

Gartner.

Licensed for Distribution

This research note is restricted to the personal use of Nabil Ben Tekaya
(Nabil.BenTekaya@mcn.gouv.qc.ca).

Implementing Segmentation for Zero Trust Networking

25 January 2024 - ID G00795830 - 38 min read

By [Nahim Fazal](#)

Initiatives: [Security Technology and Infrastructure for Technical Professionals](#)

This research will provide security and risk management technical professionals with a practical approach to creating a strategy for segmenting their networks both on-premises and in the cloud.

Overview

Key Findings

- Current interest in network segmentation is being driven by zero trust.
- Previous attempts to implement network segmentation have created significant challenges (around complexity, access control and troubleshooting) for security and risk management (SRM) technical professionals.
- Due to the challenges network segmentation presents, it has been more widely implemented by large organizations.
- Once deployed network segmentation can require a significant investment of resources to operationalize the technology.
- Network segmentation can mitigate lateral movement, but it cannot stop it altogether.

Recommendations

When implementing segmentation as an element of a zero trust approach, security and risk management technical professionals should:

- Avoid failure caused by segmenting the entire network in one big bang, by instead implementing a phased approach.
- Design your segmentation strategy initially to protect your critical assets. Identify your business-critical assets through engagement with application owners.
- Manage your segmentation strategy effectively. Oversegmenting will result in a network that can not be managed effectively. Begin with macrosegmentation and then move to microsegmentation for high-risk assets.
- Design consistent segmentation policies across on-premises and public cloud IaaS by creating segmentation policies that span both on-premises and cloud infrastructure. This will reduce operational complexities.

Analysis

Segmenting the network has come into sharp focus over the last two years in part because organizations have moved toward actively deploying their zero trust strategy. Segmentation falls under the enforcement pillar of the core tenets (see Figure 1). The spike over the same period (and longer) of ransomware attacks has also driven adoptions. Segmentation provides an effective strategy for containing the “blast radius” and limiting the ability of the ransomware to infect the entirety of a network.

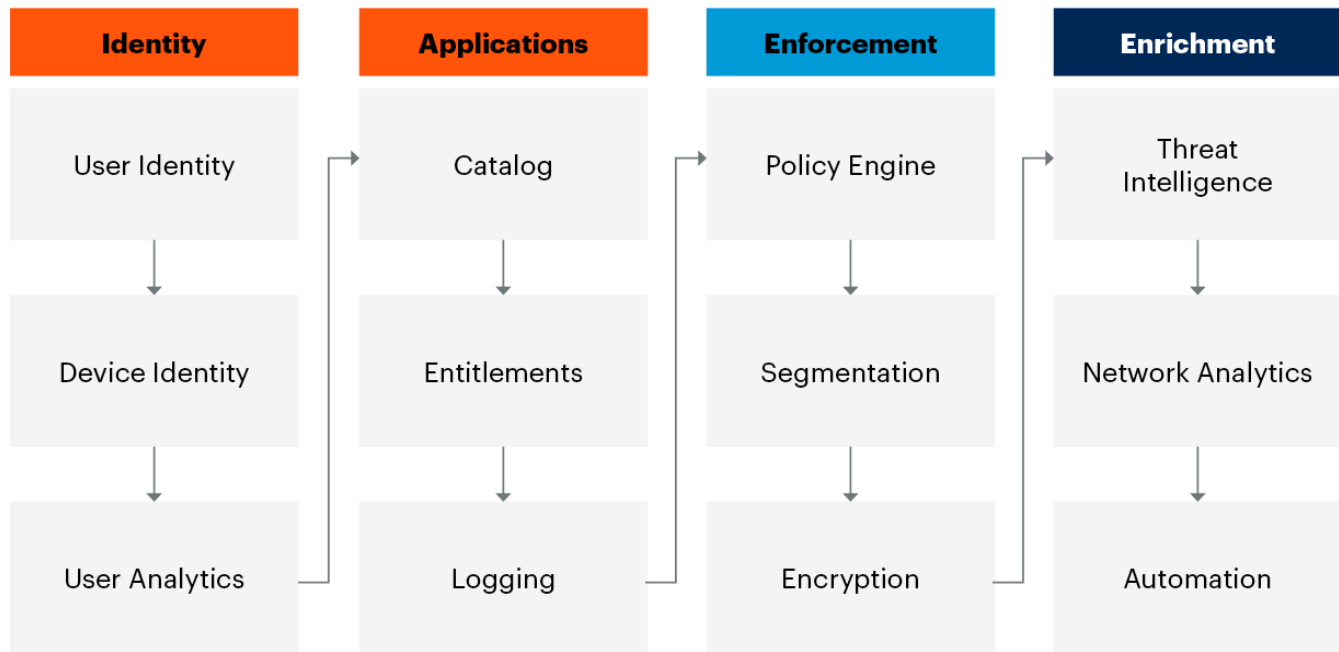
Even for those organizations that are not pursuing a zero trust strategy (aside from limiting the spread of ransomware), there is an additional compelling reason to consider segmentation. Segmentation will provide an effective strategy to mitigate the ability of an adversary to move laterally across the network. It is almost inevitable with social engineering techniques that attackers will gain access to an endpoint on the network and evade endpoint detection. However, segmentation provides SRM with an approach to contain adversaries within a single segment of the network.

Analysis of recent attacks highlights the trend that adversaries are looking to move laterally across the network as quickly as possible. For all of these reasons, adopting a segmentation policy will provide SRM an opportunity to improve the security posture of their organizations. In 2024, there really should be no reason that completely flat networks exist.

Figure 1: Zero Trust Core Tenets



Zero Trust Core Tenets



Source: Gartner
766061_C

Gartner

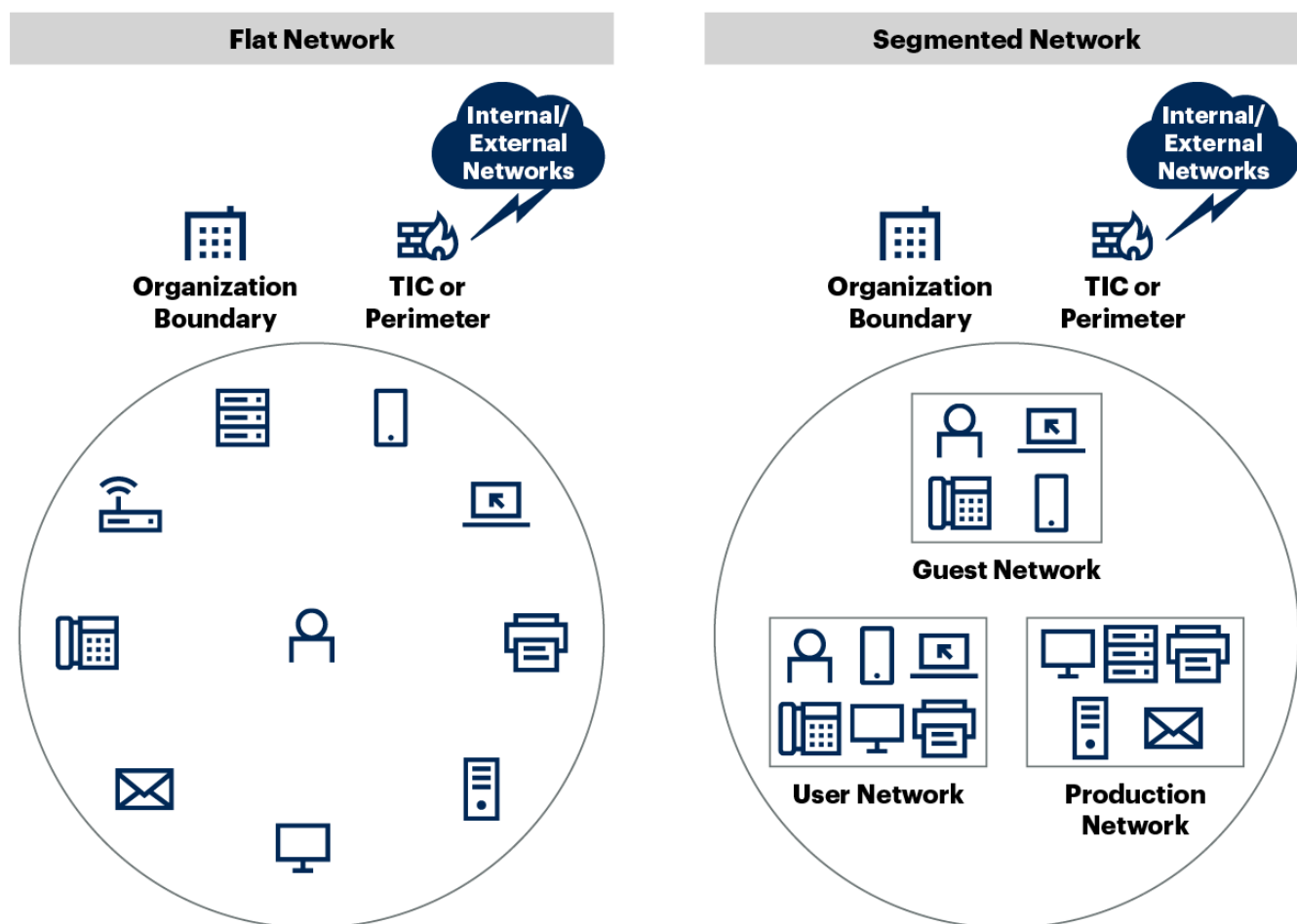
If the network is not segmented then it is referred to as a flat network. The requirement for zero trust is that there is macrosegmentation and microsegmentation of the network. The reasoning behind this is in a completely flat, unsegmented network, once an adversary compromises an endpoint, they can move across the network and to the target of their attacks. By creating segments, SRM professionals can contain the attacker within a given segment. This is shown in Figure 1 under the Enforcement tenet.

