replify
FLEXIBLE NETWORK ACCELERATION

*Replify Whitepaper*

# Zero Trust Network Access ***Acceleration***

## Overview

In this whitepaper, we look at the emergence of ZTNA; it's position in the network; the network performance challenges organisations may face adopting ZTNA; and how a Replify OEM partnership could add value to your solution.

## Audience

Targeted at ZTNA or network vendors who are looking to take their product offering to the next level.

## Executive Summary

Introducing a ZTNA solution across an organisation can present new challenges for branches and users that have limited bandwidth. Where cloud access is required for proxying protected data, ensuring reliable, fast access to the cloud ZTNA provider is key to a positive user experience.

The Replify Accelerator can accelerate traffic to a ZTNA gateway or connector, right from the end-user device. Replify offers its full WAN acceleration suite as an OEM option. Visit our website for more information and to find out how you can get started today with a free trial: www.replify.com/ztna

# The Zero Trust Paradigm vs Perimeter Security

Traditionally, protecting an organization's digital resources meant defining a perimeter around them. To get through this perimeter, possession of a key and locality to the resource would be all that is required. The key would take the form of a username and password. Sufficient locality to the resource could be established by being on the same network as the resource. Access to that network may have been possible only if you were in the same branch office as the resource. To get into that office, you may need to show personal ID at reception, or perhaps be in possession of a physical key card for the access to the building or even server room. Pretty secure, right?

That model served well for a while, with understood risks. Social engineering, malware and viruses on USB sticks, disgruntled employees, were all exploits that organizations needed to be on the lookout for to ensure their resources stayed protected. Things change when the resources need to be accessible from outside the room, building, or another network, perhaps halfway around the world. The perimeter here naturally must extend, become wider, and thus harder to protect. Potentially more individuals require access, at any time or the day, and perhaps from anywhere on the planet. This results in an explosion in the attack surface, presenting opportunities for bad actors to threaten the sensitive data.

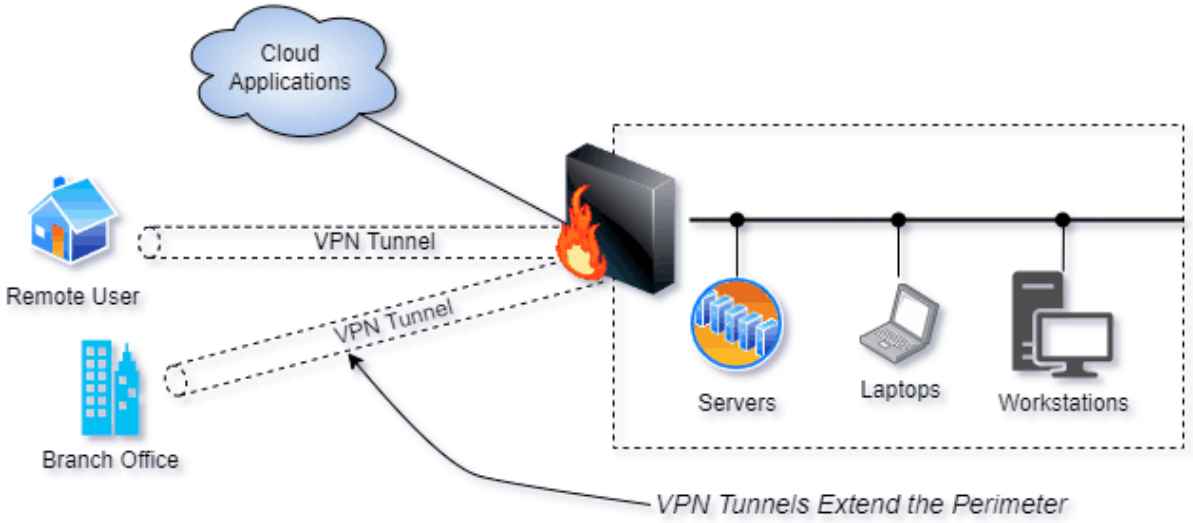


*Figure 1. With increased remote working and communication between sites, the perimeter has become difficult to define.*

The very definition of the 'perimeter' in the traditional sense is difficult in today's world. Many questions arise. Can cloud services be part of the perimeter? Do people's homes become part of the perimeter? Does there need to be a perimeter covering my Office 365 users and those accessing the file server in a branch office? What if someone inside the perimeter is malicious? Is my firewall vulnerable to attack? Can my VPN credentials be stolen?

