Zero Trust Architecture
What it is and how to get there

Alcatel·Lucent
Enterprise

eBook
Zero Trust Architecture

Contents

What is zero trust architecture?

What does zero trust architecture (ZTA) mean? It means every user and device must be authenticated and authorised before access to data is allowed. It's a 'trust no one; authenticate everything' strategy.

An analogy will help here.

If we think about traditional security as a fortress that protects a village, we would build the fortress wall (otherwise known as the firewall) around the village (otherwise known as the enterprise). Anything coming from the outside the fortress is untrusted and scrutinised but anything inside the fortress is implicitly trusted and allowed. This trust boundary is both physical and implicit, depending on what side of the fortress you are on, because as long as you're on the 'right' side of the wall, no further checks are required. If we think of this approach in terms of the enterprise network there may be some basic segmentation in the form of VLANs, SSIDs, subnets or interfaces tied to a firewall, but this segmentation is static and has more to do with network scalability and manageability than with security.

Today, however, the firewall (fortress) approach is becoming less effective on its own, for a number of reasons. The first is mobility. Users connect to other networks outside of the enterprise and can bring threats with them. Second, guests are not necessarily trustworthy. One would argue that even employees (those inside the fortress) should not be blindly trusted. Third, more and more IoT devices are being added to the network. They pose higher security risks because they may not be sanctioned and managed by IT and they usually lack security capabilities. With the traditional 'fortress/village', 'firewall/enterprise' approach, if a user or device is compromised, there's little to nothing stopping the threat from spilling over to other users and devices. Once you're inside, you're free to move about.

In the fortress analogy, if the only thing protecting the village from intruders is the wall, the day they learn how to climb the fortress wall — and they will — there will be carnage.

The question is — what can we do about it? Let's think about this from the 'trust no one', or 'zero trust' approach.

In zero trust, no user or device is trusted. Whether they're on premises or off premises, they go through the same checks. There's no such thing as trusting internal users. Every access is authenticated and authorised.

In the fortress analogy it would mean that in addition to the fortress that protects the village from outside threats, every house, every building, has its own security to protect from risks coming from nefarious actors who live inside the fortress. In terms of the Enterprise, what's known as software-defined micro-segmentation takes it another step further. On top of the fortress, and the security around the buildings, we also have personal security guards that follow us around wherever we go. And wherever we go, we will be asked for our passport. In the Enterprise network this trust boundary is fuzzy, is distributed and is mobile. It's not tied to a specific location, switch port, or VLAN. It depends on the identity, the device, the situation, and time of the day, among other things. It's software-defined and it's adjusted on-the-fly. In this approach, the components need to be managed and should be able to react and reconfigure as needed to respond to threats or changes in the workflow.

Understanding macro- and micro-segmentation

In a zero trust architecture there are two kinds of segmentation, macro- and micro-segmentation. In terms of our analogy, the fortress wall is the macro-segmentation, and the personal security guards are the micro-segmentation.

In **macro-segmentation**, the physical network is partitioned into different logical segments. These segments can be a VLAN, a combination of VLAN and VRF, it could also be a VPN when talking about Shortest Path Bridging (SPB), MPLS, or even VXLAN or GRE tunnels. Any traffic between users or devices on different segments is controlled by a firewall. All enterprises use segmentation in some way shape or form, but not always for security reasons. Quite often, this kind of segmentation is used for scalability, administrative or organisational reasons. If two devices are mapped to different VLANs but they can communicate without going through a firewall, then they are on the same macro-segment. A typical example of this kind of segmentation is running IP telephony on separate VLANs and VRF which are logically isolated from PCs.

The question is how do you map users or devices to these segments? While it can be done statically, by switch port or SSID for example, it really is an obsolete way of doing things. It's too rigid and does not bode well with mobile users. Ideally, you would have a software-defined authentication system so that when a user or device connects and authenticates they are assigned a profile. The profile will provision the user or device on the right segment regardless of the physical location, switch port or SSID.

While macro-segmentation does have security benefits, in many cases it is done for organisational or administrative reasons. For instance, cameras and door locks fall under the control of the access security group, whereas thermostats fall under the control of the building maintenance group.