Figure 2 illustrates the differences between a flat network and a segmented network.

Figure 2: Segmented Network



Network Segmentation



Source: Gartner

TIC = trusted internet connection

792127_C

Gartner

When approaching segmentation, it is important to understand that no matter what your starting point is (be it a flat network or one that is partially segmented), segmentation will involve a series of incremental steps. In addition to this, before any choice is made around approaches or technology, SRM technical professionals must have a clear segmentation strategy in place that balances security and operational requirements. Nonetheless, there will be impacts to the target operating model for the SOC.

The segmentation strategy that needs to be defined upfront should identify what will be scoped into the initial phases of the phased segmentation approach and what parts of the network can be left for the final phases. The initial phase of the segmentation strategy should focus on segmenting the business-critical assets that need to be protected from adversaries. This will be any assets on the network that are governed by laws or regulations — for example, HIPAA (the Health Insurance Portability and Accountability Act of 1996) or GDPR (General Data Protection Regulation, May 2018). Losing control of this type of asset will bring with it sanctions and the

possible removal of a license to operate. Secondly, in this initial phase, you should scope in any legacy systems that cannot have up-to-date security patches applied to them.

Once this has been completed, the next step will be to identify what is the best approach. This depends on the extent to which segmentation has already been deployed. This research provides a detailed discussion of the different approaches and their limitations. SRM technical professionals will understand how to define what strategy to adopt and what type of technical approach to take (agent or agentless). Most importantly, they will be able to identify the correct approach based on their own network and the practical security benefits of deploying a segmentation strategy. Although there are tangible security benefits to deploying a segmentation strategy, SRM technical professionals will also understand its limitations. Understanding these limitations will allow organizations to understand what additional controls need to be deployed to support segmentation. This will be vital to prevent adversaries from achieving the strategic objective of their attacks.

Benefits of Segmentation

Network segmentation brings other key benefits in addition to zero trust. Some of key benefits of implementation network segmentation are detailed below (however, this is not an exhaustive list).

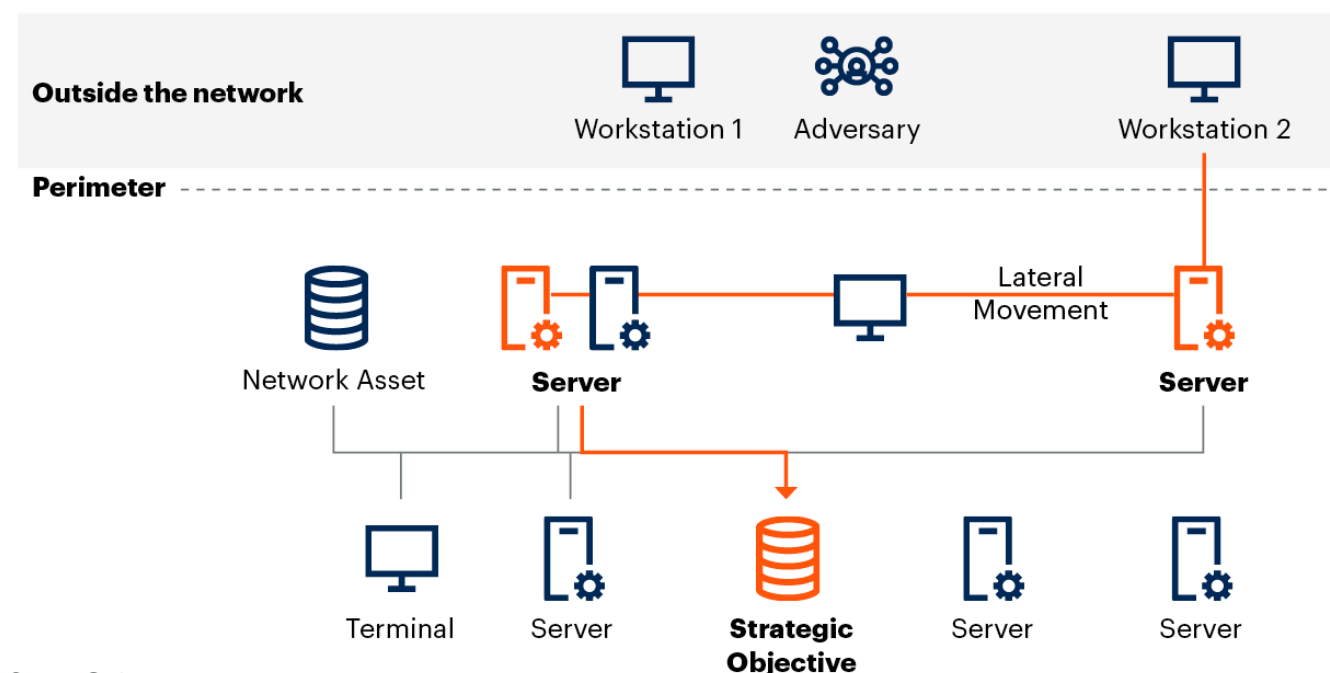
Mitigating Lateral movement

The term “lateral movement” refers to a tactic that is used by an adversary once they have compromised an endpoint on a network to move across or deeper into a network in search of the strategic objective of their attack (see Figure 3).

Figure 3: Lateral Movement Through a Network



Lateral Movement Through a Network



Source: Gartner
795830_C

Gartner

Currently, MITRE ATT&CK lists nine techniques that enable an adversary to achieve lateral movement (see [Lateral Movement](#), MITRE ATT&CK). There are three main components to performing lateral movement. These are reconnaissance, privilege escalation and finally, access. During the reconnaissance phase, the adversary is looking to build up a detailed map of the network. This can be done using built-in tools such as netstat (which shows a machine's current connections) and ARP cache (which gives information mapping IP addresses to MAC addresses).