# Zero Trust Network Access (ZTNA)

ZTNA brings the concept of *zero trust* to all an organization's applications. The focus is no longer on the nuts and bolts of networks – IP addresses, hostnames, and ports. Instead, a user requesting access to a service is authenticated, and authorized specifically for that service. Additionally, this authorization is continually verified.

With ZTNA, continual verification is of utmost importance. A user that is trusted one day, may not be trusted the next. This was harder to enforce with perimeter security, as again, once the user is within, they are trusted by default.

A point-to-point access solution, such as a traditional VPN, is not concerned with what actions are performed once a user is connected. Its role is to enable and secure a connection between one network and another. The VPN is still fit for this purpose, but the needs of organizations are much more complex than just securing a point-to-point connection.

A ZTNA client application on the end-user device can capture requests to a service. The service might be an email server in the cloud, a SharePoint site, or a file share in a data centre or office. The client application typically communicates to a controller in the cloud. This controller is where the organization has defined rules. The rules set out who has access to what, when and from where. Further, the rules can define what software must or must not be running on the client device. For example, access may only be allowed for users on systems with a particular set of anti-virus definitions.

On the end user device, when a request is made for a service, the client will authenticate with the controller. This is often done via the sharing of certificates. The client will check that it trusts the controller, and the controller will check the client's certificate. Additionally, the controller may decide that, for this service request, two-factor authentication is required. The controller will also check the client's security posture and, if satisfied, the client will be allowed to access the gateway to the service. This gateway sits between the client and the service and acts as the 'data plane'. Traffic is fed from the client to the gateway, and then from the gateway to the service.

# Proxying and Tunnelling

Typically, there are two connections when accessing a service with ZTNA: the connection to the controller; and the connection to the gateway. There may be multiple gateways, each protecting a different set of services, at potentially several locations. For example, there could be a gateway in each branch office where services are located, as well as in data centers and the public cloud.

A connector component will often sit between the gateway and the service and will maintain a connection with the gateway. Again, this means the applications are invisible, with no ports being opened on the service side as the connection from the connector to the gateway is outbound. In this case, the gateway proxies the connection from the client to the connector which then passes the traffic on to the desired service.

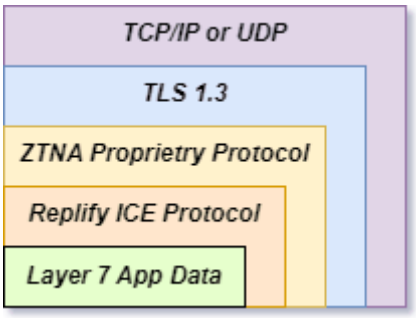


*Figure 2. Multiple Logical Tunnels*

The gateway often handles the connection to the service, acting as a tunnel. Crucially, the gateway will only accept clients that have been permitted by the controller, so the application is not directly accessible. This has the effect of making the application invisible, as it's only accepting connections from the controller. This also simplifies the firewall configuration for the services, exposing themselves just to the gateway, rather than the internet.

Ideally the gateway should be placed closer to the service than the user is. This means the latency between the gateway and the application server is lower than that between the client and gateway and that the bandwidth is higher. Often an end customer would have no control over the placement of the gateway geographically as this is typically owned by the ZTNA provider. For global coverage, multiple gateways might be required to achieve optimal performance.

# The Impact on Networks and User Experience

While ZTNA greatly improves the security of an organization's digital assets, it can come at the cost of user experience. Sacrificing security for performance is a dangerous game and goes against the motivation of embracing a Zero Trust policy. Sacrificing some application performance to increase security is a much easier case to make, but why is there an impact on performance and how significant is it?