Micro-segmentation takes things one step further. Not all users are the same and not all users have a legitimate need to access all resources. The same profile that maps users to a segment also includes a set of policies which add fine-grained control over user/device privileges which are different for different roles such as HR versus Finance. This is known as **role-based access**, and directly relates to the **principle of least privilege**. And so, even though cameras and door locks are both on the same segment, they do not need to use the same resources. The camera needs to communicate with the video recorder and the door lock with its server. There is no need for a camera to communicate with a door lock just like there is no need for a door lock to communicate with another door lock. These fine-grained permissions are implemented through policies which are part of the profile and dynamically applied to the device after authentication.

Micro-segmentation needs to be software defined for several reasons. Neither users nor IoT devices are static, they move, connect and disconnect, the policies cannot be tied to a location or a port. In fact, micro-segmentation configurations need to be dynamic based on the combination of multiple factors, including but not limited to, the identity of the user or device, the time of the day, and the location.

In summary, when communication between different segments is controlled by a firewall, it is macro-segmentation. When communication within the same segment is controlled by Network Access Control (NAC) policies associated to the device or user role, it is micro-segmentation.

Why you need both

What happens if you only use one type of segmentation?

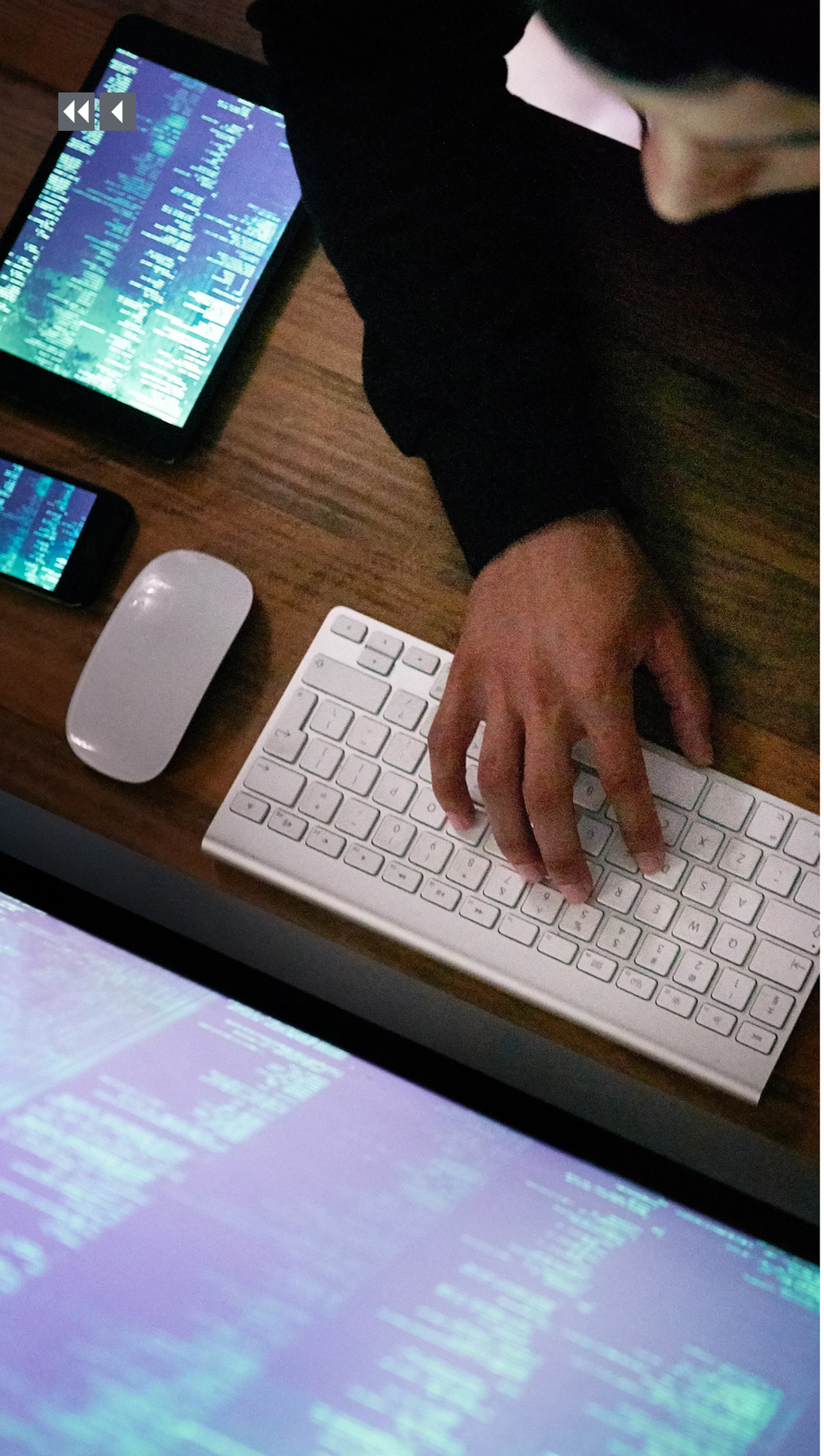
Let’s look at macro-segmentation. The problem with only using this approach is that the firewall becomes a bottleneck as all traffic needs to go through the firewall for authentication and authorisation. This can lead to performance issues. You can deploy more firewalls at the distribution layer, but this can be quite costly and may not necessarily improve performance since firewalls are not wire rate. In addition, there are now multiple policy enforcement points and multiple places to keep policies up to date, making it cumbersome to manage.