For privilege-escalation, the adversary can use techniques such as pass the hash or ticket. Pass the hash captures password hashes that, once they are authenticated, would allow the adversary to run commands on a local or remote machine. Pass the ticket uses Kerberos tickets. The adversary would look to create a Kerberos ticket that remains valid indefinitely, even following a password reset. These artifacts can then be used in the final stage of the attack to gain access and, in the process, move laterally to the target system. There has been a tendency to describe segmentation as a means to protect against advanced persistent threats (APT). It does not protect organizations against this type of attack. What it does do is mitigate lateral movement.

Containment

If an adversary or malware (for example, ransomware) is able to compromise an endpoint on the network and there is no segmentation of the network, then the attacker or malware will move freely across the network. For example, suppose a spear phishing attack is sent to a user in human resources and they click on the malicious link. In a flat network (one in which there is no

segmentation), the malware or adversary will move freely to more interesting parts of the network that contain highly sensitive data. In the absence of effective segmentation, ransomware can encrypt the entirety of the flat network and thereby impact business resiliency. Segmentation can and does prevent the spread of ransomware.

Equally important is the ability to contain the insider threat. This is something that often gets overlooked and understandably so. Even if an insider has privileged access to network resources, segmentation will limit that insider to only those resources contained within a specific segment.

Application Owners Gain Visibility

Under the application tenet for zero trust, the first requirement is to build an application inventory. Many organizations simply do not possess such data, and for others, the application inventory is, at best, partial and likely outdated. There are outliers that do have almost-complete datasets. Segmentation can help with the zero trust application pillar in addition to the enforcement pillar. Deploying a segmentation strategy not only helps to build the application inventory, it also helps to identify critical data that is governed by regulations such as PCI Security Standards Council, HIPAA and GDPR.

Cloud Workload Protection

Workloads in the cloud can be defined as consisting of the functions, services, containers, and virtual machines that will store and use data. This will also include the network resources. Before the adoption of cloud services, these workloads were to be found in the data center, so traditional approaches to securing them were applied. Recent developments in cloud services have resulted in the use of multiple microservices and database clusters, and a front end. In this instance, there is a need to apply security at the workload level as data flows through these various components and onto the identity that will interact with it. In addition, many organizations are working with multiple cloud providers and combining this public cloud with their own private cloud deployments. Using segmentation principles, these complex cloud environments can be broken down to the workload level. From there, security policies can be defined both at the segment and workload level. Further, insights can be gained into traffic flows between individual workloads.

DevOps Security

DevOps environments have two key characteristics: they change very rapidly and they have a high degree of automation.

This can lead to them being prone to security threats (for example. external adversaries and malware) and a constant rate of change to the environment can also lead to configuration errors and vulnerabilities not being addressed. Segmentation can be used to control this environment by creating self-contained segments for the different stages of the software development life cycle. The environment should be logically divided into production, preproduction, testing and research

and development. Policies can be created that constrain traffic flow within each of these environments, ensuring there is no flow of traffic to the live production environment.

Legacy Systems

Many organizations have a need to continue to run legacy systems due to operational necessity. Examples include old Solaris servers running Oracle Database, ATM networks running on older versions of Microsoft Windows that are no longer supported and running business applications on out-of-date Linux operating systems. Over time, it is also possible that workloads that use these systems have increased rather than receded. This legacy infrastructure does tend to be found more often in large enterprise networks. These systems will no longer be receiving technical support, including the deployment of security patches. Of course, this makes them a prime target for adversaries, as vulnerabilities will continue to be found in their software. Migrating these systems to the cloud or decommissioning them are not options either. It will also be operationally challenging to get these systems to comply with the core principles of zero trust.

There are a number of different approaches that can be taken to protect legacy systems. The first possible approach might be to place the legacy system into a VLAN. VLANs can be hard to maintain because they can require greater management and configuration overheads. These overheads can include trunking ports that assign VLAN IDs, creating the possibility for human error, leading to configuration errors. This can also lead to a temptation to place all similar types of legacy systems in the same VLAN, meaning that an adversary that moves laterally into said VLAN can reach all of the systems within it. Adopting microsegmentation would allow an organization to build granular security policies across each legacy system and mitigate an adversary's ability to move laterally.

Key Principles

The following key principles apply to both enterprise networks and operational technology (OT), and Internet of Things (IoT) environments and cloud. Gartner defines OT as "hardware and software that detects or causes a change through the direct monitoring and/or control of physical devices, processes and events in asset-centric enterprises, particularly in production and operations." The Internet of Things (IoT) is the network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment.

The principles to follow are:

- Planning segmentation strategy
- Identifying what to segment
- Mapping dependencies

- Defining segments
- Simulating segmentation policy
- Enforcing segmentation policy

Each of these is discussed in depth below.

Planning Your Segmentation Strategy

Central to the planning process is the formation of a cross-functional team from across the enterprise network. This will include:

- A network team
- A security team
- An application development team
- Application owners
- Business stakeholders
- Representatives from the zero trust strategy team (optional; needed only if you are implementing zero trust)

When planning a segmentation strategy, there is a need to take a pragmatic approach. Failure to do so will lead to a failed deployment and will clearly hinder your progress on your zero trust journey. Being pragmatic entails understanding that you cannot segment your entire network in one big-bang approach. That may be an end goal eventually, but at the outset, you will need a more-focused approach. You need to understand what the adversary's strategic objective is, in order to understand what to segment first. Normally, this will be some kind of data repository. There are, of course, outliers to this general statement — including crypto miners, ransomware and saboteurs.

In the case of OT/IoT environments, the targets for attacks will be systems such as industrial control systems (ICSs) or sensors that control the configuration of industrial control systems. OT/IoT environments tend to be less mature from a security and inventory perspective. The first step in the process will be to make sure there is an up-to-date inventory of the environment that ensures there is a single source of truth for the assets. Without this, it will be very difficult to identify all of the high-value assets.

The single source of truth should also identify how these OT/IoT assets should be configured. This will help with mapping out application dependencies — a topic that is discussed in much more detail below. There are two types of asset discovery:

- Active discovery involves the use of probes on the network to discover information such as configuration installed firmware, using standard industrial protocols
- Passive discovery involves using a network appliance to sniff network traffic

That said, this can include applications, services, or anything that is defined as mission-critical. If it is a challenge to identify these critical systems, then perhaps consult your disaster recovery plan, as it should have already defined what systems are mission-critical. In the initial stage of the segmentation strategy, identify your high-value assets. There will be some obvious starting points, such as your HR data or your critical applications. If there is a risk and compliance team in the organization, they can help to identify additional critical datasets.

Identifying What to Segment

Understanding what segments will be created is fundamental, and critical to the success of any segmentation project. Failure to gain clarity and stakeholder support at this stage will guarantee failure. It is not possible, nor desirable initially, to segment all of the network. The most pragmatic approach is to segment areas that:





- Would be most valuable to an adversary
- Are critical to the business
- Represent a vulnerable system that may be the target of an attack itself or could be used to move laterally to another critical system.

The Purdue model can help SRM professionals understand what to segment in an OT environment.¹

For an adversary, the most valuable thing inside a network is data. There are some outliers to this general statement, but data is key. How do you identify this critical data? The first set of data that would be deemed critical to any organization is any data governed by laws or regulations. Examples are provided below in Table 1.

Table 1: Critical Data

Regulatory Agency	Requirement	Benefit

 PCI DSS	Install and maintain a firewall to protect cardholder data.	Utilize microsegmentation to support operational effectiveness of maintaining firewall configuration and auditing.
 SWIFT	Generate a real-time application dependency map; impose segmentation; provide validation.	Utilize microsegmentation application mapping to generate the dependency map, automate segmentation and automate validation.
 HIPAA	Implement a means of access control, including username and PIN.	Microsegmentation can prevent unauthorized users from even accessing the network that the HIPAA records are stored on.
 EU GDPR	Prevent access to communication networks.	Microsegmentation can prevent unauthorized users from accessing the communication networks that GDPR-qualified records are stored on.