The communication with the new components required for ZTNA is key. The controller, gateway and connectors are essential for ZTNA. The position of these components relative to each client, and the services is an important consideration when thinking about the overall performance of the system. Given clients could be widely dispersed geographically, this can be tricky to achieve.

## Controller

The controller is considered part of the management plane. Before a user can establish a connection to a service protected by ZTNA, they need to satisfy the controller that their access is permitted. This initially requires a communication between the ZTNA client and the controller to establish the rules for that client. This typically wouldn't get in the way of the client's connection to the service, so the impact of the communication with the controller on performance should be low.

## Gateway

The gateway is a key part of the client-to-service data plane. This means that application layer traffic will be flowing between the client and the gateway. Further, the gateway then proxies the traffic between it and the connector. The position of the gateway is very important in reducing the impact ZTNA has on application performance and user experience.

*For the gateway not to have a negative impact on performance:*

- It must have as good a connection to the service or connector as the client has. This includes both bandwidth and latency.
- The connection between the client and the gateway must have as good a connection as the client has between the client and the service.

If either of the above do not hold, or, more generally, if routing to the service via the gateway adds extra latency, the client may have a degraded experience using the application. The

application may be sluggish or unresponsive. In the worst case, high latency may make the application unusable.

Ideally, a ZTNA vendor would have many points-of-presence around the globe. However, even this does not satisfy all users, especially those outside metropolitan areas, such as those working remotely or off-site.

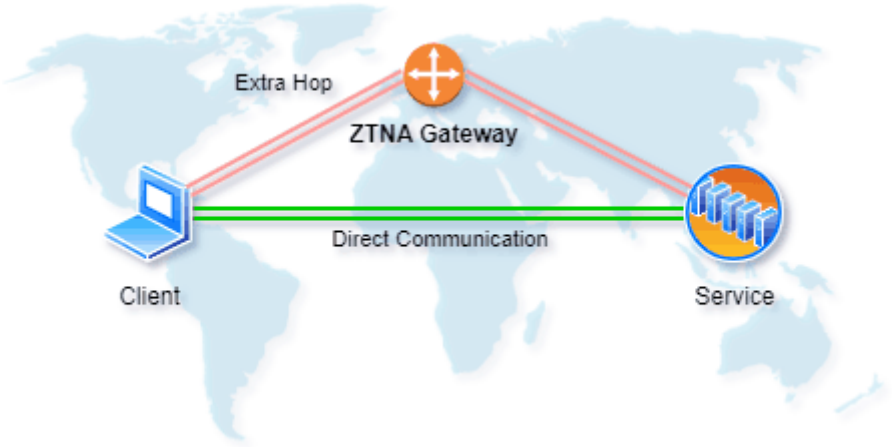


*Figure 3. The gateway introduces an extra hop, with extra processing and tunnelling required.*

# Connector

The connection between the gateway and connector is also central to the performance aspect of the ZTNA solution. If the connector is far from the gateway, or the connection from it to the gateway is poorer than that between the client and the connector, the user experience will be adversely affected.

# Accelerating ZTNA

By default, we expect our connections from the client to the gateway, controller, connector, and services to be secure and free from malicious or unwanted modification, tampering or inspection. Such comprehensive securing of the application data limits what optimization can be applied by external software. Effectively, you are left with transport layer optimization at best. There are gains to be had by optimising the transport layer, but these may not be enough to give the users the best performance while also enjoying the security benefits of ZTNA.

## Where Replify Fits In

The Replify Accelerator Client, along with the Replify Virtual Appliance, brings world-class Layer 7 optimization capability as an integration opportunity. With over a decade of experience integrating with OEMs in a diverse range of sectors, Replify can successfully integrate WAN optimization into a ZTNA solution. The off-the-shelf Replify Accelerator demonstrably improves the user experience of a number of high-profile ZTNA solutions.

