zscaler™

White Paper

Architect's Guide to Universal ZTNA

Contents

Objective

The objective of this guide is to provide several reference architectures that illustrate how both Zscaler Internet Access (ZIA) and Zscaler Private Access (ZPA) technologies can be leveraged for LAN (Local Area Network) segmentation to enforce zero trust access for users.

Historically, it's hard to close down the network to control all traffic flows, especially for local-to-local connections. Clients typically need to connect with local resources—like DNS, Active Directory Controllers, and printers—but also should not be able to connect to each other or to management ports that are required to make those resources run. IP-based filters are unsuitable because of DHCP, combined with the fact that different users require different types of access. This usually leads to a situation which, once connected, users have full access to all resources on the network and security is no longer enforced.

The traditional approach to solve this has been to stop unknown/untrusted devices from connecting to the network through 802.1x and MAC Authentication Bypass (MAB). However, this is susceptible to spoofing, it translates poorly over the WAN, and it still doesn't provide least-privileged access. More importantly though, protecting the network against unknown devices only addresses part of the threat.

ZPA provides a single policy framework that's able to control all any-to-local flows, independent of IP or location. This enables user access without requiring users to share network context (or routing domain) with the applications they need to access:

- The user can be anywhere
- The application can be hosted in any location
- There are no IP dependencies (and IP overlap is supported)
- Granular, context-based policies control application access
- Leverage any underlying network infrastructure; use the LAN or the internet
- Geographical distance and application latency determines the optimal datapath

Decoupling the user from the network is the first step toward zero trust security. Applications should be inaccessible unless the user is authorized, and the attack surface should remain invisible even to authorized users.

These characteristics are achieved by the ZPA architecture:

High-level overview of the ZPA traffic flows

The ZPA cloud consists of many globally-distributed points-of-presence called ZPA Service Edges. When the Zscaler Client Connector initiates, it finds and connects to its nearest ZPA Service Edge using an outbound TLS session which is used for signaling control channel and data transfers.

On the application side, the Zscaler App Connector is responsible for connecting to application servers and other internal resources. When it initiates, it finds and connects to its closest Service Edge and retains the TLS session for signaling. When the client wants to talk to an application over ZPA, the ZPA cloud finds which Zscaler App Connector is best suited to handle the connection, after which the elected App Connector builds a local UDP or TCP session between itself and the application server. Note that only data (without any TCP/IP information) is exchanged over the ZPA cloud.

Learn more about ZPA.

Similarly, ZIA is used to secure local-to-internet flows by applying full defense-in-depth security as a cloud solution to users, regardless of where they are, and without adding latency. In addition, access to some SaaS applications like Microsoft Office 365 (M365) can be prohibited when the user is not actively protected by Zscaler.

However, ZPA and ZIA policies are only enforced on corporate-controlled clients with Zscaler Client Connector installed. This is fine for roaming users—it just means that a non-corporate device isn't protected, but isn't able to connect to corporate resources either—but it doesn't solve exposure of the local network. In other words, it's still essential to secure the LAN against unknown devices.

Interestingly, using ZPA and ZIA can help here too. An additional benefit of using ZPA and ZIA is that it makes those connections use well-defined paths: only connections to Service Edges (from clients and App Connectors) and from App Connectors to application servers are expected. Combining Zscaler with local access controls allows customers to make the best of both worlds: global policy enforcement without complicated management and no way to locally bypass these controls. The rest of this paper focuses on how this is achieved.

Network Segmentation Challenges

One of the biggest challenges in network segmentation is the complexity of handling different access permissions. Users, printers, public terminals, cash registers, etc. all share the same infrastructure, but all need different access. Trying to segregate these leads to a high risk of migration failure, high operational cost, and difficulty in troubleshooting.

It becomes even more difficult in environments in which hardware and software versions of infrastructure and connected devices are not uniform; technologies like 802.1X (explained below) require uniform hardware and software to function reliably. Replacing and/or upgrading such environments can be costly and time-consuming.

Mapping and enforcing trust from user to device, to IP, to subnets, and to applications

Scalability limitations affect all information systems, but this varies across the different approaches to network segmentation. Therefore, it is important to select an approach that best fits your strategy and your environment. In this new world of hybrid working environments where people work remotely but still visit offices, we need to ask ourselves what our segmentation approach should be. Should it extend beyond end users and local infrastructure? Should it include traffic policies? Doing so makes a segmentation project even more complex, but limiting the scope diminishes the true value of the project as it only addresses part of the picture. We need an approach that encompasses users and workloads no matter where they reside or travel.

Set up the Zscaler Client Connector to control outbound traffic

Zscaler Client Connector (with ZPA & ZIA) can be used not only to provide secure access to local and internet-based resources, but also enforce policy blocks beyond the scope of existing applications:

- By using ZIA with an IP-based access policy for private IP addresses (like RFC 1918 ranges). This still provides access to local devices, which can be useful for users at home who need to use a local printer.
- By using ZPA with a block policy for RFC 1918 (and other) IP addresses. This also blocks access to local devices to prevent access to all unknown devices.

