How to Build a Zero Trust Architecture

29 September 2022 - ID G00766061 - 38 min read

By Thomas Lintemuth

Initiatives: Security Technology and Infrastructure for Technical Professionals

This document instructs security and risk management technical professionals how to build and deploy a zero trust architecture. It disambiguates this popular term and provides practical architectural insights.

Overview

Key Findings

- Clients rarely define their security architecture before deploying technical controls. Instead, deployments are generally initiated by external mandates.
- A zero trust architecture (ZTA) cannot be bought. It must be built. Implementing and maintaining zero trust is an iterative process.
- ZTA is rarely deployed from scratch. It will use technologies that are already deployed.
- ZTA does not require an “all or nothing” approach. It can be deployed application by application.

Recommendations

As a security and risk management technical professional focused on building and deploying a ZTA, you should:

- Define your organizational ZTA strategy and architectural principles to avoid confusion from varied interpretations of the term “zero trust.”

- Develop your identity management and governance with a view to having a single identity per user, so that access can be properly assigned and risk properly evaluated.
- Develop a service catalog in order to know what resources you need to protect by detailing applications deployed in your organization.
- Determine the characteristics that qualify a device to access your resources. Characteristics should be considered for managed devices, nonmanaged devices and nonuser devices, with context continuously assessed so that access can be adjusted based on observed changes.

Analysis

"Zero trust" is treated as the panacea for all security risks. If there is a security risk simply apply zero trust and the risk magically disappears — At least that is what we are told by many a vendor and industry pundit. Yet, Gartner finds that many clients do not understand what zero trust is. Is it a product? Is it a technology? Is it all hype? For this document we will focus the discussion on the term "zero trust architecture" (ZTA). Gartner defines ZTA as "an architecture that replaces implicit trust with continuously assessed risk and trust levels based on identity and context that adapts to risk-optimize the security posture." Many clients look for a simple way to discuss ZTA in their organizations. For that I suggest this simple phrase: "Securely connect users to applications." Various entities have laid claim to the term zero trust, yet it goes back further than most readers may realize, being mentioned as early as April 1994 by Stephen Paul Marsh in his doctoral thesis on computational security. [1]

> **Zero trust systematically replaces implicit trust with calculated adaptive trust.**

Interestingly, zero does not literally mean zero in ZTA. In order to provide access, trust must be shifted. More technically accurate terms would be "no blind trust," "zero assumptions," or even "zero implicit trust." So let's recast the term as zero implicit trust.

As a request for access comes into our ZTA, the risk to grant that access must be calculated. The risk calculation takes into consideration various signals such as device location, believability of user assertion, device hygiene, threat intelligence, time of day, day of week, and the data sensitivity of the application being requested. Access is granted when the calculated risk is less than the value of extending the access.

Believe it or not, the concepts inherent to ZTA are not all that new. Some people relate ZTA to attribute-based access control (ABAC). ABAC with continual assessment is one way to think about

ZTA. You can also describe ZTA as an architecture used to securely connect subjects to objects using context. Numerous entities, including national governments, industry trade groups and commercial entities, have set forth their requirements for what constitutes a ZTA. One example is the Cybersecurity and Infrastructure Security Agency (CISA), part of the U.S. Federal Government. According to CISA, the pillars for zero trust are: [2]

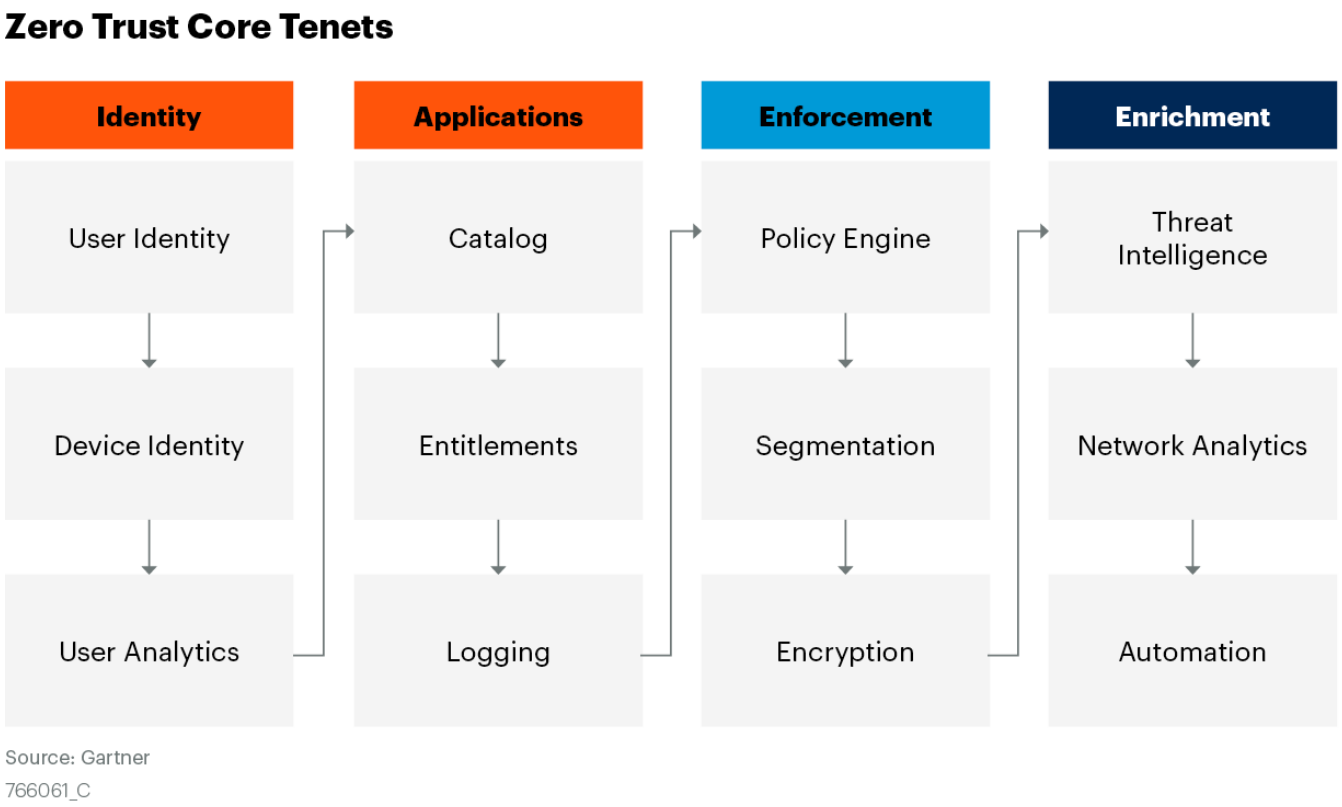
- Identity
- Device
- Network/environment
- Application workload
- Data

Gartner finds this overly complex for many organizations. Focusing on the business, application and technical architecture of an organization enables you to mitigate risk more effectively, while minimizing the complexity of deployment. With this in mind, Gartner recommends architecting ZTA using the following four pillars (see also Figure 1):

- Identity of users and devices
- Application governance and logging
- Enforcement via technical controls and encryption
- Enrichment with monitoring and automation

Download All Graphics in This Material

Figure 1: Core Tenets of Zero Trust



Identity of Users and Devices

When a user requests access to a resource, risk must be evaluated to determine if authorization can be granted. What is the level of certainty that the entity requesting access is who they are asserting themselves to be? How sensitive is the data in the application they will be accessing? Does the device the user is connecting from meet minimum requirements as defined in the policy for accessing the application?