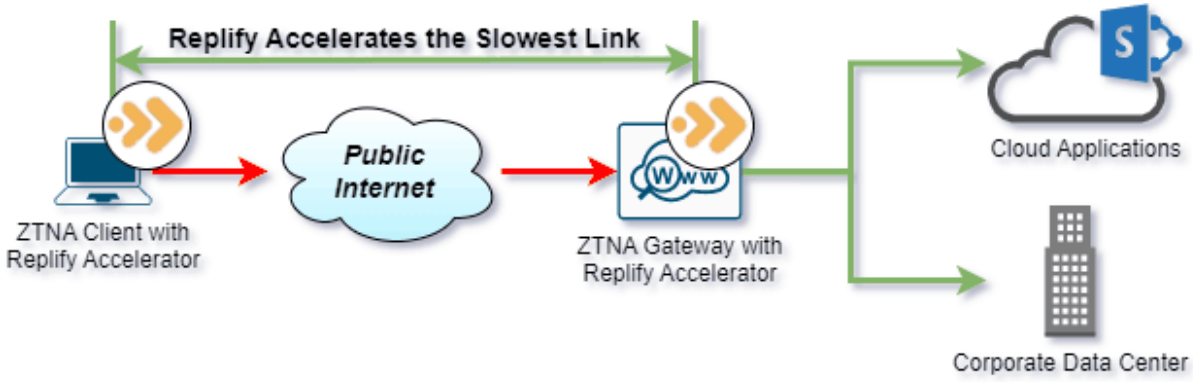


*Figure 4. The Replify Accelerator is running on the ZTNA client, optimizing traffic over the internet to the ZTNA Gateway*

The proxying and tunnelling of data means that applying the 'off-the-shelf' Replify Acceleration to traffic prior to analysis by a ZTNA platform, makes it harder to inspect and apply rules to that traffic. The data becomes opaque, Replify-optimized traffic. This is why a deeper integration is necessary.

## Tight Integration is Key

The position of the Replify Accelerator within the traffic flow is key. The Replify Accelerator needs to intercept the application traffic that is being sent on a TCP connection. The Accelerator is designed to optimize some well-known layer 7 protocols, such as HTTP/2, SMB, and FTP. To be able to effectively apply optimization, the Accelerator must be able to inspect this traffic. When the Accelerator has identified a flow to optimize, the original application data is translated into a bespoke Replify Accelerator optimized flow. This communication happens on a TCP connection between the Replify Accelerator nodes. See figure 5 for an example of the process flow.

For a ZTNA product to apply security functions effectively, a key component is the identification of the service in use, as well as the nature of data that is being communicated. Previous generation firewalls and security products used to look at the TCP connection details (i.e., the IP address and port) to decide whether to permit the traffic. The ability to look inside the traffic is one of the things that makes ZTNA so powerful, so it's important we don't lose that, when combining with WAN optimization.

Further, inspection of encrypted traffic is also important to ensuring an organisation's policies can be applied. The Replify Accelerator also requires access to the unencrypted traffic to provide the best acceleration. While the Replify Accelerator has full TLS proxy capability, performing the encryption/decryption twice, once for the WAN optimization and once for the ZTNA inspection, will reduce performance. This is another reason why combining the functions brings great benefits.

To enable all functions of the ZTNA solution, it's key to combine or build-in the WAN optimization. The ZTNA solution can apply its logic first, then pass the data to the WAN optimizer once security checks have passed. The Replify Accelerator can then perform the caching, compression and optimization and hand this flow back to the ZTNA product for redirection to the ZTNA gateway.