Both policies can be applied to specific users/groups and when (not) on specific known locations. This is particularly useful to control client access for different users, without relying on local network-based controls (e.g.: when roaming).

Note that the client still needs local (or internet-based) access to DNS in order to initially resolve and connect to the Zscaler Service Edges. Fortunately, once connected the Client Connector can use Ztunnel2 to resolve public hosts, and ZPA to resolve corporate private hosts to prevent DNS leaking and/or hijacking.

This can be configured as follows:

Use ZTunnel2 policy to block RFC 1918 addresses; use ZTunnel2 for DNS:

- Create a Client Connector Forwarding Profile to use ZTunnel2
- Create a Client Connector App Profile and link it to that Forwarding Profile
- Add a ZTunnel2 block policy for 10.0.0.0/8, 172.12.0.0/12 & 192.168.0.0/16 on all ports except 53. Optionally also exclude UDP/67 to optimize DHCP
- Define "*" to forward all DNS through ZTunnel2
- (Optional) Assign this App Profile to those users that need access to local resources when roaming
- Make sure there's no ZPA App-Segment for the RFC 1918 addresses (see below)

Initially, local DNS requests (required for the Client Connector to discover its closest Service Edges) are still allowed, but once ZTunnel2 is connected all public DNS traffic is picked up by ZIA. When active, all IP-based traffic (except L2 adjacent) will be blocked by the ZTunnel2 policy. Since devices on the same subnet can still be accessed it's useful when users should still be able to access local printers when roaming (which won't be accessible through ZPA).

Pro/Con:

+ Allows for L2 access while blocking all other (L3) IP access

– Exposes the client to L2-based bypasses
(like using a local proxy to pick up browser traffic before it is picked up by ZTunnel2)

+ also works for roaming users

+ can be turned on for unknown locations and for specific users/groups

Use a ZPA IP policy to block RFC 1918 addresses; use ZTunnel2 for DNS

- In ZPA: create an app-segment for 10.0.0.0/8, 172.12.0.0/12 and 192.168.0.0/16 on all ports except 53. Optionally also exclude UDP/67 to optimize DHCP
- Create a Block policy for that app-segment
- (Optional) Create a ZPA Forwarding Policy to bypass the segment for users you want to provide local LAN access to (as per above)
- Create a Client Connector profile to use ZTunnel2 to pick up all public DNS (see above)

Combined it means all IP-based traffic (including L2 adjacent) will be picked up by ZPA and blocked. Initial local DNS requests (required by the Client Connector to discover the closest Service Edges) are still allowed, but once connected all public DNS traffic is picked up by ZIA.

Pro/Con:

+ Also blocks access to L2-adjacent hosts

– stops users from using "their own" resources (like printing at home)

o Otherwise the same advantages & disadvantages as the ZTunnel2 setup

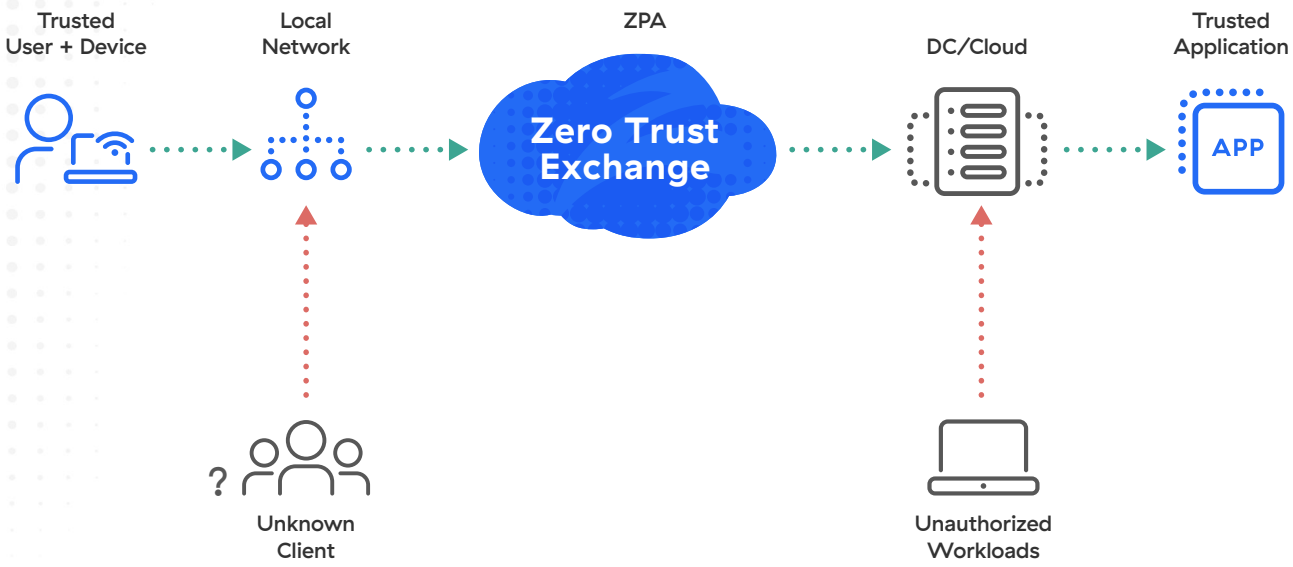
As mentioned, both can be applied per user (or user-group) and based on whether the user is connected to a (un)known location, but the expectation is that the number of exceptions would be limited.