The other option, only using micro-segmentation, is also problematic. If the only policy enforcement is done through NAC policies, then these policy lists will become very long and complex, and you may exhaust the device capacity limits.

The bottom line is, it’s better to have a balance between these two forms of segmentation. Let the firewall control any traffic between different segments (vertical) and let the NAC policies control traffic within a given segment (lateral).

By combining these two, you can act on security threats that spill over from one security segment to another, as well as the ones that move laterally across the same segment. In more tangible terms, micro-segmentation is what stops an attacker who has managed to compromise a camera, from using the breach as a pivot to compromise other resources such as a door lock.

The goal is to authenticate every connection and assign permissions to each user or device. That means using segmentation to prevent threat propagation through lateral movement, and continuous monitoring and quarantining of any user or device that becomes non-compliant.



How to get there

In a greenfield situation it would be relatively easy to build a zero trust architecture using micro-segmentation from the ground up. But, in brownfield situations, retrofitting the network with micro-segmentation can result in users, devices, and applications being locked out of the network due to failed authentications or incomplete policies. It would be difficult or even unlikely that an enterprise would be able to migrate in one go — in a single refresh cycle.

In brownfield environments, there will be a period during which non-zero trust and zero trust architectures will coexist, and migration will happen one layer or one location at a time. What's important is that you make sure that the infrastructure elements you deploy, and the way in which they are deployed, are flexible and capable of operating in a zero trust or micro-segmented mode when other infrastructure elements are ready. This means that the infrastructure will need to interoperate with existing and future components.

The methodology

There are five steps toward a zero trust architecture – monitor, validate and assess, plan, simulate, and enforce.

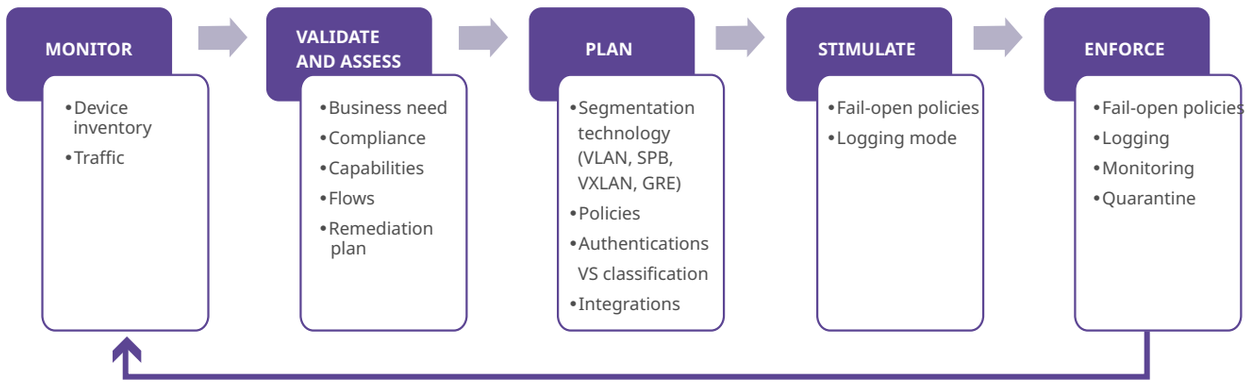


Figure 1 - The methodology

Step 1: Monitor

Before doing anything else, you need to start monitoring and building a map and inventory of what you have in your network.