Legacy architectures typically grant access based on a one-time analysis, though possibly duration-limited, of inputs provided. ZTA strives to extend access based on continually evaluated risk.

Users

Perhaps the most important function that a ZTA must perform is ensuring the proper level of authentication of a user or entity who is requesting access. Different requests will frequently require a different level of access and corresponding checks to access different levels of functionality.

Consider the number of times identification must be provided in the following example. Imagine yourself as a researcher at a company for a top-secret project. You must enter the parking lot of the campus, move through the campus, and finally end up in the research lab. Notice the variety of access verifications that this may involve:

- Access gated company campus:
  - Vehicle identification must match
  - User badge visually verified
- Access main building:
  - Go through metal detector
  - User badge electronically verified
- Access to research lab:
  - User badge electronically verified
  - Biometric eye scan verified
  - Biometric fingerprint verified
- Access to research computer:
  - Username
  - Biometric verification
  - Multifactor hardware one-time password

In this example we see how a given user may need to authenticate themselves multiple times based on the sensitivity of the place on campus they are trying to access.

Applications also have different levels of sensitivity. The asserted identity should provide a level of assurance that is commensurate with the sensitivity of the application to which access is being requested. The industry has used principles like least-privilege access and multifactor authentication (MFA) to accomplish this. Going forward Gartner proposes that an entity asserting an identity should be verified according to continuous adaptive trust (CAT). A low-risk resource, such as accessing a cafeteria menu, may require only basic authentication. A high-risk resource, such as corporate intellectual property, would require strong authentication. CAT enables an organization to decrease user friction for low-risk applications, while using appropriate authentication requirements for high-risk applications.

Decrease user friction for low-risk apps, increase authentication for high-risk apps.

“Identity” has grown up in an IT world where most systems maintained their individual identity stores. For instance, there may have been a mainframe, numerous UNIX- and/or Linux-based systems, Active Directory, and many SaaS-based applications each utilizing their own identity store. We need to move from many user identities to a single identity that allows visibility into, and control of, access while mitigating risk. We have been moving toward single identity, or federated identity, as many single sign-on (SSO) technologies, such as SAML, OAuth and even Shibboleth, have been more broadly adopted. ZTA forces the issue, making it a requirement.

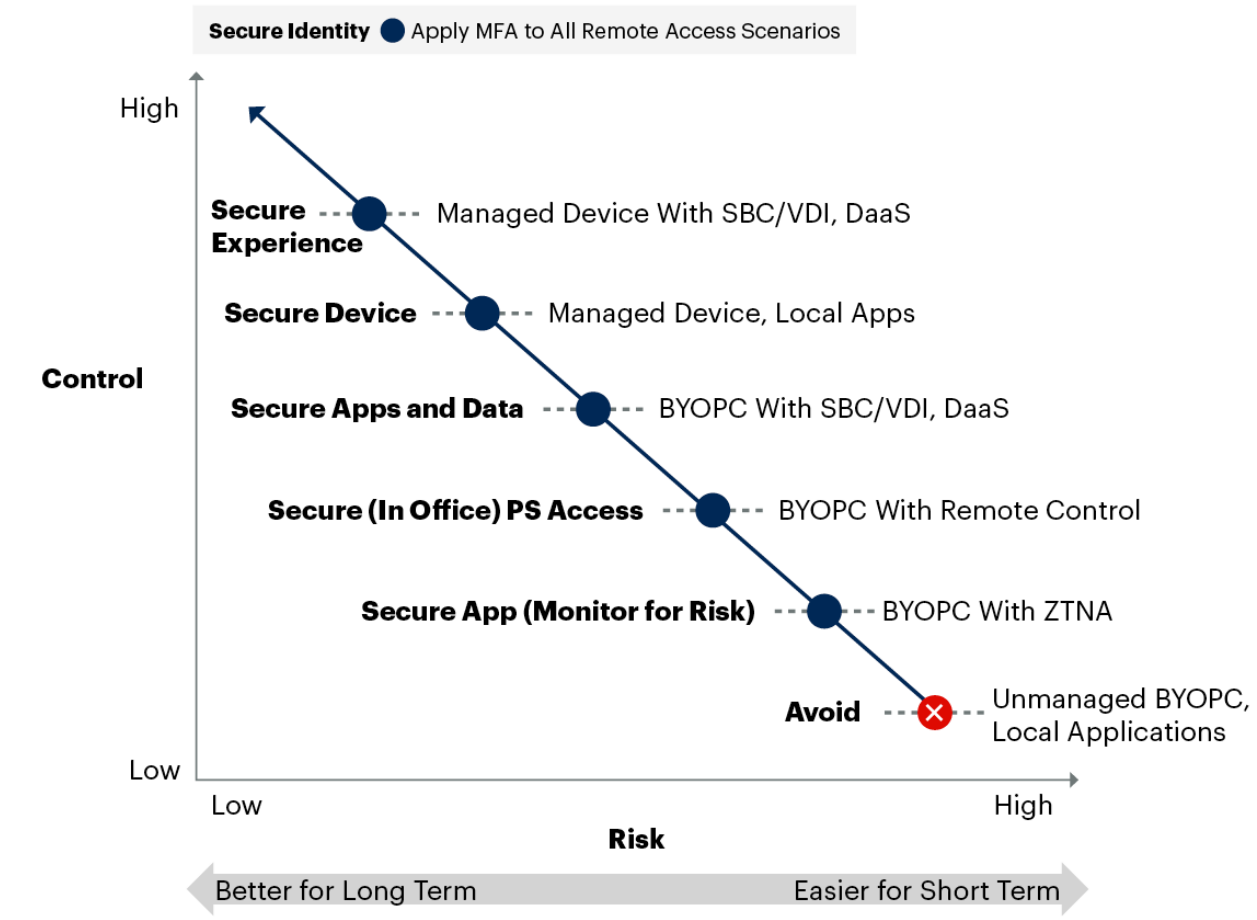
It is also critical to understand what entitlements a given user in a specific role should have. Gartner often finds that organizations have few, if any, roles documented about what a user should have access to. New users are added based on tribal knowledge or possibly even by simply duplicating an existing user’s entitlements. To move forward with ZTA an organization must document who should have access to what, based on role, and translate this information into technical controls that ZTA can enforce. Longer term, an organization should implement a review process to ensure users only have entitlements appropriate to their role. Some organizations do this manually, but there are third-party products that can automate the process.

Devices

The more information that is known about devices used for access, the better we can manage risk. What is known about the device that the user is requesting access from? Is it a managed device? Has it been observed accessing applications previously? Is the current user typical for this device? Is the device jailbroken? Does it meet basic hygiene requirements? The answers to these questions are fed into the policy engine to make a decision about access. Nonmanaged devices may be permitted by policy to access some applications, but prevented from accessing other applications. Figure 2 compares the degree of risk and level of control associated with different policies for device security.