Enforcing Zero Trust on the Local Network

Of course, the policy above is only applied to controlled devices; it doesn't protect against access from devices not using the organization's ZPA solution. But since all authorized connections are well defined, closing down the network becomes much more straightforward once ZPA is deployed and the aforementioned controls are put in place:

- Only DNS and DHCP should be directly accessible
- Otherwise, users should only access Zscaler Service Edges (public or private)
- Other internal resources will only be accessed by App Connectors

This means you can leverage ZPA to create and enforce granular access policies and use simple infrastructure policies only to make sure ZPA can't be bypassed:



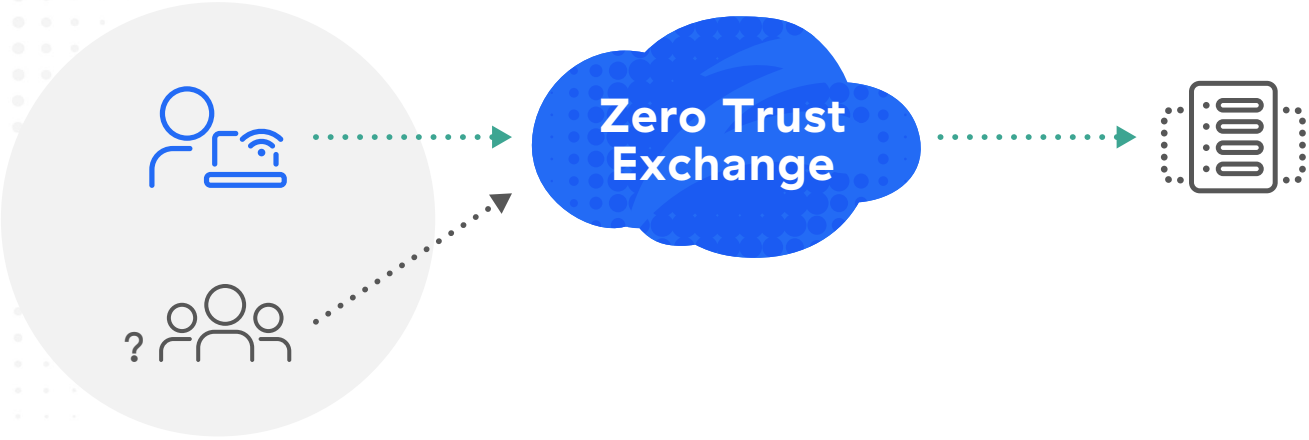
Preventing unauthorized access through the network

To ensure that the Zscaler policies are enforced and no alternative datapath towards on-prem resources exist we can:

- Ensure no resources exist on the local network
- Prevent non-controlled devices to connect to the LAN by using 802.1x
- Segment-off the users from the rest of the network
- Prevent unauthorized connections through inline Path controls
- Control vlan communication
- Prevent unauthorized access through destination controls

Note that in order to apply this universally, both ZIA and ZPA are to be enforced through Client Connector when roaming and when on-prem.