Migrating to ZTA requires detailed knowledge of the assets (physical and virtual), subjects (including user privileges), and business processes touching or riding on the network. Incomplete knowledge will most often lead to failure where access is denied due to insufficient information. This is especially an issue if there are unknown “shadow IT” or “shadow IoT” within the organisation.

Start monitoring devices and traffic flows. Create an inventory report with all devices seen on the network, categorised by device type, manufacturer, model, operating system, among others. The report should also show where and which switch port or SSID the device was last seen. This information can be gathered from elements such as MAC address, DHCP signature and HTTP user agent.

The majority of third-party tools will only provide an IP address, and not the equipment type. It would be ideal to have a tool to create an IoT/device inventory to facilitate quick and easy NAC profile creation for each type of device.

The other piece of information required for creating policies is traffic flows. Depending on the equipment you have, you may obtain this information from flow monitoring tools such as sFlow, Netflow, or Deep Packet Inspection (DPI).

This process will be iterative. The first time you enable monitoring, the reports may not be as meaningful but as you progress to the other steps, they will become more specific and useful. The information you gather in this step will be key for the next steps.

Step 2: Validate and Assess

The next step is to validate these findings. Assess the business needs. Any shadow IoT that cannot be justified should be eliminated as it unnecessarily increases the attack surface. For the rest, you need to identify the subjects (users and devices), the traffic flows and the workflows because these will need to be reflected in your policies. For example, who will have access to specific assets, and what they will be allowed to do with those assets. Apply the principle of least privilege.

Micro-segmentation does not mean relaxing other security policies such as password policies, or firmware updates. Individual capabilities need to be evaluated. Can these devices support certificate-based authentication? Is there a management tool that will allow those certificates to be issued and applied? What are the required traffic flows? You may need to get this information from the manufacturer, but you should also contrast it with your own traffic analysis reports. If you find assets that do not comply with company policy, you need a remediation plan to bring them to compliance, or you will need to put additional controls in place.

Step 3: Plan

At this stage you already know the assets, the subjects (users), the traffic, and the workflows. You now need to turn this knowledge into authentication and security policies to implement the required micro-segmentation architecture. As mentioned previously, for best results you should include a combination of macro- and micro-segmentation. Remembering that in most brownfield scenarios you will be constrained by the architecture that is already in place.

For macro-segmentation, there are multiple options such as VLANs, VRFs, SPB VPNs, VXLAN or GRE tunnels and special features such as private VLAN. Each of these have their pros and cons and can be useful in different situations. For micro-segmentation, you will need to know what policies to include in the profile for each user or device type. Lastly, you will need to define how to map users and devices to their segment and policies. This comes down to authentication or classification, which can include device fingerprinting. Ideally, you should invest in technology (software-defined segmentation) that allows you to create flexible authentication policies so you can easily update network profiles.

We suggest you design the authentication flow in this order:

1. Authentication through 802.1x certificates using the RADIUS server. The authentication generates an authentication record. This is information that can be shared with a firewall.
2. If you can't authenticate the device through 802.1x certificates, then you should try MAC authentication next. MAC authentication is not nearly as secure as 802.1x but it is better than no authentication at all.Use it until you are ready to progress to 802.1x.
3. If no profile is returned, you can attempt fingerprinting which you can also use to map to a profile segment and rules. This does not generate an authentication or accounting record but does get registered in the IoT inventory database.
4. Lastly, you can have a default "catch all" in case profiles are not returned, or if everything else fails. In the early stages you will still need to map the device to a profile which will lead to the same segment and rules and will log the device in the inventory database.

This flow should be structured in a flexible manner so you can tweak it as you progress. For instance, you may want to eliminate MAC authentication at first and add it on later, after you have the list of MAC addresses obtained from the inventory report. And as you refine the process, you can, for example, change the segment and rules associated to the default profile to very restrictive rules allowing access to a bastion host only.

You may also want to share the device role with the firewall so that firewall rules can be based on device role and not just subnet/IP address. The advantages of this integration are two-fold. First, the firewall can apply fine-grained policies to those IoT devices. Second, firewall policies are now user- or role-based and so they are no longer tied to a subnet or IP address, this enables future re-design and re-segmentation of the network.

