The Top Five Risks of Perimeter Firewalls and the One Way to Overcome Them All

Firewalls have long been an integral part of the enterprise network architecture. But with the shift to digital business models, the once-sturdy firewall has gone from a security staple to a security risk. Here's why.

In a traditional perimeter-based security architecture that leverages firewalls and VPNs, security is constrained to the perimeter, or zone of trust. Any user or application inside the perimeter or trust zone is considered good, and those outside are considered bad. This worked well when most users and applications were inside the perimeter. Anyone outside the perimeter had to be brought onto the network by extending the perimeter to them. This was treated as an exception to include them in the trusted zone.

Business has changed significantly since the introduction of the firewall. Today's enterprise employees can work anywhere and everywhere—at a home office, in shared workspaces, at branch offices, and beyond—as long as there's an internet connection and a power source. Extending the perimeter for each remote user no longer works when the exception of distributed users and applications is now the norm. The concept of a trusted zone is no longer relevant because applications and users can now be anywhere which requires switching to a zero trust model. However, firewalls and VPNs are not capable of achieving true zero trust and pose several risks when attempting to do so.

In this whitepaper, we'll detail the five major risks brought about by firewalls in a cloud and mobile world and how to relegate them to the past with a modern zero trust approach.

RISK #1

Large Attack Surface

An attack surface is the sum of all exposed points, such as IP addresses, that may allow adversaries to discover vulnerabilities, route to them, and exploit them to access a system and extract valuable data. Simply put, the smaller the attack surface, the harder it is for attackers to gain access. But the distribution of applications in the cloud and a mobile workforce has exponentially expanded the attack surface leaving organizations more vulnerable than ever before. The use of perimeter-based physical and virtual firewalls not only fails to solve this problem, but also makes things worse by increasing your organization's attack surface, which enables cybercriminals to gain a foothold into your network or cloud instance.

How? Firewalls publish the IP addresses of your servers and applications to the internet so that they can be found by your employees and partners—but this means that attackers can find them too. Every internet-facing firewall, whether it's in the data center, cloud, or a branch office, can be discovered, attacked, and exploited. Virtual firewalls are just as risky as their physical counterparts, because they, too, expose IPs to the internet, often in a much larger number than physical firewalls, increasing the risk further.

How to eliminate the attack surface with zero trust

Successfully eliminating your attack surface is the secret to securing your network, applications, and—most importantly—your data. A true zero trust offering makes applications non-routable entities which are invisible to potential attackers, so your resources can't be discovered on the internet. A true zero trust platform puts itself between the user and application so all communication goes through the platform, and nothing reaches applications without the platform allowing it.

This approach is fundamentally different from perimeter firewalls because only inside-out connections are allowed versus the traditional outside-in approach that requires addresses to be published. By making apps invisible to adversaries and accessible only by authorized users, the attack surface is practically eliminated, and access to applications—on the internet, in SaaS, or in public or private clouds—is always secure.

Detecting attack surface

Attack surfaces can be hard to find manually, but services like the internet attack surface analysis provide visibility into overall attack surface, uncovering the servers, namespaces, vulnerabilities, and cloud instances that are currently visible to the internet. The assessment queries public sources to reveal any areas of exposure that put you at risk. This way, organizations can assess attack surfaces, analyze them, and eliminate them with zero trust.

RISK #2 Choking Application Performance

Users have grown to expect a certain standard of responsiveness and uptime from the cloud applications they use in their personal lives. However, employees often experience a significantly reduced experience when accessing corporate applications using the company's network access solutions, because they no longer have fast and direct access to the cloud applications. In fact, users lose productivity and the ability to effectively collaborate with their counterparts due to lagging application performance. This compels many users to bypass security controls, which is particularly risky when people are using unmanaged devices or unsecured Wi-Fi and home networks. End-user performance issues also arise due to SaaS or cloud application availability, device capacity, network path outages, or network congestion which cannot be easily isolated and diagnosed by the operator.