Ensure no resources exist on the local network

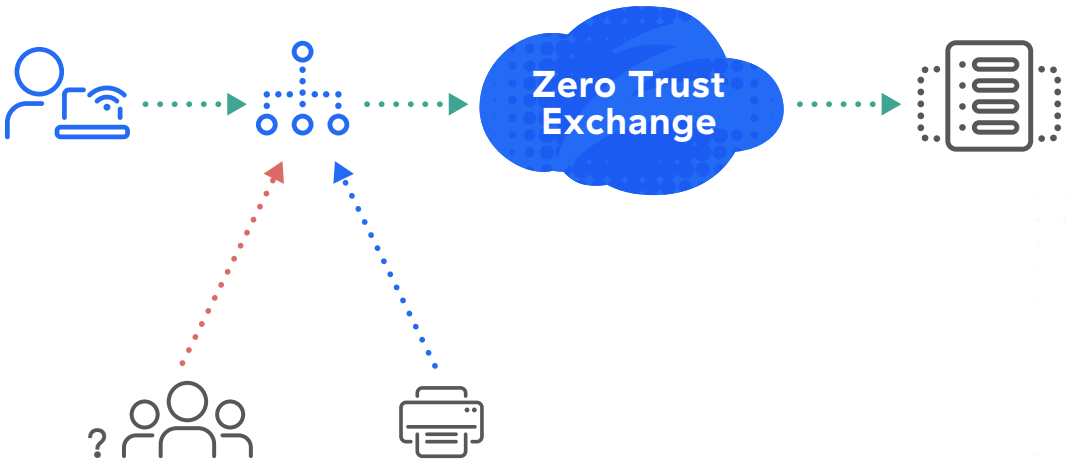


The local network as an extension of the internet

The easiest way to protect against unauthorized access is not having any local resources in the first place. Small (or larger) branch locations without local devices (like printers) don't require any local communication, don't expose any resources and therefore don't need any controls (including firewalls). This is also true for shared networks, as found in shared office spaces.

The only resources that need protection are the clients themselves, and they are protected by Zscaler, identically as when at home. Company resources like Active Directory and internal applications can be reached through ZPA, so the local network doesn't need site-site or SDWAN technologies either. In other words: the only purpose of the local network is to provide internet connectivity, and no additional controls are needed.

Prevent non-controlled devices to connect to the LAN by using 802.1X



Using 802.1x & MAB to block network access of unknown devices

In this scenario 802.1x is only used to prevent unknown devices from connecting to the network. As ZPA is used for user-level granularity, the complexity of combining 802.1x with IP-based access lists, Security Group Tags, and policy enforcement in the LAN/WAN is removed. However, some consideration must still be taken for non-802.1x capable devices (using MAB) which could open up the network to unauthorized access through spoofing. Ideally, access of these devices should still be limited which:

Pro/Con:

+ Stops unknown devices from connecting to the local LAN

+ Integrates w/ ZPA to allow for easily defined local LAN policies

– Requires consistent hardware/software across the LAN

– Requires support on switches and clients

– Requires radius

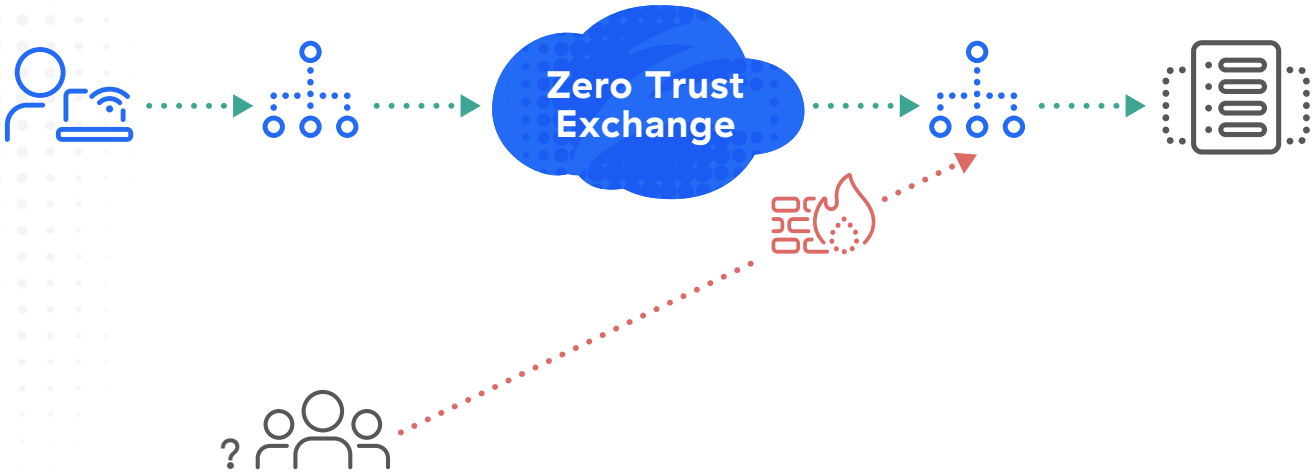
– Requires certificates, user interaction (password), or MAB list

Looking at these disadvantages, combining ZPA with 802.1x is mostly interesting for environments already set up w/ 802.1x:

- By using ZPA as the policy framework in environments where 802.1x isn't set up with access-lists
- By using ZPA as a way to simplify the datapath so existing access lists can be simplified and decoupled from user access (also see the next section)

Also, since ZPA will be responsible for distinguishing between users, Radius no longer has to push user-specific attributes which removes the complexity of managing users in multiple Identity Access Management (IAM) environments. Instead, these attributes are inherited (and maintained) from the primary user management environment (like AzureAD).