Figure 2: Device Access: Risk Versus Control, by Policy

Device Risks Versus Controls



Source: Gartner
766061_C

Gartner

When managed user devices are being used to request access, you should fully vet them. Increasingly, Gartner finds the resident endpoint detection and response (EDR) tool being used as a source of device hygiene data for other security products. Some vendors are looking to leverage integration with products such as OPSWAT for hygiene assessment. Products that have an endpoint focus can often provide detailed endpoint checking, which provides more signals to determine the risk that an endpoint presents to an organization.

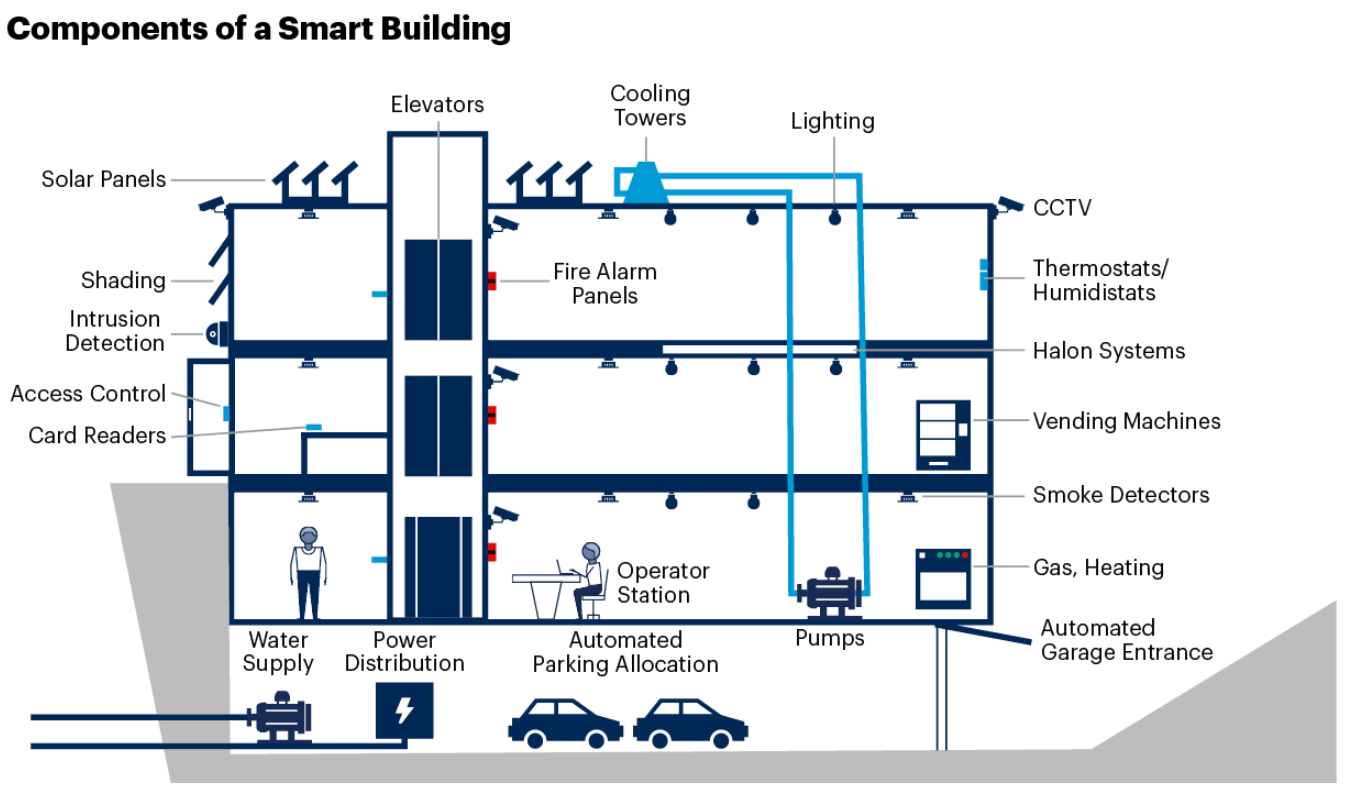
Standard hygiene checking could start by looking for something quite basic. Is the device rooted? As the identity of a device could be critical, is there a machine certificate on the device that can be verified? Has the device been customized, so that it is no longer within the organization's standard image? What software is running on the device, and is the software fully patched? Are there known vulnerabilities with the installed software? When was the device last used for access and by whom? What is the state of the software controls on the device?

It is important to consider that nonuser devices also initiate access to resources. Recal, the simplified definition of ZTA may be reworded slightly to be "Securely connect entities to resources."

Nonuser devices are entities that need access. As nonuser devices, and in particular Internet of Things (IoT) devices, are increasingly deployed, ZTA must also be able to take device identity into consideration. As a starting point, your policy should, by a default, deny devices access to the internet. Access should be allowed according to the standard access request policy. This policy should document the resources these devices are permitted to access and how access will be enforced. As these devices are commonly on managed networks, network access control has been used by many organizations for identification and enforcement. All device access must be authenticated and authorized.

Keep in mind that nonuser devices will provide less information than user devices and likely will not support modern encryption algorithms, certificates, complex passwords or any other control applied to more-user-centric devices. While scanning tools may provide information, some devices lose their network connection if subjected to scanning. In these instances, a product that monitors network traffic, perhaps via Internet Protocol Flow Information Export (IPFIX), can assist with profiling devices to ensure they are compliant. Such a product can also generate, or help maintain, an asset inventory. Figure 3 shows the various types of nonuser devices found in a smart building that ZTA may need to account for.

Figure 3: Common Nonuser Devices in a Smart Building



Source: Gartner

718675_C

Gartner

User Analytics

ZTA must use continuously assessed context data to mitigate the risk of account takeovers and insider threats. You must verify that entities are doing only what they have been authorized to do, and doing so within the proper limits. For instance, a salesperson who normally downloads data on 10 clients a day from the CRM application would be flagged if that person attempted to download the entire customer database. Risk needs to be evaluated each time an entity requests access to a resource. As a user makes a request to access new resources, the risk should be calculated based on the most up-to-date assessed context.

Keep in mind that many components of ZTA are built on software, and that software will have vulnerabilities. Threat actors will discover and exploit these vulnerabilities. Analytics will help determine when a threat actor finds a vulnerability, is actively exploiting it and is acting with trust.

We must transition away from making one-time decisions regarding access. As contextual information changes, the level of trust extended may also change. A very simple example:

1. A user has gained trust to access an application.
2. The user's device suddenly reports a virus infection, with the signal being fed to the policy decision point (PDP).
3. The device is disconnected and required to run an ad hoc scan to clean the infection.
4. Trust is again granted, so that the device can access the application again.

Application Governance and Logging

ZTA must securely connect users to applications. By definition, we must know which applications users are trying to access.

Applications

How many applications exist in your organization? Do you have an application catalog? If you do not, we recommend building one that contains the items in Table 1.

Table 1: Application Catalog

Application Name	App lication Owner	Data Sensitivity	Criticality	Network Protocol	Busines Users
	[Name]	[Confidential, for internal use only (FIUO), public]	[High, Medium, Low]	[HTTPS, SQL, SMB, etc.]	[Busines groups]

Source: Gartner (September 2022)

Not all applications expose an organization to the same level of risk. While your catalog may initially contain only the names of your applications, you should work to ensure that, over time, it includes the other items shown. In particular, recording data sensitivity in this table will greatly assist in determining risk.