Source: Gartner

If it is a challenge to identify this data on the network, speaking with the internal risk and compliance team should help the segmentation team identify which servers, workloads and applications it exists in. Controlling access to this data is critical. If an organization loses control over it (or if it falls into the hands of adversaries) sanctions will be imposed by regulators. This would likely lead to partners and customers losing confidence in the organization.

There should be some datasets, systems and workloads that are very common to the vast majority of organizations that should be the focus of your segmentation strategy. This would include HR datasets, an obvious target for any adversary. If there is no risk and compliance function within an organization, the next place to look for help in identifying critical datasets and systems is your disaster recovery and business continuity documentation. If you discover these do not exist, stop the segmentation project and remediate this issue immediately.

Banking networks will also lend themselves to identifying systems that need to be scoped into the initial stages of a segmentation strategy. The parts of the bank's network that host the ATM network and the Swift (Society for Worldwide Interbank Financial Telecommunication) would therefore be scoped into the initial stage of the segmentation project.

Legacy systems (outdated hardware or software), because they are no longer supported, should also be scoped into the initial phases of a segmentation strategy. These systems are vulnerable to attack because there will be no security updates for them to patch known vulnerabilities. Because of the threat they represent, access to these systems should be severely restricted, as should any network flows between these legacy systems and other workloads on your network.

To begin your segmentation strategy, focus on "quick wins" to get the momentum and the trust of key stakeholders in the early stages of the project. This will also provide an opportunity to clearly map out how segmentation will improve the organization's security posture and limit lateral movement to critical data and systems.

Quick wins can also include environmental segmentation. Different parts of the network will be used for testing, research and development, disaster recovery testing and user acceptance testing (different environments). There should be no traffic flows between preproduction and the production network. These are two very different environments. The quick win here would be to separate the two environments. This approach can be taken and applied to all of the different environments that exist within the network.

Mapping Dependencies

The next step, once the high-value resources have been identified, is to map out any application dependencies. These are the network connections that allow the application to function, so this will include any data stores, middleware components and additional servers that the application interacts with. This information should be included in any standard build documentation. This is the documentation that is normally handed over to the server support team once the server goes into the production environment.

During this process, you can also verify that the information is contained in your disaster recovery documentation. In the absence of this documentation, the alternative approach would be to talk to key stakeholders such as business users, server support, the network team and the developers. If all of the above approaches fail, then it is possible to use the segmentation agent – which is deployed on the server hosting the application – to discover these dependencies. However, the limitation here is that the agent may not discover all of the dependencies. Some processes may only run quarterly, or during end of year processing. If the agent is run in discovery mode for three months to map out these dependencies, the dependency that is only used by a system once every six months may never be captured. These exact same principles will be followed in OT and IoT environments.

Once application dependencies have been mapped, ensure that these are validated with the key business and technology stakeholders and your disaster recovery documentation. It is vital that any segmentation tool is able to support the environments that exist in an organization — be they virtual, bare metal or hybrid cloud environments. Ensure that legacy operating systems and services are documented. It is critical that you do this before choosing a segmentation tool. Following this strategy will ensure that postsegmentation, an organization has a centralized policy management across all environments.

Defining Segments

Once the application dependencies have been mapped, the next step will be to define the logical segments. A logical segment includes servers and systems that are grouped together. A simple guidance principle to apply here is to group together all of the servers and workloads that support the same business function. Business functions could be divided into HR, engineering, marketing and finance. The business function will be unique to each organization. Within each of these business functions, there will be a mixture of high- and low-value assets. Prioritize the high-value assets. Treat each business function as a separate project that will have its own set of objectives. However, this will be just part of your overall project to implement microsegmentation as part of your zero trust journey.

Simulate Segment Policy

To begin with, the segmentation tool should be deployed in test mode to observe traffic. The objective here will be to identify the traffic flow, what ports the traffic is utilizing, and how consistently this is happening over a given period of time. Verify that the correct traffic is being allowed. Check the results with the application owners and business stakeholders. It is possible to simulate policy creation in order to visualize whether the proposed policy being created will break any of the dependencies of the application. Always simulate the policy creation before deploying into live mode. Finally, ensure that the operations and security teams receive the required reports, dashboards, and alerts. At this stage, deploy policies that are general in nature, based on the traffic seen. Once this has been operationalized, SRM technical professionals can fine-tune and apply more granular policies.

Enforce Segment Policy

When enforcing the policy, it will be necessary to communicate the date and time in advance to all key stakeholders. Let everyone know how to escalate issues if something goes wrong. Be ready to fix problems, or even turn off enforcement if problems occur. Remember that policy enforcement will take place on each segment that is defined one at a time. There will be a need to scale up operational support for segmentation once all of the defined segments have moved into a live enforcement stage. Periodically review applications and communication flows and engage with DevOps as new applications are being developed.

When approaching the topic of segmentation, it is important to understand that for most organizations, it will be a journey. If the starting point is a flat network, then creating basic subnets using firewalls is a good first step. You can then move on to environmental segmentation, before finally arriving at segmentation at the workload level. Where a zero trust strategy is being adopted, SRM technical professionals must ensure that segmenting at the network level is only one component of their overall zero trust strategy.

Once the key principles have been defined, the next step in the segmentation strategy is for SRM technical professionals to understand what type of segmentation to deploy.

Types of Segmentation

If the network is not segmented, then it is referred to as a flat network. The requirement for zero trust is that there is macrosegmentation and microsegmentation of the network. The reasoning behind this is that in a completely flat network without segmentation, once an adversary compromises an endpoint, they can move across the network and to the target of their attacks. By creating segments, SRM technical professionals can contain the attacker within a given segment. Segmentation does not prevent lateral movement, it is a mitigation. It is still possible that an adversary could compromise an endpoint and identity that allows them to traverse a segment that has been created. At this point, SRM technical professionals will be relying on the other core tenets of zero trust that they have implemented to help prevent an adversary from achieving the strategic objective of their attack.

Macrosegmentation (Network Zoning)

This approach has its limitations because it only focuses on north-south traffic, which is traffic that goes from the client to the server. As data comes from outside the network, network segmentation is able to examine and filter it. However, if malicious activity is happening within the network, it could go undetected with traditional segmentation.

This approach may isolate all highly confidential applications in the same VLAN or subnet with access controlled via a firewall. One of the main challenges with this approach is one of scalability and the operational overhead it imposes. Access control lists (ACL) can be used in place of a firewall, but they offer lower overall security, while maintaining a high operational costs. For cloud infrastructure, network security groups can be used to create rules that will either allow or deny access to or from cloud resources.

Microsegmentation

Gartner defines microsegmentation as the ability to insert a security policy into the access layer between any two workloads in the same extended data center. Microsegmentation technologies enable the definition of fine-grained network zones, down to individual assets and applications.

Core capabilities include:

- Flow mapping, which is the ability to gather and show north-south and east-west traffic flows and use them in the policy definition (it can present this data in a visual manner)
- Workload isolation, which is isolation from other workloads based on security policy
- Policy enforcement, including the definition of rules based on different factors
- The ability to deploy in virtualized and infrastructure-as-a-service environments

Some of the most frequent optional capabilities of microsegmentation technologies include:

- Threat detection – based on threat intelligence, Layer 7 protocol inspection and anomaly detection.
- Integration with cloud infrastructure to ease deployment, enforce rules or automate policy updates when new assets are deployed.
- Asset discovery – adjacent to the flow mapping, microsegmentation tools can show more advanced context for the assets.
- Policy recommendation engine – complementary to the asset discovery, microsegmentation technology can suggest policy rules to authorize discovered flows.
- Interoperability through direct integration with third-party products, such as a firewall, and hardware, such as switches and routers.
- OT/IoT coverage – that is, the solution supports microsegmentation for IoT/OT infrastructure.
- Kubernetes/Container coverage – that is, the solution supports microsegmentation for containers/K8s.