Prevent unauthorized connections through inline path controls



Using firewalls to only allow connections from ZPA sources

Traditional firewall policies are very difficult to manage due to the many different source addresses that potentially need access; especially for connections through the WAN. ZPA solves this for user traffic by providing a singular datapath: all user-sourced connections will be initiated by ZPA App Connectors (with DHCP and (initial) DNS being the only exceptions). This means that server-side Policy Enforcement Points (PEPs) no longer need to consider different users which may or may not be mapped to different IP ranges. By the same token, client-side PEPs no longer need to accommodate for different server/application environments; they only need to allow clients to go to their Zscaler Service Edges (private or public).

Control VLAN communication

Defining ACLs can be complex in environments where users, servers, and other network connected devices (like printers, IP-cameras and door-access systems) share the same network. And completely segmenting the network to allow for more simple filters is difficult as it typically involves manually reconfiguring those devices. However, just moving the users to a different VLAN is usually straightforward as it can be done centrally: define another subnet, change/add the DHCP scope, and then change the VLAN on all (user-) access ports. An ACL can then be used to prevent all traffic not involving either App Connectors or ZPA Brokers.

The main challenge will be to distinguish between users and office devices like printers, as they'll share the switch access-ports in the office space. Depending on the organization's risk appetite, printers can be moved along with the users, but they would then still be exposed to non-authorized local access, and would require exceptions in the ACLs.

This situation can be improved by deploying the infrastructure-based Zscaler Edge-Connectors which can pick up this traffic and apply the same ZPA principles to non-user traffic. This is discussed later in this document.

An interesting alternative to segregating users is to make sure there is no intra-VLAN communication at all. This is a fairly standard option in wireless networks but also available on many switches (using port-isolation). Clients can only communicate with devices on specific predefined switch-ports (typically their default gateway) which means there's no complex configuration to distinguish between devices on access-ports; all devices are treated the same. Users can still communicate with local devices, but only through ZPA.

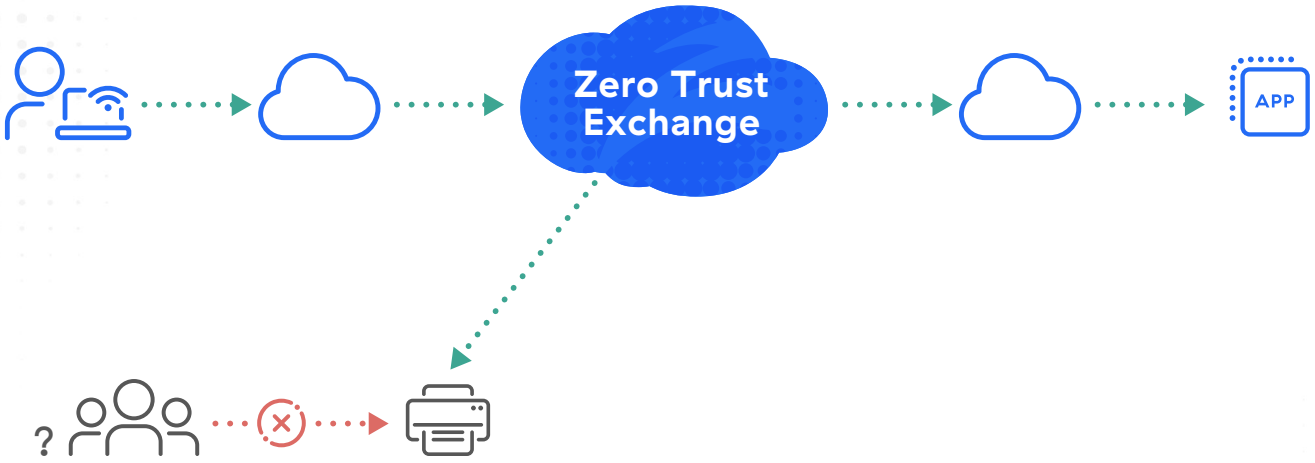


Figure X: Blocking all local access

Again, only very limited ACLs or the use of Edge Connectors are required to prevent all traffic not involving either App Connectors or ZPA Brokers (and DHCP plus DNS); all other traffic passes through ZPA.

Pro/Con:

+ Easy to set up, deploy and manage

+ No IP changes required

+ Covers all client-side use cases

– Requires support on Switches

– Server-server traffic not covered (requires additional segregation)

Prevent unauthorized access through destination controls

Similar to inline path controls, server-side firewall policies (Windows Firewall, IP tables) are extremely hard to manage. Even more so, as local server-server traffic must also be taken into consideration (and ZPA doesn't solve server-server complexity) and in most cases there won't be a way to centrally manage policies for different servers and server types.