As you develop your application catalog, keep in mind that, except for the top applications run in your organization, applications are rarely well-documented. Be prepared to spend considerable effort interacting with the IT organization and the various business groups to find out what is really being used. Obviously, if you have a physical data center, you will find applications there. You will also find applications deployed in infrastructure as a service (IaaS) cloud environments. One of the more difficult environments in which to find applications is software as a service (SaaS). No doubt you will have sanctioned some SaaS applications, and these are therefore straightforward to determine. But nonsanctioned SaaS applications may take some work to identify. Security service edge (SSE) products can help identify both types of SaaS applications. Finally, you may have users developing their own applications, which are sometimes referred to as user-defined applications (UDAs). These may only be found by interacting directly with users. In the end, do not be surprised if you find hundreds, if not thousands, of applications being used.

The ability to automate this discovery process would be welcome, but Gartner has not found a product that consistently does this well. The configuration management database (CMDB) should have information you can use to find some of the applications, but this will still likely be a manual hunt. A newer technology that may enable some automation is cyber asset attack surface management (CAASM). CAASM products connect to current systems, such as your dynamic host

configuration protocol (DHCP) server or server inventory system, via API, polling them for information about your assets. But even if you have a CAASM product deployed, at best it will provide only a list of applications. You will still need to determine the other critical pieces of information indicated in Table 1. Some network detection and response products can also produce a list of applications seen on the network. In short, even if you deploy some products to assist you, you will still have a fair amount of manual work to do to identify all your applications and crucial information about them.

Entitlements

Once you know your applications the work really begins. Which users have a business requirement to access which application and at what level? This stage of the application journey is undertaken alongside the user identity governance stage mentioned previously, but focuses on the application side. Many applications have internal controls that provide in-application entitlements. Can we authenticate users with some type of SSO technology while authorizing specific access at the application level? It is most useful if you can tie this internal control system to your identity provider (IdP), so that you can utilize a single identity to authenticate users and then use internal application controls for authorization.

Certain industries and applications have a strong need for governance. How do you prove that only certain users have access to certain functions in an application? How do you prove that only approved users can authenticate to an application? How do you know it is the right users that can authenticate to an application? Basic IdP products are looking to add capabilities, whether natively or in an adjacent product, to assist with governance. While basic entitlement reviews may have begun by identifying user entitlements, now is the time to automate the entitlement governance process. More information can be found in Gartner's Market Guide for Identity Governance and Administration.

Beyond just determining who should have access to what, the more significant challenge is to decide how these entitlements will stay up to date. Consider building a process to keep entitlements accurate as you go about determining what current entitlements should be. Once a reasonable manual process has been proven, look to automate this, as trying to maintain entitlements manually will quickly overburden your staff.

Logging

Do you enable logging at the application level? Likely you do. A more interesting question is what is appropriate to log at the application level? Should all logging be turned on for every application? Can too much be logged? The answers are — it depends, probably not, and yes. There are certain items that should be universally logged at the application level. Your security team should be able to identify these. A good starting point is to record the date/time when a subject connects, the identity of the subject, and when the subject disconnects. Just remember that recording date/time is only as good as the date/time on the server. Ensure you sync your servers to a common time

server. It may be appropriate to deploy an internal time server for this specific function. Few investigations are as frustrating as trying to correlate logs from different servers that have out-of-sync times.

Here are some best practices for log management:

- Ensure logging is active, and send alerts if logging stops.
- Ensure logs can be accessed only by approved administrators.
- Confirm logs cannot be modified.
- Ensure logs are routinely backed up.
- Send local logs to a security information and event management (SIEM) platform.

Enforcement via Technical Controls and Encryption

By now you should have user management under control, you know your applications, and you know who has access to what in your applications. Next we must be able to enforce the proper access.

The U.S. National Institute of Standards and Technology (NIST) uses the terms policy enforcement point (PEP) and policy decision point (PDP) to describe devices that provide enforcement. Vendors typically do not use these terms, but often mimic the functionality in their products. Common terms are controller, broker, identity-aware proxy, microsegmentation and even firewall. A PDP is a device that ingests contextual signals, evaluates the risk, and decides whether to extend trust. A PEP receives the decision made by the PDP and enforces access.

Firewall

Many zero trust ideas can be traced back to firewall enforcement. A key aspect to consider is whether a policy can be deployed that allows user identity to be used as a “source” in the firewall policy. Another option, allowed by some firewalls, is the use of tags at the network level that can be enforced at the firewall. A simple zero trust deployment could ingest user and device information with network access control (NAC), appropriately tag the traffic as it egresses the campus switch, and perform enforcement using the tag at a firewall that “front ends” the applications. An analysis would need to be completed to ensure there was no back-end way to access the applications.

A weakness of this approach is the lack of continuous assessment of users or devices. On the other hand, most organizations have an identity-aware firewall. Many of these organizations also have NAC deployed. So, despite the lack of continuous assessment, many organizations could take steps toward ZTA by pursuing this basic approach using products already deployed in production environments.

Microsegmentation

Microsegmentation is the ability to put a security service between any two workloads in your infrastructure, whether those workloads are in the same broadcast domain or half the world away from each other. Microsegmentation can isolate workloads within a segment. It can also enforce access from end users to a workload.

An interesting idea for deployment would be to have a firewall at the data center perimeter and microsegmentation at the workload level. This would verify user identity first at the firewall and a second time at the workload level, before access would be allowed. Microsegmentation on workloads could also limit lateral movement once they are in the data center.

Proxy

Forcing all user access to applications via a proxy is another type of PDP/PEP. Interestingly, one major IT company has even published whitepapers on how it fully deployed ZTA using this method. [3] NAC is deployed to identify devices connecting on the managed network and then check them for hygiene. All applications are accessed via web protocols such as HTTPS. Users connect to the proxy, which verifies contextual signals appropriate for the application being requested. Adding remote browser isolation here enhances zero trust as it isolates the end user's device from the browser accessing the internet. Implicit trust has been removed between the endpoint and the webpage being accessed.

ZTNA

Zero trust network access (ZTNA) is a promising technology that utilizes PDP- and PEP-type devices, for policy decision making, and policy enforcement, respectively, to control user access to applications.

Gartner tracks more than 40 vendors in the ZTNA market. Some of their products are based on the proxy concept and primarily support web-based applications. Other products that use an agent can support web applications, as well as most TCP- or UDP-based applications. Although these products have traditionally been deployed for remote access, some vendors are expanding their technology to provide access to applications whether on- or off-network.

Organizations that have deployed ZTNA products could consider using them for more than just remote users. Longer term, we expect vendors to offer what some have called "universal ZTNA," which will provide consistent access whether the user is on-premises or remote.

Perimeters

Perimeters do not go away with ZTA. If anything you see an increase in perimeters. NIST indicates that, optimally, you should place the PEP at the workload level. In that scenario you could have one perimeter per workload. That would likely be excessive for most applications, although the idea certainly underscores the point that perimeters remain in a ZTA world.