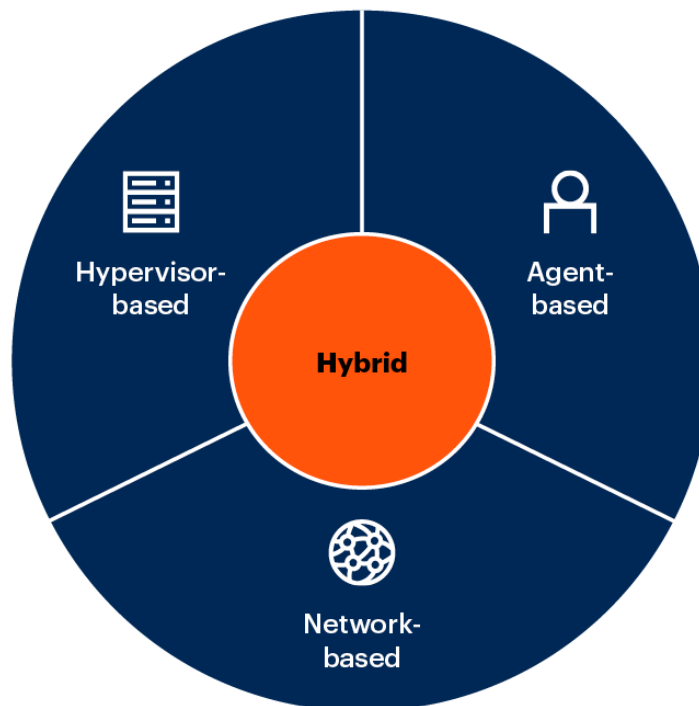
Microsegmentation can apply security policy to traffic that is already within a network, moving east-west between internal servers. Microsegmentation can segment applications within a VLAN from each other. For example, an organization has a CRM and the HR systems in the “highly confidential” VLAN. Microsegmentation will separate the two systems, so servers supporting the CRM system have no connectivity to the servers supporting the HR system. When implementing zero trust, the enforcement pillar specifically addresses macrosegmentation and microsegmentation.

Microsegmentation Techniques

Figure 4 shows the different techniques for microsegmentation.

Figure 4: Microsegmentation Techniques

Microsegmentation Techniques



Source: Gartner
795830_C

Gartner

Each of the three microsegmentation techniques comes with its own set of limitations (detailed in the relevant sections below). In most deployments, a combination of agent and agentless will need to be adopted. There may be some outliers to this general statement, but most SRM technical professionals will be faced with either legacy equipment or infrastructure that will not support the deployment of an agent. Where the network is very heavily composed of a virtualized environment (based on Gartner client interactions), the default approach is for SRM to adopt a hypervisor approach.

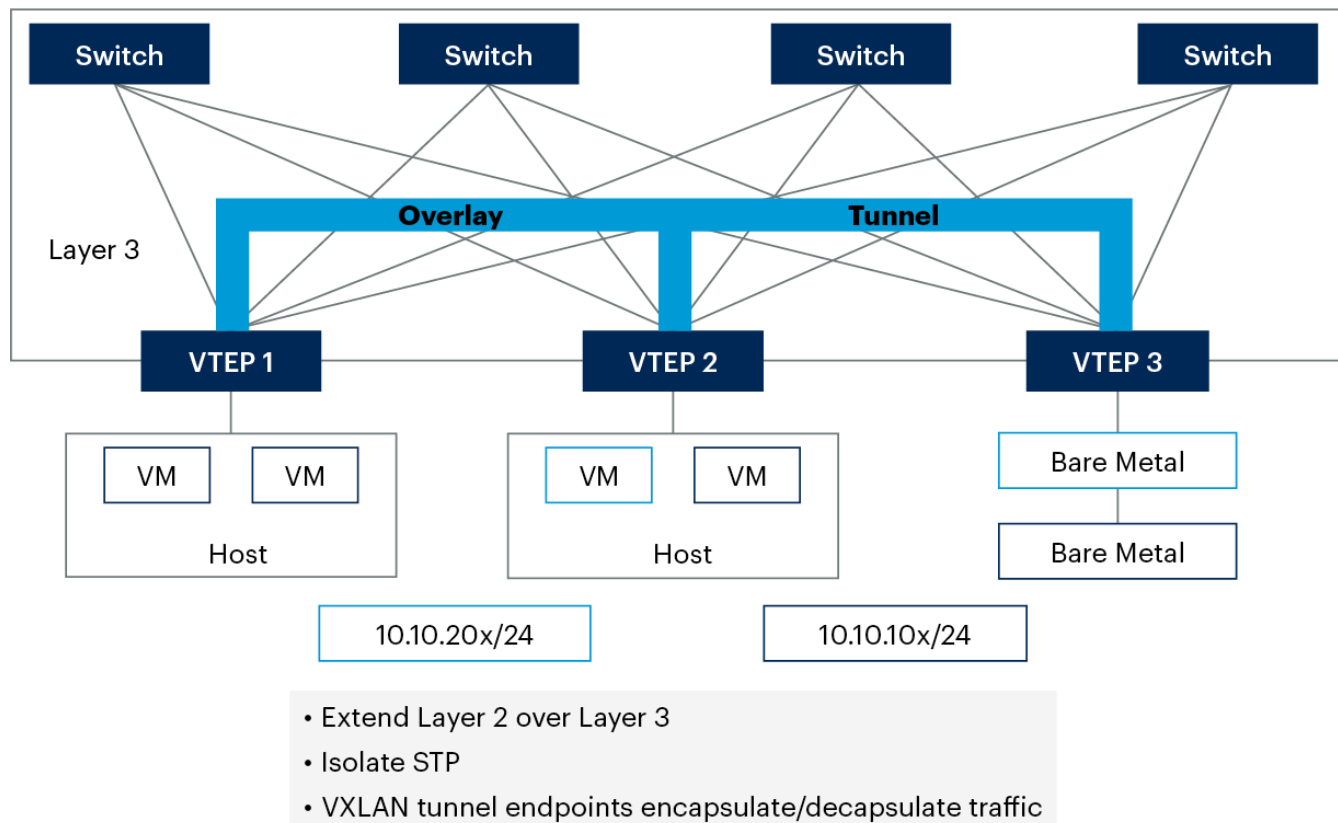
Agentless (Network-Based)

In the type of approach shown in Figure 5, an appliance will be deployed in line with the workloads to be protected. It will use third-party controls to collect and control the traffic. In order to do this, span ports or netflows from existing infrastructure can be used. In a cloud environment (discussed in more detail below), cloud APIs can be used too. Some of the strengths and limitations of this type of technique are detailed below.

Figure 5: An Agentless (Network-Based) Approach



An Agentless (Network-Based) Approach



Source: Gartner

STP = Spanning Tree Protocol; VM = virtual machine; VTEP = VXLAN tunnel endpoint; VXLAN = virtual extensible LAN
795830_C

Gartner

Advantages

- In theory, this type of technique should be simpler to deploy because it does not require the deployment of agents on the endpoint. SRM technical professionals will need to have an asset database identifying the critical endpoints in the network. This also removes the necessity of managing agents on the endpoint, and it removes the agent as a potential attack vector.
- It can be argued that the agentless approach alleviates the challenges around scalability because it does not necessitate the deployment of additional software on the endpoints. This could be particularly useful in very large networks where maintaining agents on a large number of endpoints would impose a significant operational overhead.
- Agent-based solutions can consume more resources on the endpoint (such as memory and CPU) than agentless solutions. In environments where devices have limited resources, agentless solutions may be more suitable.