Pro/Con:

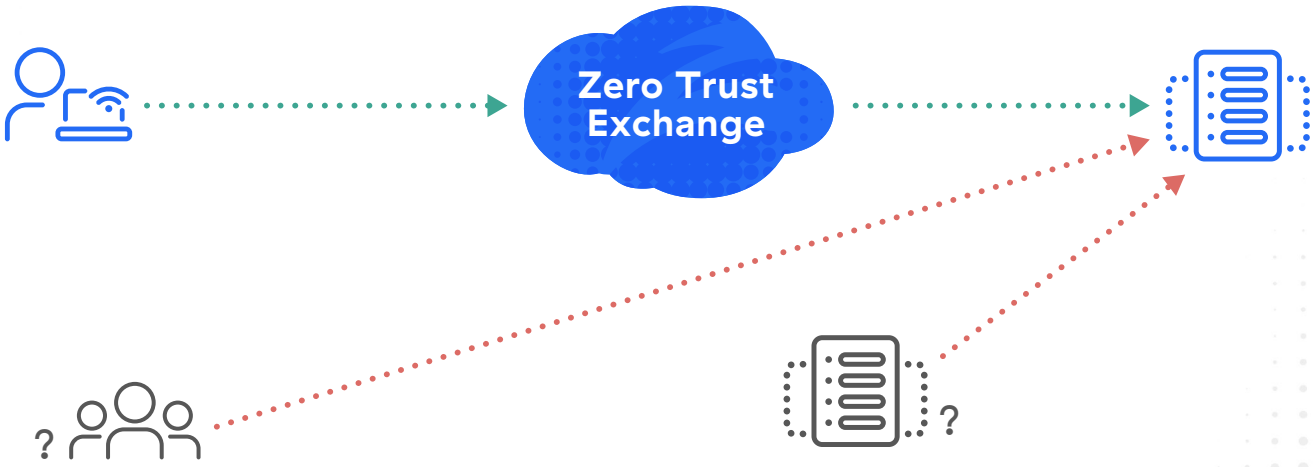
+ Easy to set up (especially at AD joined servers)