However collapsed we decide to deploy perimeters, isolation is key to an effective ZTA. Consider this, if the designers of the Titanic had used the principle of segmentation more effectively, one of the most infamous civilian naval disasters of all time could have been avoided. Submarines are designed with many airtight compartments, as it is expected that at some point there will be a leak. You too will have a leak, or breach, at some point in your network. It may come from a disgruntled insider or a malicious outsider. Stop the lateral movement and you have possibly neutralized the attacker. You will definitely slow the attack and force the attacker to leave more traces in your network, which will increase your ability to detect the attacker.

Where does one start the long journey toward segmentation? [4] Macrosegmentation should be your first step. It is somewhat common in modern networks, if not inconsistently deployed. If you have not benefited from segmentation, consider the following segments to start with:

- Users versus servers
- Production environment versus nonproduction environment
- Development versus quality assurance
- Information technology (IT) versus operational technology (OT)
- Manufacturing versus office
- Laboratory versus classroom
- Data center A versus data center B

Macrosegmentation is commonly enforced with network firewalls and their cloud virtual networking equivalents (VNets and VPCs).

> **Segmentation is key to an effective ZTA.**

Next, determine where to place smaller perimeters. An attractive place to start is with the deployment of new applications. Some organizations will deploy microsegmentation products in application development environments to observe the network traffic flows. Once an application is to be promoted from a nonproduction to production, a segmentation policy can easily be devised from observed traffic patterns. Commercial off-the-shelf applications typically have connectivity requirements that can similarly be used to devise a segmentation policy.

When you need to develop microsegmentation for your current applications, consider developing policy for your most important applications. For instance, applications may contain highly confidential data or perhaps they fall under governmental regulation. These applications would be common starting points for microsegmentation. It is not necessary to microsegment every server, nor every application that you manage. Use the application catalog you developed above and start moving through the applications one by one. There is an amount of effort required to manage a microsegmentation policy. The business value of protecting the application should be greater than the administrative cost of management.

Encryption

As it is not possible to prevent all insidious activity on a managed network, it is imperative that attackers' ability to intercept data is limited. The first step is to minimize data exposure by encrypting all data in transit. Most applications can utilize some type of Transport Layer Security (TLS) encryption. Simply enable this for applications that support it. A common question posed to Gartner is "which version of TLS should I use?" Numerous entities dictate what is an acceptable minimum requirement. Recently, a large government entity indicated that a form of TLS that uses ephemeral keys should be deployed. This suggests that TLS 1.3 would be preferred. But you may be subject to different regulations. When in doubt refer to the standards that are most applicable to your industry, such as, perhaps, the Payment Card Industry Data Security Standard (PCI DSS) and NIST Special Publication 800-52. At minimum, TLS 1.2 should be used. Applications that do not support this level should be documented. A subsequent project should update or replace these applications. (Also bear in mind that, when quantum computing becomes mainstream, all current recommendations will be updated.)

When applications do not support native encryption, it might be appropriate to enable network-level encryption, such as IPsec. Then again, it may not. As a guide, we recommend using IPsec to encrypt transmissions for applications that process high-value data when TLS is not available. Security controls should always be deployed in accordance with the value of the data that is being protected. Do not deploy IPsec unless the value of the data requires it.

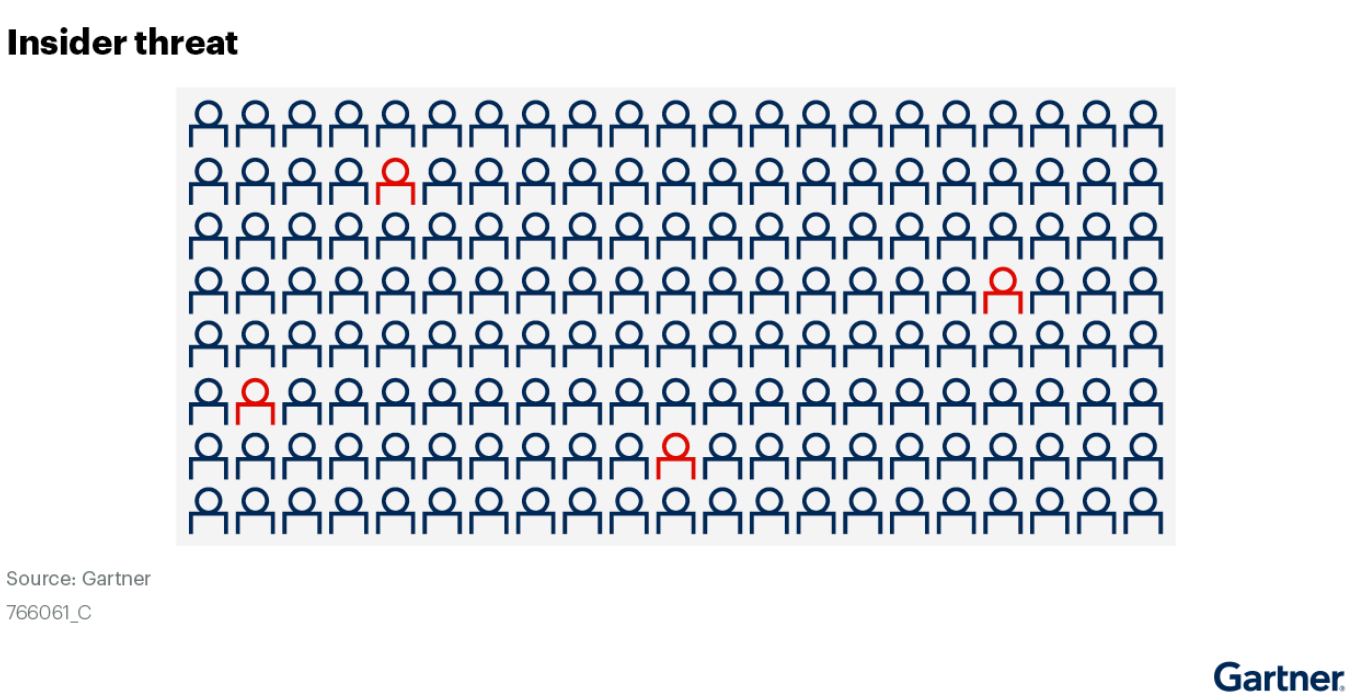
Enrichment With Monitoring and Automation

Do you trust your users? The answer is, of course, "it depends." Take your pick of headlines showing disasters that have arisen from trusting insiders: we read, for instance, that 74% of data breaches start with privileged credential abuse; [5] that 22% of security incidents involve insider threat; [6] and that 77% of IT security professionals polled indicated that it would be easy for them to steal sensitive information if they were to leave their organization. [7] How do you know when a trusted insider has turned into a malicious insider?

Consider this illustration. We use terms like "information superhighway" to describe data communication networks. Literal highways have many law-abiding users, but at times some travelers use the highways to act unlawfully. Our managed networks are in some ways similar. We

have many users that utilize the managed network to conduct legitimate activities. At other times, the network is used as a threat vector. Just as a literal highway has an enforcement agency looking to combat unlawful activity, we must have a way to detect illegitimate activities on our managed networks. There is a difference between hostile actors on the network and a hostile network. We must identify and remove hostile actors to keep our network from becoming weaponized (see Figure 4). We must also minimize threat actors' ability to intercept data.

Figure 4: Insider Threat



Detection and Response

Trust must be extended for associates to perform their roles. Associates can be counted on to perform their roles in a trustworthy manner — until they cannot. Network detection and response (NDR) with user and entity behavior analysis (UEBA) capability helps information security teams find the crossover point.