The process will be iterative and you will need to tweak, tune, and refine as you become more mature in your use of authentication and segmentation.

Step 4: Simulate

No matter how much you plan, it's unlikely you will get it right the first time. Any error in the design of the authentication scheme, any omission you make in the security policy "allow list" will result in a broken business process. You will need to apply authentication and access policies in a "fail-open" mode. What this means is that devices and users that fail to authenticate will still be allowed on the network, and unexpected traffic flows will still be allowed. But all of this will be logged and with these logs you can refine the authentication and policy schemes.

Step 5: Enforce

After some fine tuning, you will no longer observe authentication failures or denial of legitimate flows. You can then move those policies from "fail-open" to "fail-close" which means that rogue devices will be blocked, and unexpected flows will be dropped.

It goes without saying, you will need to continue monitoring for any unexpected devices and traffic flows — repeating the whole cycle as necessary.

One more thing

As part of the continuous monitoring, logging and quarantining, we recommend that you also invest in an external Intrusion Detection System (IDS). Although there's a range of Distributed Denial of Service (DDoS) attacks that can be identified directly by the switch itself, an external IDS can also detect a broader range of attacks such as viruses, or other anomalies. You may remember some years ago, several video surveillance cameras were infected with the Mirai malware, or the day that these devices launched a coordinated attack on global DNS servers that affected services such as Twitter, Spotify or Paypal. These attacks may not be detected by your switches but a dedicated IDS most certainly will.

Once the attack is detected, the IDS would inform your network management system (NMS) of the IP addresses of the affected devices. Ideally, your NMS would be capable of locating these devices in its database and would change their profiles to a "quarantine role".

The "quarantine role" is a very restrictive role, typically it would only allow communication with a bastion host so the device could be remediated, for example, by setting a strong password, or updating its firmware, among others.

Why ALE?

Alcatel-Lucent Enterprise Digital Age Networking solutions incorporate robust and flexible software-defined segmentation with dynamic DPI NAC policies that allow a phased evolution towards a zero trust architecture.

The Digital Age Network is the Alcatel-Lucent Enterprise blueprint that enables businesses and organisations to enter the digital era and grow their digital businesses. It is based on three pillars:

- An Autonomous Network that easily, automatically, and securely connects people, processes, applications, and objects. The ALE Autonomous Network is based on a streamlined portfolio complete with a true unified management platform, delivering common security policies across our LAN and WLAN. The Autonomous Network also provides deployment flexibility indoors, outdoors, and in industrial environments. Network management can be delivered on premises, in the cloud, or in a hybrid deployment, depending on the customer preference.
- Secure and efficient onboarding of IoT devices: Segmentation keeps devices in their dedicated segments and minimises the risk of having the device and network compromised. IoT segmentation can help businesses easily and automatically understand if the device is behaving properly, or not, and help to keep the network safe.
- Business Innovation through workflow automation: Integrating user, applications, and IoT metrics in real-time, with geolocation data, into collaboration platforms, simplifies the creation and rollout of new automated digital business processes and services including notifying security and network administrators of any breaches as they happen.

Do you have a tool to create an IoT/device inventory? Do you have a tool that allows you to monitor application flows? Are your current switches and wireless access points ready for software-defined segmentation? If you don't currently have these tools, please contact us and we can help you get there.

What we know for sure

Let’s wrap up with some key takeaways:

- To have a truly efficient ZTA, you must use both macro- and micro-segmentation
- There are five steps to a ZTA: monitor, validate and assess, plan, simulate, and enforce
- ZTA based on micro-segmentation rests on three pillars: Authentication, with 802.1x EAP-TLS as the gold standard; differentiated policies associated to the user or device role which go back to the principle of least privilege; and continuous monitoring and quarantining
- In hybrid and mobile environments, micro-segmentation must be software-defined, meaning, it must be dynamic and policy-based, not statically defined, otherwise it would be impractical

Migrating to a ZTA through micro-segmentation is a process, it is unlikely for an enterprise of any significant size to get there in a single refresh cycle. But in every refresh or redesign or continuous improvement cycle you can get closer to that goal if you put the right infrastructure and design in place.

Alcatel-Lucent Enterprise is committed to developing networking technology and solutions that help organisations realise their business potential through digital transformation.

Alcatel·Lucent Enterprise