+ Works for L2 and L3

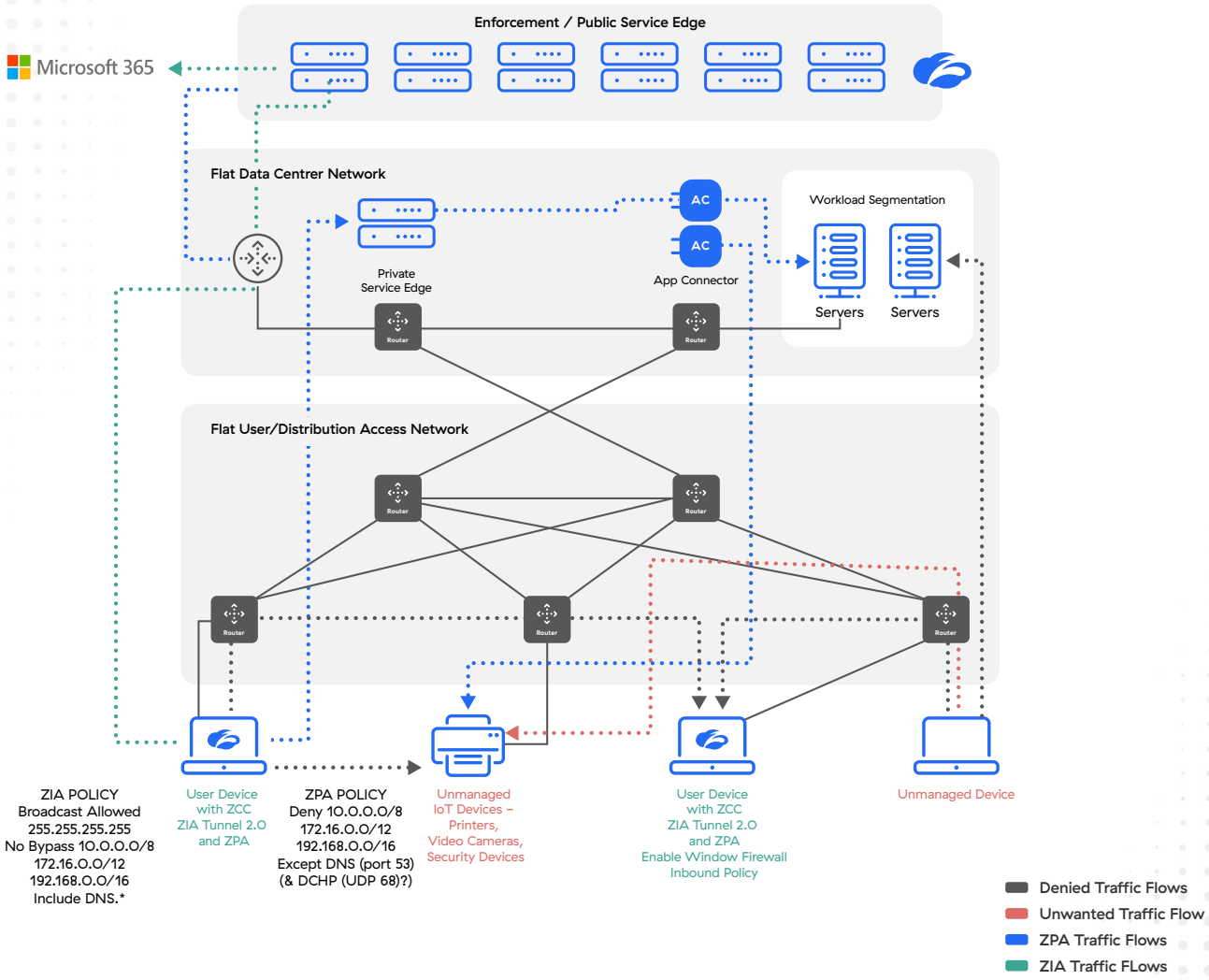
– Hard to manage on non-Windows devices

– Only works for supported servers/devices (not every printer supports ACLs)

Conclusion: only for very homogenic environments (like Windows-only environments)



Zero Trust Connection Flows



Place users in a separate VLAN from the rest (servers, printers, ...)

Segmenting the network is considered hard to do, especially when it involves renumbering servers. However, moving users to a different VLAN is usually straightforward: define another subnet, change/add the DHCP scope and then change the VLAN on all (user–) access ports. An ACL can be introduced to prevent all traffic not involving either App Connectors or ZPA Brokers (similar to #4).

The main challenge is distinguishing between users and office devices like printers, and depending on the risk appetite, printers can be moved along with the users (but they would then still be exposed to non–authorized local access, and would require exceptions in the ACLs)

Pro/Con:
+ typically easy to set up

+ covers most (but not all) of the risk

+ only a single change required (and therefore only impact once)

– still requires some IP–based ACL

– doesn't cover L2–adjacent traffic

– some devices could require exceptions

– server–server and server–other (non–client) traffic not covered (requires additional segregation)

Conclusion: straightforward first step to close down the network from unauthorized access, but requires other security controls inline.

To prevent any traffic inbound to a VLAN/routing domain, either filter inbound TCP–SYN packets (doesn't cover UDP), or use NAT/PAT to hide the entire IP subnet.

Extending Zero Trust beyond the User

Segregating the users from local resources significantly reduces the exposure of those resources. But, as mentioned above, we expect there will be a number of non–user devices that still need to communicate with those resources as well. Instead of opening up the network we should see to which extent we can apply the same controls to them.

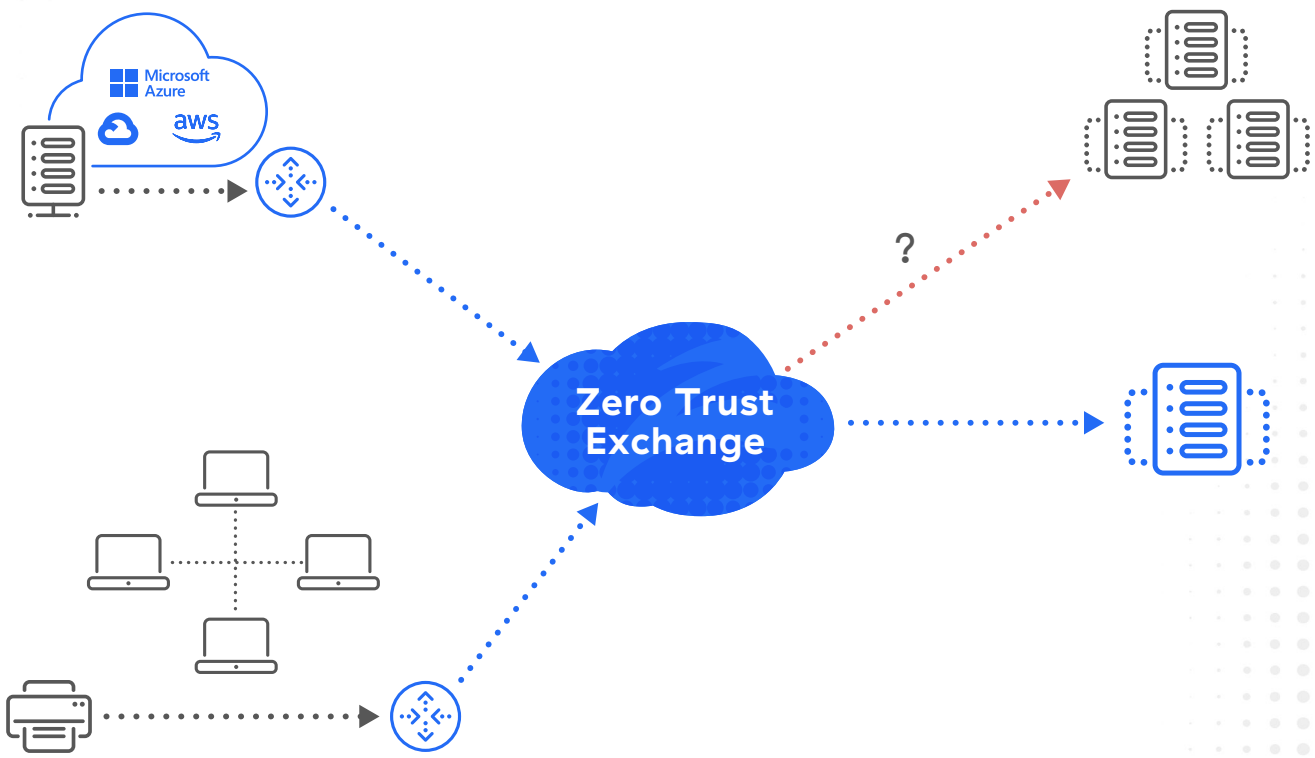
In order to look at the options we must first look at some of the limitations of those devices:

- typically you can't install an agent
- broad range of applications/protocols that don't allow in–the–middle authentication
- there's no user to authenticate, or to base policies on

Without the Zscaler agent, the device can't set up its own connection to the Zscaler Cloud, which is necessary to forward traffic to ZPA destinations. Instead, we need an infrastructure component that does this for us. The Zscaler Branch– and Cloud–Connectors (ie. Edge Connectors) are created to do just that. Similar to the Zscaler Client Connector they receive traffic, determine policy and then block, bypass or forward it to the Zscaler Cloud.

Of course, the lack of agent and user authentication also means that the first two principles of zero-trust (user identity and device posture) can't be applied, but their use does add to overall security. First of all, the ZPA Architecture itself prevents network-based discovery of servers and services, which stops most land-and-expand techniques. On top of that you can define a policy that limits the resources these local devices can access. To avoid this from becoming overly complex to manage and maintain, it's helpful if different devices can still be recognized using server-assigned TAGs (in cloud), integration with Network Asset Management tools (on premises) or through device-based authentication like OAuth.

The Edge Connectors should be placed in the datapath of the non-user devices. Looking at the VLAN-control scenarios above this means they become the default-gateway for all traffic, or for all traffic destined to the central resources. The Edge Connectors themselves follow the same traffic patterns as Client Connectors, which means that only traffic to the Zscaler Service Edges is required



Comparison of Segmentation Technologies

	Multi-VRF	EVN	MPLS	TrustSec	SDA	ZPA
Scalability	Low	Medium	Large	Large	Med-Large	Large
VN Scale	8	32	4K	1K	4K	10K
Operational Complexity	Medium	Low	High	Medium	Medium	Low
Provisioning	Hop-by-Hop	Hop-by-Hop	LDP & mBGP	ISE & dACL	DNAC	Name Space Routing
Encapsulation	802.1Q + IP	802.1Q + IP	MPLS or GRE	IP	VxLAN + IP	TLS
L2VPN Extension	No	No	Yes	Yes	Yes	No
WAN Extension	GRE, LISP, DMVPN	GRE, LISP, DMVPN	MPLSoGRE, L3VPNoMGRE	IP, L3-TF, DMVPN	IP, L3-TF, DMVPN	TLS
Multicast	Native	Native	mVPN or GRE	Native	Native	Not supported
QoS	COS/DSCP	COS/DSCP	EXP	COS/DSCP, SGT-QoS	COS/DSCP,SGT-QoS, App Policy	QoE
MTU Considerations	No	No	Yes	L3TF	Yes	Yes
Wired support	Yes	Yes	Yes	Yes	Yes	Yes
Application (or Host) Security	No	No	No	Yes	Yes	Yes
WFH	No	No	No	Yes	Yes	Yes
Hardware Dependencies	No	No	Yes	Yes	Yes	No

Looking at the table above we see that the approach towards ZTNA is to create more segments. **In other words:** instead of moving away from the castle-and-moat model, they build smaller castles within castles.

NAC vs SDP

Network Access Control	Software Defined Perimeter
Doesn't extend to cloud	Cloud enabled
Based on VLANs	Micro-segmentation
No Encryption	End to end encryption
Per Network Segment	Per application control
Remote User not supported	Replaces existing VPN solution
Static IP based policy	Dynamic application based policy
Limited context	Device posture and context
Requires Hardware & Firmware	Software based solution
Always Trust Model	Zero trust model
Least privilege	Least privilege
Single routing domain	IP independent
Services/applications are discoverable	Services/applications are not discoverable
Supports mcast & p2p	Only supports Client-Application
No ubiquitous end-to-end encryption	End-to-end encryption between client and Cloud/DC
ZTNA compliant	Full SASE compliant (includes ZTNA)
	Per connection BW limitation (500Mbps)

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