NDR establishes a baseline for device activity, while UEBA is more geared toward user activity. Using the baselines normal versus what is anomalous can be determined by user or device. Here are just some of the things that can be alerted on:

- Kerberos ticket-granting service (TGS) requests
- Abnormal connections from users
- Abnormal session durations
- Connections from unknown devices

- Abnormal amounts of data being uploaded
- Suspicious Lightweight Directory Access Protocol (LDAP) queries
- Suspicious administrator activities

To enrich zero trust, you need access to the intelligence generated by analytics tools. For brevity's sake we simply recommend sending this intelligence to the SIEM platform or whichever platform is able to provide alert aggregation and correlation. Such correlation will be most effective if you can analyze network analytics, user analytics and endpoint analytics in the same platform.

Monitoring and logging are critical to ensuring that ZTA is successful over time. Users are granted access as a point-in-time assessment of risk. User risk will change. For instance, phishing may plant malicious software on a user's endpoint. Or internal users may be persuaded to assist external threat actors or may decide that they can make more money selling company data than by being good data caretakers. Thus, all network traffic must be continuously monitored for unusual activity.

The phrase "assume breach" comes up frequently in discussions of ZTA, but should be changed to something better aligned with security: "assume you are breachable." Once breached, find and remove the intruder. Consider your network similar to your home. Is it possible for an intruder to get into your home? Yes, but you do not assume an intruder will be lying on the sofa every time you return home. Rather, if an intruder is discovered in your home, you have them removed. So, rather than simply assuming you have been breached, you need to be cognizant that threat actors are active and can eventually breach a layer of defense. You should actively look for breaches and remove the threat actor, when discovered.

On the other hand, the most dangerous threat actors are on our networks every day: our users. They are commonly referred to as insider threats. Interestingly, insider threats are not necessarily malicious. Have you ever received a ticket to restore a database after an oblivious insider mistakenly executed the "del *" command? Unfortunately, however, malicious insiders are becoming more common. One recent threat actor bought user credentials, and another paid users to install remote-control software on their corporate machines. Malicious insiders are one of our most dangerous threat actors. [8] Whether attackers are malicious insiders or insidious imposters, utilize NDR to monitor your network, detect unusual activity, and issue alerts that actions should be taken.

Threat Intelligence

StackWatch reports that 20,195 common vulnerabilities and exposures (CVEs) were tracked by MITRE in 2021, with an average severity of 7.1 out of 10. [9] Your users are accessing your organization's applications from devices that have, or have had, exploitable software. How do you keep up to date on what is happening in the real world? What are the latest zero-day exploits? What

exploits are being heavily utilized? A threat intelligence product enables an organization to answer these questions.

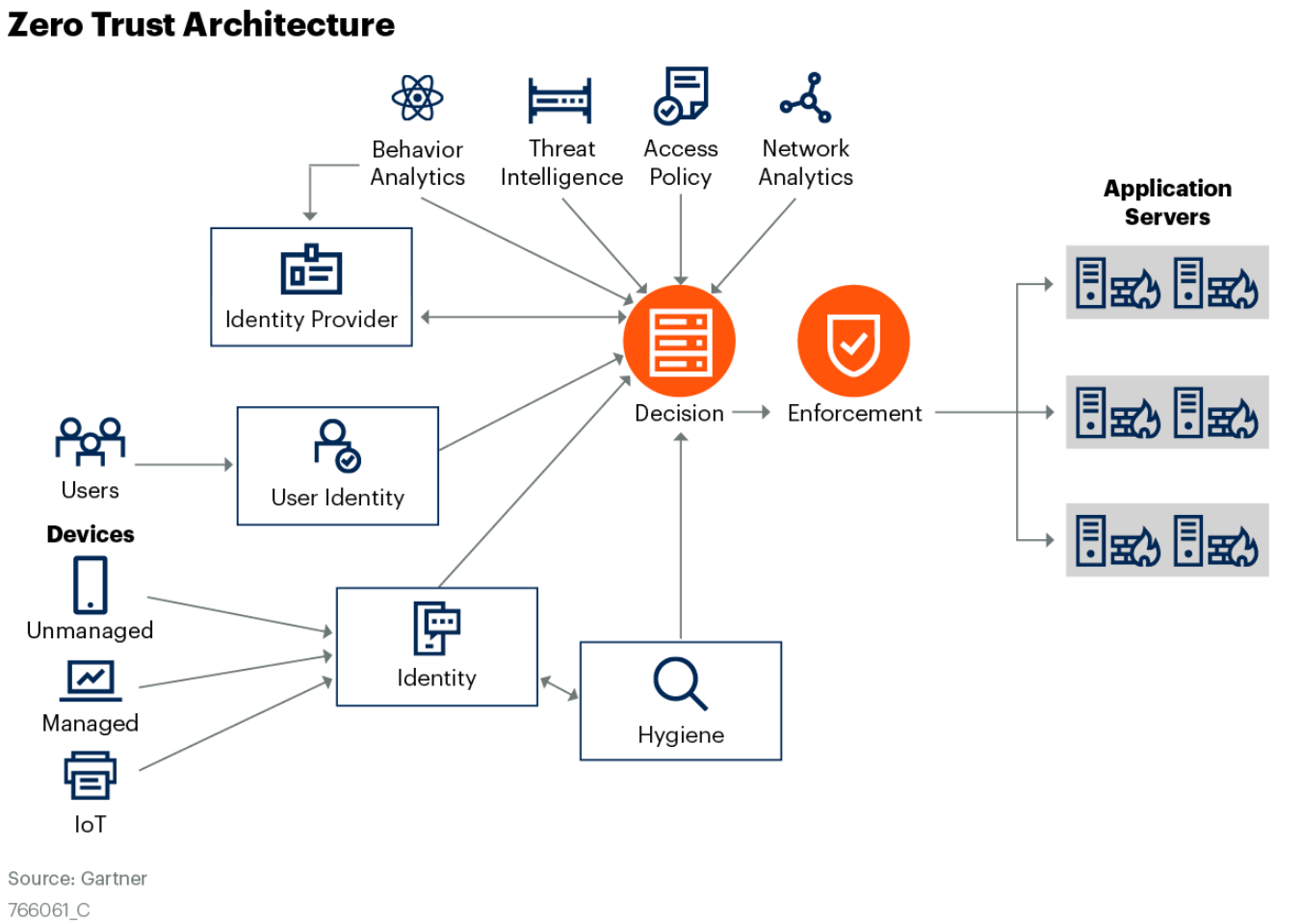
Take this real-world example. A user requests access to an application from a managed device. Upon receiving the request the system performed a hygiene check, which determined that all software was “in tolerance” for patch levels, including an application that we will call “App Z.” While the user is actively working, connected to an internal application managing highly sensitive data, a zero-day vulnerability is announced for App Z. This vulnerability is rated Severity 10. How long would it take your organization to receive information on this new vulnerability? Would you be able to perform a risk analysis to determine its impact?

ZTA should be designed to ingest threat intelligence, determine which users and/or devices are impacted, and decide whether access should be allowed. While not inherent in all products, these types of capabilities are being added to agent-based ZTNA products. Even if you cannot automate decisions on current access, you should view threat intelligence as an integral part of the process of determining the risk posed by a user from a device in the long term.