Why? The "hub-and-spoke" network architecture requires remote and branch offices to connect back to the central office (data center) through firewalls with MPLS and to remote users with VPN. This architecture creates a flat network that extends to all locations, requiring all network traffic to flow to a central security stack. Sending traffic from a remote user through the data center and out to the cloud before returning to the user, and following the same path in reverse, significantly increasing latency, hence degrading the user experience. Virtual firewalls in the cloud suffer the same fate as the traffic has to be redirected to them in the same way that it is for physical data centers since they are not inline with the application servers.

Cloud applications were designed to be accessed directly, with the fewest hops possible, to maximize performance. As such, many SaaS application vendors (such as for Microsoft 365) specifically call out that firewalls should not be put in their path to be fully supported.

How to overcome performance problems with zero trust

A zero trust architecture moves away from the traditional hub-and-spoke network and castle-and-moat security. It provides direct connectivity with applications and reduces risk while providing a better user experience.

An effective zero trust platform enforces policy inline, at the edge, so no extra hops are needed, and direct peering with application companies enables direct connection based on availability and capacity. By operating in the data path, a zero trust platform can also monitor every connection and automatically pinpoint and remediate performance issues. This capability is crucial for low-latency applications, such as Unified Communications as a Service (UCaaS) applications like Microsoft Teams and Zoom. The ability to monitor these applications and remediate issues quickly with Digital Experience Monitoring (DEM) capabilities allows organizations to identify and resolve problems before users are aware of them, subsequently improving employee collaboration and productivity.

Measuring user experience

User experience can be measured using an advanced monitoring tool that provides digital experience insight to understand, diagnose, and improve user experience issues within your organization. The score helps you identify performance anomalies using machine learning and receive actionable alerts.

RISK #3 High Operational Complexity and Costs

Utilizing firewalls, MPLS, and VPNS, or even virtual appliances is not a realistic approach to implementing zero trust. Managing and deploying perimeter firewalls to deliver consistent security across all users, all applications, all devices, and all locations is too operationally complex and costly. Staffing can not scale to manage perimeter policy deployment, updates, and patches. Hardware and virtual firewalls need to be purchased and deployed for worst case scenarios and backhauling traffic to a single security stack utilizes unnecessary bandwidth and security capacity.

Capacity planning calls upon CIOs and CISOs to accurately predict the future to plan for hardware requirements and the bandwidth consumption costs of sending all traffic over MPLS to the data center for inspection. Underestimating the network's needs chokes performance, and on the flip side, overestimating results in unnecessarily high costs and equipment sitting idle. Not to mention, it isn't practical to deploy the exact same stack of appliances at every location, which results in a collection of disparate products strung across your infrastructure. Collecting and curating logs for these many devices is another challenge, and operators often overlook critical logs, leading to a potential security risk. A staggering 75% of operators agree that it is challenging to manage firewall hardware, upgrades and deployments.[2]

And this is just a fraction of the challenge. This fragmented approach requires security personnel to use separate subscriptions and management platforms to implement different policies and manage different zones with network segmentation. Additional effort goes into stitching together visibility by user, application, and location. Employees must focus their efforts full-time on implementing patches, security updates, hardware refreshes, and managing policy across the mismatched collection of firewalls and security appliances. The result is a drain on your finances and productivity that cannot be sustained.

How to avoid complexity with zero trust

Instead of multiple hardware-based solutions or point product cloud solutions that are hard to manage and maintain, an integrated zero trust solution secures all SaaS, internet and private applications with a single platform. Zero trust eliminates the need for costly MPLS networks that need complex routing, switching, network segmentation etc. with fast, secure, direct-to-cloud access, and secure cloud-to-cloud connectivity. It essentially eliminates the need to backhaul traffic to the data center for inspection. A unified zero trust platform with a single management console is much quicker to configure, easier to manage, has simplified policies, and offers more security than traditional perimeter security.

A cloud-based zero trust solution places security controls where the users and applications are: in the cloud. With visibility across all users, clouds, and workloads, zero trust simplifies operations and troubleshooting. Transitioning to the cloud reduces the burden on the IT team to purchase, manage, maintain, and oversee firewalls and other security hardware, opening up additional time to focus on other projects. More importantly, a cloud based zero trust solution makes it easy for enterprises to scale quickly as the volume of users and applications increases.