- This technique is well-suited to infrastructure that cannot, due to resource constraints, have an agent deployed to them. This typically includes supervisory control and data acquisition (SCADA) systems, programmable logic controller (PLC) and healthcare systems.

Challenges

- This type of technique can require the use of hairpinning — forcing the network traffic through a single (or multiple) gatekeeper device on the network. This can add additional cost and complexity (bandwidth constraints) if there is a need for a number of these gatekeeper devices, and by its very nature, this approach will not scale well. The gatekeeper approach will not be able to control the container layer.
- There will be no real Layer 2 network visibility or Layer 7 process visibility into the session. The ability to deliver real-time monitoring is limited, as an agent can continuously monitor the endpoint and deliver real-time data.
- It will not be possible to create highly granular policies for each and every endpoint within the scope of the project. This will cause particular issues if there is data that has compliance requirements.
- It will prove to be challenging to maintain a consistent set of granular policies deployed across the hybrid network environments that exist today (on-premises, virtual and in the cloud).
- This technique is dependent on third-party controls. An agentless approach is dependent on API, netflow and log data to be accurate and up to date. In addition, there may be issues with integrating all API and log file data. There may also be compatibility issues, which limit the detailed security and netflow data that can be collected.

This type of segmentation technique is most suited for environments where general network-level controls are needed and there is no operational need for granular access control. This will be better suited to OT or possibly campus environments. That said, when deploying a zero trust program and interacting with a network resource, one of the key principles is to deploy the principle of least privilege. The ability to deliver this via an agentless approach will be constrained due to the lack of ability to apply granular controls. Working toward this principle will require using other control sets in the other pillars that go up to make the zero trust framework.

Agent-Based

With an agent-based approach, agents will be deployed on endpoints (see Figure 6). The agents will then analyze the traffic. If SRM technical professionals do not have any up-to-date data on their

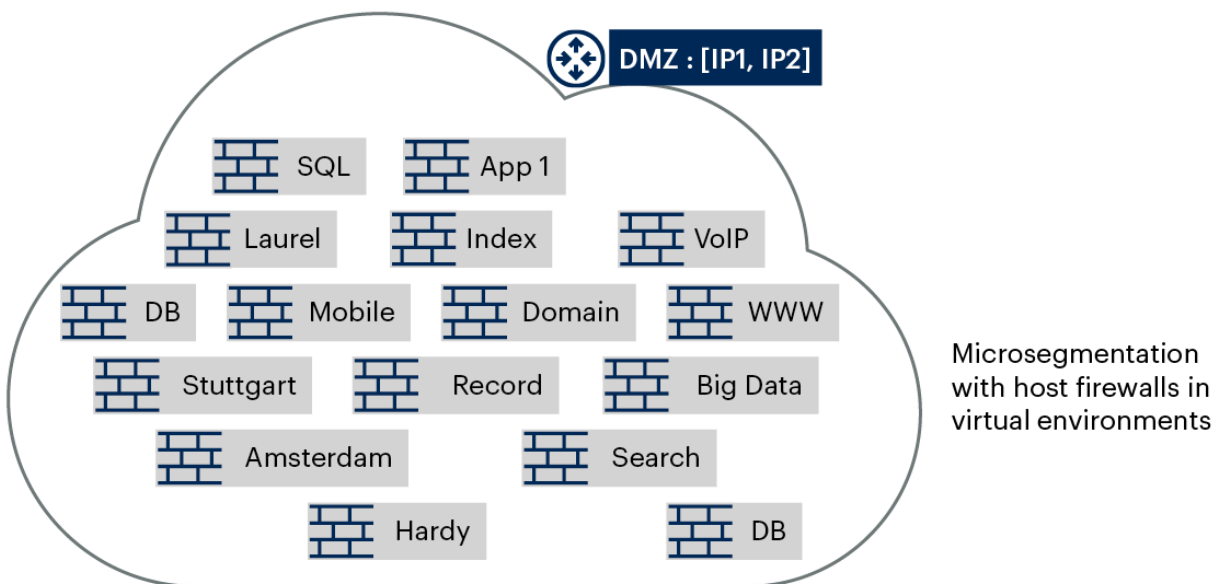
workloads or which workloads are critical, the agent-based approach can allow for a monitoring phase to deliver this data. In this phase, the agent will collect data on the network flows and help to build out a detailed map of dependencies too. This is, of course, only a partial solution to the challenge of defining critical data or applications. The business and application stakeholders will still need to provide context.

Agent-based solutions generally use the Microsoft Windows Filtering Platform (WFP) to enable either the firewall that is built into the OS or the proprietary firewall that is part of the agent software. The WFP filters traffic on OSI Layer 3/Layer 4, and process data collected from WFP is used to make intra-application communication maps and policy decisions. On Linux systems, iptables or nftables can be used to get the connection state table. This can be used for flow monitoring and visibility purposes and to implement Layer 3 and 4 policies, either to allow or deny traffic.

Figure 6: An Agent-Based Approach



An Agent-Based Approach



Source: Gartner

DMZ = demilitarized zone; VoIP = voice over Internet Protocol

795830_C

Gartner

Advantages

- If there is an operational need for visibility into real time process-level activity on the endpoint, then the only way to achieve this is via an agent-based approach. If you are working toward adopting a zero trust approach, then this level of visibility is what SRM technical professionals should be working toward. Initially, granular control is not going to be implemented, but in the long term, that is an objective that SRM technical professionals will be working toward.

- Agents can be deployed across multiple hybrid environments such as bare metal, infrastructure as a service (IaaS) cloud, virtual and hybrid cloud environments (including those of Alibaba Group, Amazon Web Services [AWS], Microsoft Azure, Google Cloud Platform [GCP]). They can also provide visibility into the container-level independently of the container orchestration.
- Where there are compliance requirements such as GDPR or HIPAA, an agent based approach can provide the granular level of control that will help meet these compliance requirements.
- Agent-based microsegmentation products have low complexity and high flexibility. They generally offer network layer (that is, Layer 3) filtering and process verification for all workloads where the agents can run. Devices that can't support an agent, such as IoT devices and mainframes are frequently excluded.

Challenges

- It can become a complex deployment because of the need to deploy agents on endpoints. In large enterprises, this can represent a significant burden. Once the agent is installed, there will be an ongoing operational overhead to ensure it is fully patched. Monitoring will need to be deployed to ensure resource consumption is carefully tracked. Integration into existing ticketing and monitoring systems will need to be designed into the solution and this will have an impact on security operation center (SOC) and support resources.
- The agent itself can become an attack vector and will likely introduce new vulnerabilities.
- Agent support for legacy systems is weak. Careful consideration must be given to the oldest operating systems that are in existence in the network. Some vendors will support older versions of Windows all the way back to Windows XP, but any legacy system running versions older than this would not have out-of-the-box support. SRM technical professionals must have accurate and up-to-date audit data on their legacy operating systems. These statements apply equally to any Linux, UNIX or other operating systems that are deployed, and custom-made operating systems running on OT equipment.