Speaking of threat intelligence. Do you know which of your users have had their passwords exposed? Do a search on haveibeenpwned.com and you will likely find one of your accounts exposing your username and password. Better still, deploy a digital risk protection service that can automate this check for your organization as a whole. Has there been a recent exposure of passwords of users that work for your organization? It is likely that some of those users reuse work account passwords for personal accounts. ZTA should be designed to ingest threat intelligence regarding user breach data as a signal and to take action when negative activities could increase the risk of a user’s access. Possible actions include forcing a password reset or requiring multifactor authentication to increase the level of confidence in a user assertion.

Putting everything together, Figure 5 shows how your architecture may look.

Figure 5: Zero Trust Architecture



Strengths

ZTA will enable you to:

- Replace implicit trust with continuously assessed explicit trust, based on identity and context, supported by security infrastructure that adapts to risk-optimize your organization's security posture.
- Securely connect entities to resources. This lets us to securely connect users to applications, while limiting nonusers to authorized resources only. User and network analytics are used to detect abuse in real time.
- Enforce a consistent and dynamic approach to protect resources, using granular access decisions based on context.
- Visualize which users are accessing which applications, which will help you adhere to the principle of least privilege via "just in time" and "just enough" access.

- Drive consistent security posture and consistent access policies, regardless of user, device or location, throughout your organization, while minimizing user friction by matching authentication with risk.
- Limit lateral movement of threat actors or malware within your organization by pushing access decisions as close to the requested resource as possible.
- Enhance system confidentiality by enforcing encryption for all data in motion.

Weaknesses

Weaknesses of ZTA include the following:

- There will be resistance from some associates. As ZTA limits users, even administrators, to “just enough” access, some associates will find this overly restrictive. Administrators will complain that they do not have the access rights they need to perform their usual tasks. Business users will also be reluctant to give up rights to applications that they may have played a key part in developing or implementing. The fight to limit access will be real and should not be underestimated. Staff will come up with all types of “what if?” scenarios to justify overprovisioned access.
- ZTA will take a long time to deploy. MIT Lincoln Laboratory has indicated that it took three to five years to implement zero trust at the companies it has worked with. [10]
- There is no “one size fits all” approach to developing ZTA. Risk analysis and ROI exercises should be executed to determine which resources are valuable enough to “front end” with ZTA.
- Legitimate requests for access will occasionally be incorrectly blocked. False positives can arise whenever contextual information must be analyzed. This can have serious ramifications if certain individuals’ access is incorrectly blocked. Contextual decisions may also incorrectly drop user connections that have been properly established, causing users to lose access randomly or perhaps at the most inopportune times.
- ZTA does not:
  - Prevent software supply chain attacks.
  - Protect public-facing applications.
  - Offer 100% assurance the right user is behind the account.
  - Generally protect access to SaaS applications.
  - Fix or compensate for poor access policy.

- ZTA complexity ramps up quickly, to the point of being operationally overwhelming, if automation and good processes are not ingrained during deployment.
- There must be a well-developed exception process that can be executed quickly for critical access. Someone has to know what is a valid exception and what is not, and be in a position to change access rule changes quickly.
- The PDP is a single point of control that could provide inappropriate access if compromised.
- Requirements for privileged access management must still be addressed.
- Technical debt may prevent zero trust from being implemented everywhere, reducing the overall impact on risk mitigation.
- Enforcement depends on access to resources being available only via the PDP. Few companies actually test to see if all other access to resources is blocked.

Guidance

Develop a ZTA around the four pillars mentioned earlier:

- Identity of users and devices
- Application governance and logging
- Enforcement via technical controls and encryption
- Enrichment with monitoring and automation

Start by improving user identity systems and develop an application catalog. Once key applications that should be accessed via zero trust have been identified, determine how you will deploy a PEP to limit access to only authenticated and authorized users.

Try not to become overwhelmed by the magnitude of the project. Start with one application and iterate over time to move all required applications to zero trust.

How to Get Started

There are nontechnical tasks to perform before initiating a zero trust journey. To succeed, you must do the following before diving into the technical aspects:

- Define the strategy and ZTA principles that your organization wants.
- Identify the risks you are trying to address with zero trust, such as minimizing lateral movement or controlling the spread of ransomware.

- Identify the scope of zero trust and the target for initial implementation, including any gaps and redundancies in technologies.
- Secure the budget required to implement or refresh technologies for the ZTA.
- Know your business processes. As you proceed on your journey, be ready to add, change and possibly retire processes to maximize the effectiveness of ZTA within your organization.

Zero trust enables you to securely connect users to applications. It stands to reason that user identity and applications are where we need to start. Next move on to devices and enforcement.

User Identity

In the short term, you should:

- Create a master directory with an identity governance and administration (IGA) function to ensure appropriate access.
- Have a central IdP integrated with SSO/SAML/OAuth.
- Ensure all access requires appropriate authentication.
- Develop an automated process to review user entitlements.

In the long term:

- Ensure access decisions include user activity, especially anomalous activity.
- Utilize continuous adaptive trust to enhance multifactor authentication. [11]
- Control privileged access through a proxy, and monitor appropriate activity.
- Provide access based on role.
- Get to grips with privileged access by establishing a privileged access management practice and applying best practices.

Applications

In the short term, you should:

- Start building an application catalog detailing every application that any user requires to fulfill their role.
- Determine the business importance of the top 50 applications.

- Enable logging to track critical application functions.
- Enable encryption for critical applications at the application layer.

In the long term, you should:

- Ensure entitlements are documented and can be automatically provisioned.
- Run applications in an enclave appropriate to data classification and criticality.
- Allow access via just-in-time and contextually based risk analysis.
- Ensure access decisions take into consideration known vulnerabilities of the application platform and workloads.
- Test the isolation of applications behind zero trust enforcement points.
- Check that just-enough access is enforced by your organization’s policy.
- Test that the right level of authentication is enforced for the resource being accessed.

Device Identity

In the short term, you should:

- Authenticate managed devices via a certificate from a trusted certificate authority.
- Ensure device hygiene is considered in access decisions.
- Ensure end-user devices are provisioned with a “gold” image and verified via posture assessments.
- Ensure managed devices are acquired only via authorized channels.

In the long term, you should:

- Enforce the deprovisioning process with a clearly defined data destruction policy.
- Ensure an automated process exists to validate the security posture of user and nonuser devices.
- Make real-time analytics of endpoint hygiene available to help make access decisions.
- Ensure the vulnerabilities of endpoints are known and considered in access decisions.
- Deploy automated systems to discover new assets and remove decommissioned assets.

Enforcement

In the short term, you should:

- Utilize currently available security controls that can enforce access based on user identity.
- Deploy products that can analyze network traffic to assist with mapping users to applications.

In the long term, you should:

- Watch the market for ZTNA products that seamlessly support access from where your users are located, whether on-premises or remote.
- Plan to replace security controls based on enforcement at the underlay with overlay models that enforce policy based on granular, context-based signals integrated with unified endpoint management (UEM)/EDR and IdP systems.

Zero trust is an iterative process, especially when trying to retrofit access to existing applications. Strive to make progress, rather than attain some arbitrary goal of zero trust maturity. A great place to begin zero trust is with the deployment of new applications, as traffic flows and user entitlements can be defined during deployment.