Cost sensitivity

The 2021 VPN risk report survey concluded that high cost of security appliances and infrastructure was the second biggest challenge organizations face with their remote access solution. Organizations that have adopted zero trust through the Zero Trust Exchange have received 139% ROI and $4.1M in benefits on average with increased productivity, reduced incidents, and reduced appliances.[3]

RISK #4

Lateral Threat Movement

Attackers use a variety of means to gain access to an organization's network, often through phishing attacks or malware infections. Once on the network, their goal is to move laterally through the organization looking for access to sensitive data to exfiltrate, encrypt it for ransom, or cause other disruptions. Lateral movement allows an attacker to avoid detection and retain access, even if discovered on the machine that was first infected. And with a long dwell time, data theft might not occur until weeks, or even months, after the original breach.

Organizations have relied on a "castle-and-moat" security approach—also known as "perimeter security"—to protect data from malicious attacks. Like medieval castles protected by stone walls, moats, and gates, perimeter security invests heavily in fortifying network perimeters with firewalls as well as other tools. Perimeter security guards the entry and exit points to the network by verifying the data packets and the identity of users that enter and leave the organization's network, and then assumes that activity inside the hardened perimeter is relatively safe.

Traditional security architectures are incapable of stopping these sophisticated attacks because once their user, good or bad, enters a "secured" network, they become trusted users and gain lateral access to all applications even if they shouldn't. Reducing east-west lateral movement in perimeter-based architectures requires network segmentation (internal perimeters), which is an operational nightmare as it requires organizations to deploy and manage more firewalls with more policies without properly solving the underlying problem.

How to overcome lateral movement with zero trust

Zero trust prevents lateral movement as it connects users and workloads to applications directly—never to the corporate network. This means that threats cannot propagate laterally to infect other devices and applications, without the need for complex network segmentation. This doesn't just apply to users accessing applications but can apply to all connections within the organization from IoT machines to applications talking to each other where an application in one location (cloud or data center) can securely connect to another application wherever it resides. Thanks to these secure one-to-one connections, the risk of lateral movement is eliminated.

The zero trust model begins with the assumption that everything is hostile, and only establishes trust based upon identity and context. This approach authorizes connections based on knowledge of the entities connecting and the context of their connections ensures that access is limited to what is needed only, at all times. This removes a significant burden from security and IT teams because this happens automatically and can dynamically change when the conditions change for those entities and their connections.

Finally, zero trust provides granular controls with conditional access. An admin can configure policies so that users can access certain applications only if their traffic originates from a trusted location, such as a corporate network, and users have provided multifactor authentication. The admin can also block user traffic originating from certain locations or geographies, from an untrusted device, or if the data being requested is beyond a user's specific permissions. All connections are based on context and as context changes, trust is reassessed.

The new segmentation

*The cost, complexity, and time involved in network segmentation using virtual firewalls outweighs the security benefit. **Workload segmentation** is a new way to segment application workloads. With one click, security can be enhanced by allowing workload segmentation to reveal risk and apply identity-based protection to workloads without any changes to the network. The identity-based workload segmentation technology provides gap-free protection with policies that automatically adapt to environmental changes.*

RISK #5 Data Loss

Data is crucial to organizations for strategic, financial, and security reasons, among others; and in some cases it can be critical to national security. Even with network security perimeters in place, data can be leaked due to lack of awareness, unintentional user actions, system glitches, and increasingly sophisticated malicious activities. This can cause a range of problems, including fines, customer loss, legal ramifications, regulatory noncompliance, and damage to the company's brand. Let's look into different types of data and how they could be at risk:

- **Data in motion:** Data in transit via the internet constitutes most of the data in motion today as applications are now primarily accessed through the web; this is true for SaaS applications, apps in the data center and those in public clouds. When users access the internet and risky destinations where sensitive information can be exfiltrated, it is a threat to enterprise data. Firewalls can't follow users off-network or secure their critical web traffic in motion. Less data and fewer applications remain on endpoints, placing more importance on securing the data that flows between endpoints, cloud applications, and storage with a data-in-motion solution.
- **Data at rest:** Data residing within data centers, SaaS applications, and public clouds accounts for the vast majority of data at rest. In particular, securing data at rest in SaaS apps is critical for security; even if secured with firewalls, it only takes a few clicks to share data with an unauthorized user through apps such as Microsoft OneDrive. Additionally, cloud breaches can be caused by dangerous misconfigurations or permissions. As SaaS and IaaS are highly dynamic and often configured by individuals who are not security experts, such gaps are frequently overlooked and exploited.

