre de clients qui abandonnent le site.[6]

Les flux massifs sur le réseau sont vulnérables aux attaques

Les applications d'intelligence artificielle (IA) et de machine learning nécessitent des volumes particulièrement importants de données, atteignant souvent plusieurs téraoctets [7] pour élaborer et tester les algorithmes. Les industries pharmaceutiques, biotechnologiques, génomiques et des hydrocarbures ont toutes besoin de ce big data pour leurs recherches. Pour traiter et analyser ces données, les départements de recherche doivent être en mesure de transmettre efficacement ces mégadonnées sur leur réseau. Et la transmission efficace de ces méga-données nécessite un débit pouvant atteindre 100 Gbps.

Théoriquement, pour bénéficier de telles performances, les instituts de recherche devraient être en mesure d'exploiter des architectures réseau hyperscale construites sur des routeurs et des switchs. Toutefois, ces dispositifs ne suivent pas l'état des sessions et n'offrent pas de sécurité de la couche 4. Par ailleurs, ils sont vulnérables à des attaques DDoS toujours plus nombreuses.

Qui plus est, les données qu'ils transmettent sont souvent sensibles et réglementées par des lois, comme le règlement général de la protection des données (RGPD) de l'UE. Ce cadre réglementaire impose différents contrôles des accès, ce qui nécessite de faire transiter le trafic réseau via des pare-feux et le chiffrement des messages. Pour autant, la plupart des NGFW (pare-feu de nouvelle génération) ne peuvent gérer de débits supérieurs à 10 Gbps, ce qui ralentit considérablement les travaux de recherche, mais empêche également les entreprises d'optimiser le retour sur investissement de leurs liaisons WAN existantes, acquises pour transmettre des données à 40 Gbps et 100 Gbps alors que les NGFW en place ne supportent que 10 Gbps.

La latence des pare-feux peut entraîner de lourdes pertes financières

Pour les opérations boursières, les jeux en ligne et les activités similaires, la maîtrise de la latence du réseau est essentielle. Même les microretards dans les allers/retours du trafic réseau peuvent avoir un impact significatif sur la rentabilité ou les performances.

Par conséquent, les sociétés financières investissent généralement dans des infrastructures de réseau qui offrent une latence extrêmement faible pour leurs data centers, inférieure à 5 µs.[8] Dans les contextes sensibles à la latence, de nombreuses sociétés configurent leurs NGFW en mode surveillance, sacrifiant ainsi la sécurité au profit des performances du réseau.[9]

Les performances des data centers à haut débit nécessitent une connectivité IPsec à haut débit

Pour les prestataires de services Cloud et les sociétés exploitant des réseaux de fourniture de contenu (CDN), la réplication des données sur plusieurs sites régionaux est essentielle. Ces sociétés utilisent ces sites en local pour héberger des copies complètes des données stockées, renforcer la résilience de leur infrastructure, diminuer la latence des réponses aux demandes clients et alléger les charges du data center principal.

Pour ce faire, les sociétés ont besoin d'interconnexions entre les data centers et de liaisons haut débit entre les sites régionaux pour assurer la synchronisation du réseau.[10] Comme les prestataires de services Cloud et les réseaux CDN transmettent souvent des données sensibles ou privées, les liaisons sont souvent déployées via des tunnels IPsec. Mais, dans le même temps, la sécurité réseau de la couche 4 exige que les NGFW traitent le trafic IPsec avec le même débit que les liaisons du réseau. Comme la plupart des NGFW existants ne peuvent offrir des performances IPsec supérieures à 10 Gbps, le transfert de volumes importants de données entre les data centers est ralenti.

Conclusion

Les efforts d'innovation, destinés à améliorer l'efficacité et l'expérience clients, font évoluer l'infrastructure réseau. Les data centers hyperscale sont conçus pour supporter des flux massifs de données, des pics de connexion et bien d'autres scénarios.

Si de nombreuses sociétés ont déployé une architecture réseau hyperscale, assurer la sécurité à très grande échelle est plus difficile. Désactiver les NGFW ou les placer en mode de surveillance, pour éliminer les goulots d'étranglement sur le réseau, laisse la porte ouverte aux attaques et rend la conformité réglementaire aléatoire. Ne pas segmenter les applications et l'infrastructure facilite la propagation d'une attaque jusqu'au cœur du réseau, suite à une simple intrusion en périphérie. Et les conséquences sont bien pires si les attaques proviennent d'utilisateurs internes et de confiance.

Les data centers hyperscale exigent de repenser l'approche aux solutions de sécurité, une approche qui doit s'adapter aux demandes évolutives des entreprises. Une solution de sécurité hyperscale doit pouvoir gérer des connexions d'utilisateurs à très grande échelle, traiter des dizaines de millions de sessions par seconde, prendre en charge des flux massifs de 100 Gbps, segmenter efficacement des environnements virtuels d'envergure, protéger l'edge réseau de l'entreprise via à une sécurité L4 performante et prévenir les attaques DDoS. Ceci afin de déjouer les cyberattaques qui visent à perturber les activités métier, ternir l'image de marque des entreprises ciblées, voire mettre l'activité de certaines organisations à l'arrêt.

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

[2] Rajiv Kohli et Nigel P. Melville, « Digital innovation: A review and synthesis », Information Systems Journal, 29 janvier 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 septembre 2019.

[4] « What is DCI? » Ciena, 16 mai 2019.

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

[6] « Find Out How You Stack Up to New Industry Benchmarks for Mobile Page Speed », Google, mars 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 septembre 2019.

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

[10] « What is DCI? » Ciena, 16 mai 2019.

décembre 7, 2020 11:36 AM

744553-A-1-FR