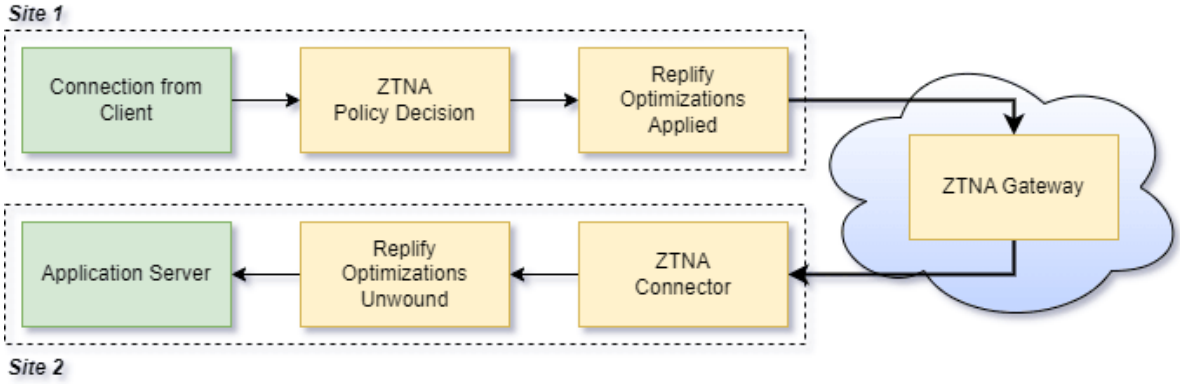


*Figure 5. An example of where the Replify Engine would be added if the entire WAN link is optimized.*

The ZTNA client can include the Replify optimization engine, and direct traffic to it after deciding that the client is permitted to access this resource or service. This optimized data is then sent on to the gateway, where it can be proxied to the relevant connector or service. At the endpoint, either in the gateway, or connector, the optimised traffic can be directed to a Replify Virtual Appliance. Here, it is restored and sent on to the service.

# The Replify OEM Model

Replify's uniqueness in offering the whole solution as an easy-to-integrate option is what gives us the edge over competitors. Offering an accelerated ZTNA solution could be what your ZTNA product needs to stand out.

Replify has a long and successful history of OEM integrations with a variety of vendors in the Visibility and SD-WAN spaces. In Replify lab testing of multiple popular ZTNA products, we've proven the Replify Accelerator can bring a much-improved user experience. That comes from a reduction in bandwidth usage, and a boost in effective throughput resulting from Replify's Intelligent Caching Engine.

Both the **Replify Accelerator Client** and the **Virtual Appliance** share the same optimization code and capability, with both being suited for deep integration into an existing product stack.

An integration of the Replify Accelerator would offer your ZTNA solution:

**Last Mile Optimization**
Optimizes data right from the client to the server with a client-based integration for Windows, Mac, Linux or Android.

**Acceleration of the Slowest Link**
The route from the client to the ZTNA data gateway or access point, without acceleration, is often the slowest, even with hundreds of global PoPs. The Replify Accelerator can even be used between the gateway and the application if that is the focus.

**Over 90% Offload of Data Possible**
The Replify WAN Accelerator's caching, compression and protocol optimization can give over 90% offload of layer-7 traffic. This reduces the load on the network, provides application acceleration and, importantly, improves the user experience.

**Easy Integration**
Built to be easy-to-integrate and cross platform from the start, the Accelerator Intelligent Caching Engine works on many different hardware platforms, hypervisors, containers and cloud platforms.

**Comprehensive APIs**
As part of Replify's focus on flexibility, comprehensive APIs allow the Accelerator to be tightly integrated into your existing management interfaces and traffic processing chains.

**Retain Visibility of Flows**
The Replify Accelerator provides visibility of optimized flows and traffic so you can track optimized connections from end to end, without disrupting other network processing of the traffic.

**Flexible Licensing**
A key tenet of Replify's OEM partnerships is our flexible OEM licensing model. This puts the OEMs in full control of pricing, so they always find an approach that works for their business and customers.

**Access to Replify's Core Dev Team**
OEMs get access to the core Replify development team who have extensive experience of working with network vendors and understanding their needs and products.

**Dedicated Code Stream**
A code stream is created for your integration so bespoke changes can be made, tracked, and maintained. You'll benefit from free updates to the Replify core product applied to your code, for the lifetime of the OEM agreement.

# Conclusion

The Zero Trust security paradigm is here today. While organisations are having to get more comfortable with the fact that the classic perimeter doesn't fit in today's diverse network, the need to overcome new challenges grows. One of those challenges is ensuring business is not disrupted due to an underperforming solution.

Replify is ideally placed to bring their own, world-class WAN optimization capability directly to your solution, giving your product the edge over competitors, and ensuring your customers can enjoy both the optimum performance and security that Replify + ZTNA can bring.

Get in touch now via the website at www.replify.com or contact us at contact@replify.com