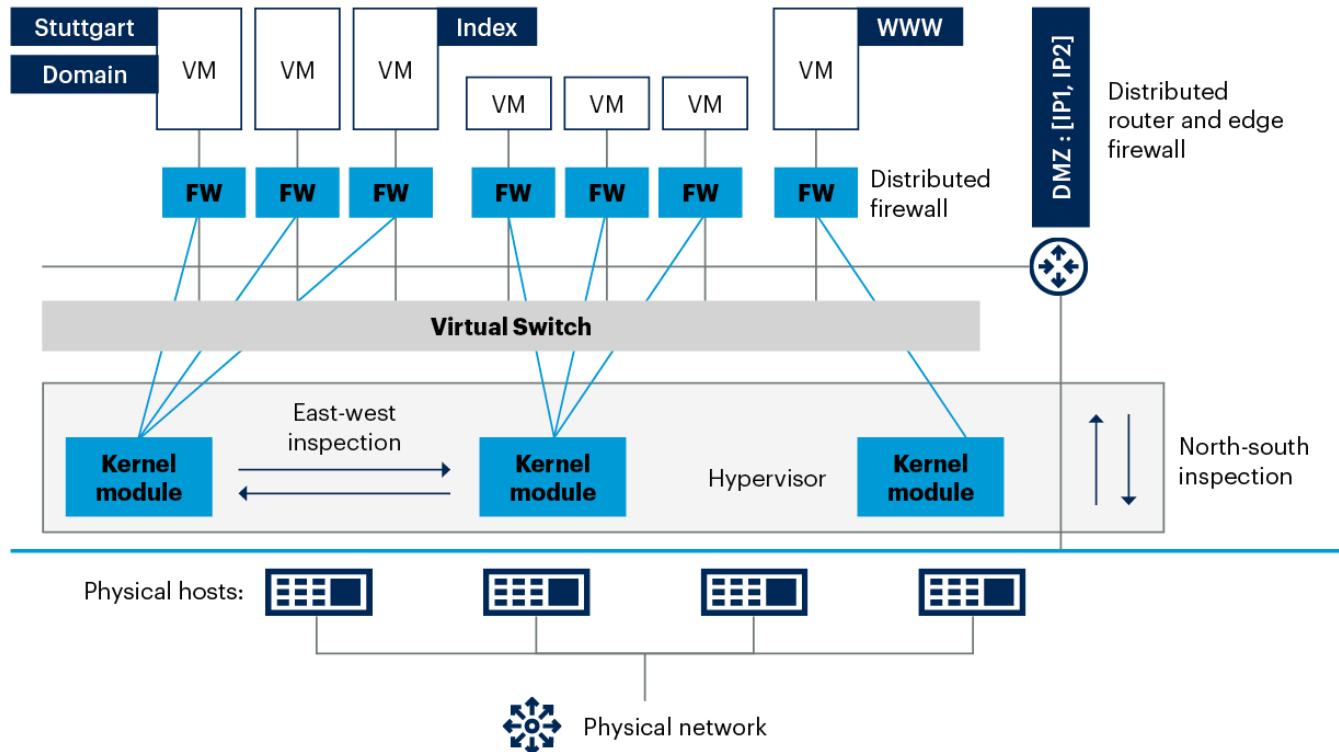
Hypervisor-Based

This is implemented by making use of the hypervisors (also referred to as a virtual machine monitor [VMM]) in the virtualized environment(see Figure 7). The VMM is the software component that will create and run the virtual machine. The VMM allows one machine to host a number of virtual guests by managing the host machine's underlying resources, such as memory and CPU.

Figure 7: A Hypervisor-Based Approach



A Hypervisor-Based Approach



Source: Gartner

DMZ = demilitarized zone; FW = firewall; VM = virtual machine

795830_C

Gartner

The hypervisor will create overlay networks (a virtual or logical network created on top of an existing physical network). Virtual networking results in virtual network cards or switches, which create an overlay that allows for communication between virtual machines or between the hypervisor and the rest of the network.

It is the overlay networks that are used to enforce microsegmentation. Hypervisor-based microsegmentation is very similar to network-based segmentation in the sense that there is no agent to deploy. The main difference is that the hypervisor approach will use hypervisor devices instead of network devices. This approach will allow SRM to use existing firewalls, and policies can be moved to new hypervisors seamlessly.

Advantages

- It does not require changes to network hardware. This can make it easier to deploy, particularly if the environment is VM-based (where over 75% of the environment is virtualized). It is the overlay networks that are used to enforce microsegmentation and there is no requirement to deploy an agent, which can make it potentially easier to deploy and operationalize

- Embedding network security controls inside the hypervisor allows fine-grained policy controls to be exerted over east-west traffic. Policy can be enforced at the virtual network interface card (vNIC) for individual virtual machines. The policy can also be decoupled from IP addressing on individual virtual machines. APIs can be used to allow traffic inspection and control to be embedded as a service in the programmable hypervisor kernel networking stack.
- Its policy constructs are easy for teams to learn because they are similar to network-based segmentation.

Challenges

- The main limitation of the approach is that it does not support bare metal servers, container workloads, or public cloud environments. For cloud infrastructure, customers will not have access to the hypervisor.
- It doesn't provide host-level visibility into its software, processes or vulnerabilities. If you are working toward a zero trust strategy, one of the key foundations is to have a good understanding of what software is already deployed on the network. In many organizations, this information has not been captured, so the lack of visibility will be a significant limitation.
- When deploying segmentation, there is an implicit assumption made that any infrastructure on the network has not already been compromised. The ability to identify processes that are already running on a system can potentially provide insight into whether that system has been compromised. The lack of this visibility limits the ability to identify such threats.
- The type of segmentation technique deployed will be heavily influenced by the type of environment to be segmented.

Environments

The type of segmentation (agent or agentless) technique and tool you choose will be informed by your understanding of the environment into which you are deploying your segmentation strategy.

Data Center

For the data center, all three types of segmentation techniques could potentially be used. For most organizations, the likely approach will be a mix of agent-based and agentless. The agentless approach would be used for legacy systems, IoT devices or any other piece of infrastructure that does not support agent deployment. Select architectures that support proprietary hypervisor APIs if more than 75% of workloads are virtualized and use a VMware hypervisor, when you need microsegmentation with the lowest possible latency

Cloud

So what are some options to deploy segmentation on cloud infrastructure without using a dedicated segmentation tool?

A virtual network (VNet) is an isolated network within Azure and is comparable to Amazon Virtual Private Cloud (Amazon VPC). Each virtual network is isolated by default. Traffic can be controlled by using a number of different options, including (but not limited to):

- Network security groups (NSG) is composed of rulesets that can control the type of traffic that is allowed to communicate with the cloud resource. It is used to secure communication between resources within a VNet.
- Application security groups (ASG) are used to define traffic rules for the underlying VMs that run the workload. This allows for control over source and destination traffic. Here, the rule is created in the context of a label rather than an IP or network address.

These features can be used to create broad zoning and a basic level of segmentation, but this is not microsegmentation. One of the stated objectives of a zero trust strategy is to implement microsegmentation that is controlling east-west traffic. In AWS, there are security groups, which offer broad segmentation. However, there is only resource tagging, which provides a simple means of east-west isolation.

Kubernetes Segmentation

The default setting within Kubernetes does not limit communication between nodes, clusters and pods. Therefore, these components can communicate within the same namespace or between them. It is possible, however, to use network policies or infrastructure layers like service meshes to work toward network segmentation. Network segmentation is not going to allow for granular policies to be applied and that is a limitation of this approach. The pods will need to communicate with each other and SRM will need to identify all the pods where that communication needs to be allowed. SRM will also need to define a list of allowed and restricted ingress and egress public internet communication.

Many SRM technical professionals believe that cloud-native tools for discovery and visualization within an IaaS environment can be weak. In Azure, there are built-in tools that will help SRM technical professionals to segment their infrastructure: resource flowcharts, Vnets and application security groups.

A weakness of using cloud-native options is that there can be very limited granularity options. When deploying zero trust principles across critical infrastructure, the more granular the segmentation, the easier it is to control what entities access the infrastructure and what actions they can perform on it.

With a dedicated microsegmentation product, the asset discovery and visualization process can be automated, and dynamic labeling can be used. This approach will allow SRM technical professionals to group together assets that share certain common features. This will streamline the work involved in creating and managing microsegmentation policies.