The ultimate goal of any security technology is to protect sensitive data, but firewalls are not capable of efficiently identifying and controlling data in motion or at rest, and that puts the organization's data at risk. Most importantly, they are unable to inspect encrypted traffic effectively—more than 90 percent of all traffic[1]—allowing SSL/TLS-encrypted traffic to pass uninspected.

How to avoid data loss with zero trust

A true zero trust platform can inspect all traffic, both on and off the network, including all encrypted traffic. It closes gaps in visibility and inspection to provide effective data loss prevention (DLP) and cyberthreat protection. It is able to decrypt all data, determine the sanity of data, and then authorize connections using context such as a user, geolocation, IP address, device posture, time of the day, etc. DLP policies of a zero trust solution protect data in motion, while users everywhere get fast, consistent security.

An inline zero trust solution delivers full shadow IT discovery and control. It secures web-based threats and data with browser isolation that enables unmanaged device access without the performance challenges. Browser isolation works by streaming data as pixels from an isolated session in a containerized environment, enabling BYOD, but preventing data loss via downloading, copying, pasting, and printing. Out-of-band DLP and advanced threat protection (ATP) remediate risky file sharing and malware at rest in the cloud. To protect cloud data, it also remediates potentially fatal misconfigurations, compliance violations, permissions, and entitlements. In short, zero trust provides consistent, unified security for data at rest and data in motion—including encrypted traffic—across internet, SaaS, and public cloud applications at scale irrespective of the user device.

Web browser vulnerability

The percentage of encrypted web traffic on the internet has steadily increased from 50 percent in 2014 to a staggering 95 percent today.[1] Even then, web browsers are the top target for attackers, as 98 percent of attacks are carried out over the public internet and 80 percent of those attacks target end-users through browsers as per Gartner. Browser isolation tools such as Cloud Browser Isolation help mitigate these vulnerabilities—particularly where they can be deployed without client-side software installations, making them a better fit for unmanaged devices accessing corporate IT resources.

Achieve True Zero Trust with Zscaler

Zscaler delivers zero trust with the Zscaler Zero Trust Exchange, a cloud-native platform—operating across 150 data centers worldwide—that leverages the largest security cloud on the planet to provide fast and secure connections and allows your employees to securely work from anywhere, on any device, using the internet as the corporate network. Unlike firewalls and VPNs, the Zero Trust Exchange is founded upon the principle of least-privileged access, and the idea that no user or application is inherently trusted. Instead, connections are authorized based upon the user's identity and context, including user location, device security posture, application being accessed, and content being exchanged.

How? The Zero Trust Exchange starts by terminating the connection to enable deep content inspection—including encrypted traffic, executing deep data and threat analysis. It then determines identity and device and verifies access rights using business policies based upon context; including user, device, and application being requested and the type of content. Once the business policy is verified and enforced, the Zero Trust Exchange brokers the connection between the intended resources. Users and devices are connected directly to applications, never to the corporate network.

Learn more

To learn more about zero trust and how Zscaler can help, visit the **Zero Trust Exchange** page.

Sources

[1] *Google Transparency Report https://transparencyreport.google.com/https/overview?hl=en*
[2] *Zscaler Networks Security Survey 2020*
[3] *ESG Economic Validation Study 2021*

About Zscaler

Zscaler enables organizations to securely transform their networks and applications for a mobile and cloud-first world. Zscaler connects users to applications and cloud services, regardless of device, location, or network, while providing comprehensive security and a fast user experience. All without costly, complex gateway appliances.

zscaler™