Frequently Asked Questions About Zero Trust

What is zero trust?

Zero trust is a paradigm for securely connecting to resources. It is commonly used to associate a product with enhanced security capabilities. Without context, the term zero trust is ambiguous. Zero trust architecture (ZTA) is a design that securely connects entities to resources.

Do I need to get rid of firewalls?

No. Firewalls support macrosegmentation, which continues to be an important element for security. Firewalls must remain.

Does ZTA increase user friction?

Zero trust does not have to increase user friction; in fact, eliminating authentication challenges for low-risk access could reduce user friction. CAT also shows promise in reducing user friction overall.

Should I assume a breach?

Assume there will be constant attacks and that some will be successful. Deploy products to find and remove threat actors. It is reasonable to assume that you will be breached. No network is

100% safe. Deploy controls to identify and then remove attackers when they get in. Segment your networks to limit damage until attackers can be removed.

Does ZTA reduce the perimeter?

In some respects ZTA pushes the perimeter further out. Authentication, and therefore access, to applications has typically been controlled at the application server, with limited network controls to the server. ZTA says we need to secure the network level. Now users must identify themselves before they are granted connectivity to the application server. We have, in effect, extended the perimeter to the user.

Does ZTA mean I will have a "coffee shop" network?

Zero trust is about providing access to applications. A "coffee shop" network allows users access to a network. One is not intrinsically related to the other. Using the internet as a connectivity layer is a valid design, but it does not necessarily simplify connectivity. Security controls must be maintained whether a user connects from a coffee shop, an enterprise-managed network or a home network. While the idea of a coffee shop network may seem appealing, it primarily deals with user connectivity. What nonuser connected devices do you have to secure — printers, scanners, smart speakers, video conference systems, home/building automation systems? It is hard to imagine a building such as the one pictured in Figure 3 being managed like a coffee shop network.

Why don't you feature a data pillar in your architecture?

Zero trust is not a replacement for a data governance program. This research takes the view that data characteristics need to be an input into determining user entitlements.

Why is data at rest not included in your encryption section?

Encrypting data at rest has been a recommended practice for many years. It should already be part of your information security policy.

Evidence

[1] Formalising trust as a computational concept, Stephen Paul Marsh

[2] Zero Trust Maturity Model, Cybersecurity and Infrastructure Security Agency (CISA)

[3] BeyondCorp: Design to Deployment at Google, Google Research

[4] The 6 Principles of Successful Network Segmentation Strategies

[5] 74% of Data Breaches Start With Privileged Credential Abuse, Forbes

[6] 2021 Data Breach Investigations Report, Verizon

[7] Privileged Account Practices Are Poor, and IT Security Teams Know It, Help Net Security

[8] A Closer Look at the LAPSUS$ Data Extortion Group, KrebsOnSecurity

[9] 2021 Security Vulnerability Report, StackWatch

[10] Zero-Trust Architecture May Hold Answer to Cybersecurity Insider Threats, MIT Lincoln Laboratory

[11] Shift Focus From MFA to Continuous Adaptive Trust

- Memorandum on Improving the Cybersecurity of National Security, Department of Defense, and Intelligence Community Systems, The White House
- Zero Trust 1.0, The National Cyber Security Centre
- The Singapore Cybersecurity Strategy 2021, Cyber Security Agency of Singapore

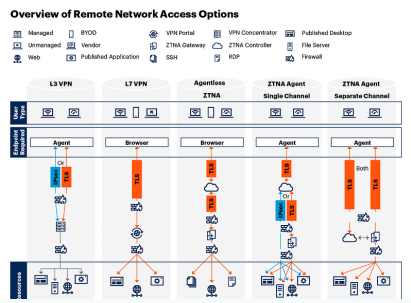
Industry-related:

- Zero Trust Security Architecture, The Open Group
- Transitioning to Modern Access Architecture With Zero Trust, Microsoft
- BeyondCorp: A New Approach to Enterprise Security, Google Research
- An Overview of Zero Trust Architecture, According to NIST, Cisco

Recommended by Author

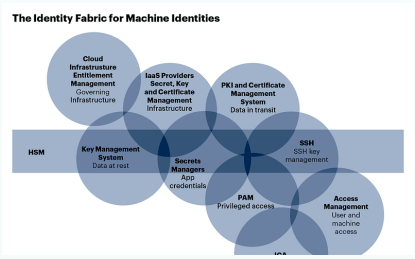
Remote Access Options for Enterprise Endpoints

RESEARCH ▪ 11 April 2022



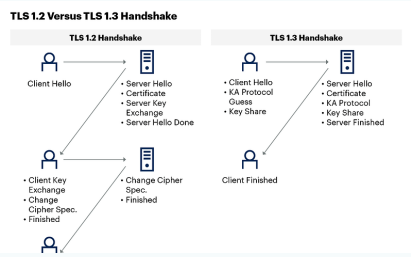
Managing Machine Identities, Secrets, Keys and Certificates

RESEARCH ▪ 16 March 2022



Overcome the Challenges of Implementing and Supporting TLS 1.3

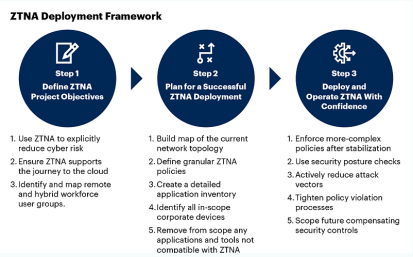
RESEARCH ▪ 22 December 2021



Your Peers Also Viewed

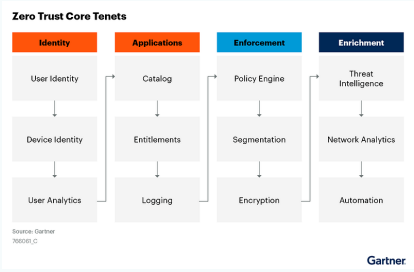
Overcoming the Challenges of Implementing Zero Trust Network Access

RESEARCH ▪ 24 November 2022



Implementing Segmentation for Zero Trust Networking

RESEARCH ▪ 25 January 2024

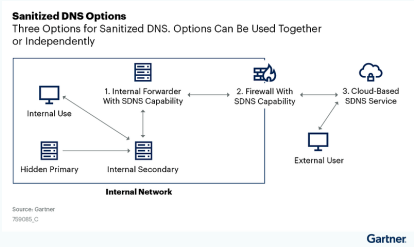


2024 Planning Guide for Security

RESEARCH ▪ 4 October 2023

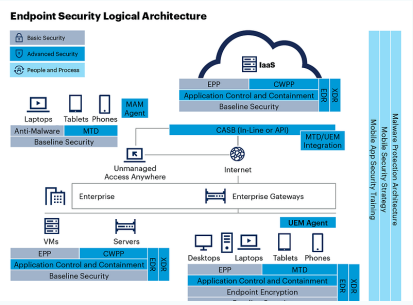
Guide to Network Security Concepts

RESEARCH ▪ 13 July 2023



Guide to Endpoint Security Concepts

RESEARCH ▪ 1 November 2022



Supporting Initiatives

Security Technology and Infrastructure for Technical Professionals

Follow