Operational Technology

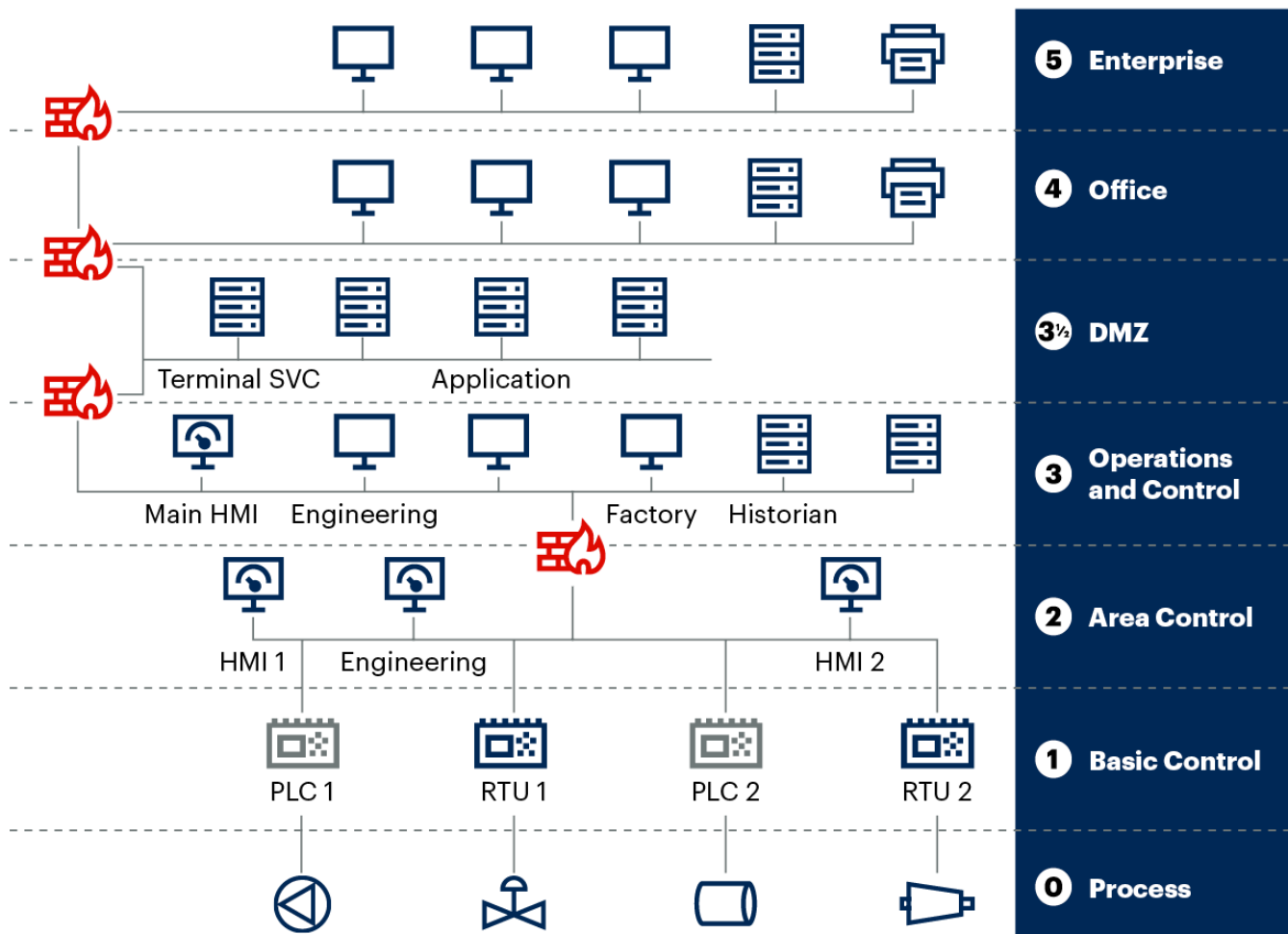
In an environment where OT/IoT is deployed, the segmentation technique that will have to be adopted is the gatekeeper approach, deploying an appliance typically in-line with the infrastructure being protected. Remember here the ability to deploy granular controls will be limited. Figure 8 shows the zones of the Purdue model. The model has defined the standard for creating network architecture in a manner that supports OT security.

In this model, there are six zones that contain the IT and OT systems. The objective here is to implement effective access controls. This segmentation framework, which is specific to OT environments, complements the zero trust approach. This is because both frameworks are helping SRM technical professionals to work toward effective segmentation, and where possible, both frameworks apply the principle of least privilege when identities access networked resources.

Figure 8: OT Segmentation



IT – OT Architecture



Source: Gartner
729470_C

Gartner.

Strengths

Adopting a segmentation policy will:

- Help organizations work toward deploying a zero trust framework.
- Improve lateral movement security.
- Promote containment. This will allow segmentation of the network and constrain an adversary to a certain part of the network, limiting the infection of other parts of the network. Ransomware attacks are a good example of a type of attack that can be constrained effectively by a coherent segmented network.
- Help organizations gain insights into critical applications.

Weaknesses

The weaknesses of a segmentation policy are:

- It can be complex. Segmentation will require extensive planning and will involve multiple stakeholders, such as the application developers and business stakeholders.
- There is a risk of partial deployments. Due to the challenges surrounding deploying an effective segmentation strategy, there is a real risk that it may be abandoned because it is deemed to be too challenging.
- There is a requirement for policy life cycle management. The process of implementing, monitoring, reviewing and updating the segmentation policies created will create a resource overhead. Unless this is effectively managed, the segmentation policies can become dated and no longer fit for purpose.
- Microsegmentation will provide adoption and operational challenges. Initial deployments may run into stakeholder objections.
- Defining policies that address the needs of each internal system can be complicated, and managing the policy life cycle can be one of the most challenging aspects of the postdeployment phase of microsegmentation.
- Lack of human capital will impact an organization's ability to operationalize segmentation. The focus of this research has been on how to create a segmentation strategy, but an equally important postdeployment challenge is the impact the segmentation tool will have on the target operating model for the networking team.
- Postdeployment consideration also needs to be given to how exactly security and policy issues will be divided between the SOC and networking teams. SRM technical professionals must establish where exactly the lines of demarcation are. This will prove even more challenging where organizations may have adopted a hybrid approach to separating out SOC and networking functions.

Guidance

- Use tagging and context of applications, workloads and services, rather than an IP address for segmentation.
- Base segmentation strategies on business risk grouping workloads and data of similar risk levels.
- Start with a network flow mapping project to understand application and server flows.
- Architect for consistent segmentation policies across on-premises and public cloud IaaS.

- Ensure that your segmentation strategy extends into containers and container networking environments.
- Segmenting your network involves a series of steps. First understand what you want to segment, map dependencies and then apply your segmentation rules.
- Prioritize your vulnerable and business-critical applications and systems.
- Begin with taking small steps – for example, understanding what ports are not used.

Acronym Key and Glossary Terms

API	application programming interface
BMU	broadcast, unknown-unicast and multicast traffic
CPU	central processing unit
PLC	programmable logic controllers
SCDA	supervisory control and data acquisition

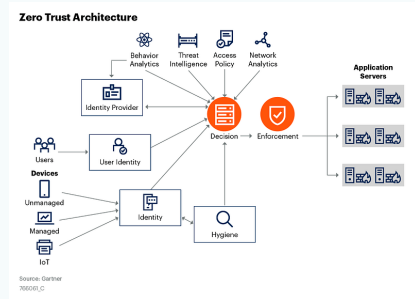
Evidence

¹ [Purdue Model Framework for Industrial Control Systems and Cybersecurity Segmentation](#) (PDF), U.S. Department of Energy.

Recommended by Author

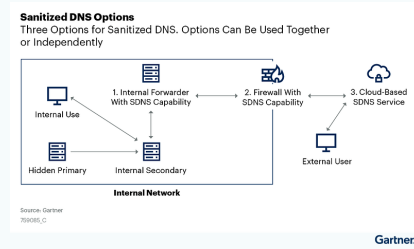
How to Build a Zero Trust Architecture

RESEARCH ▪ 29 September 2022



Guide to Network Security Concepts

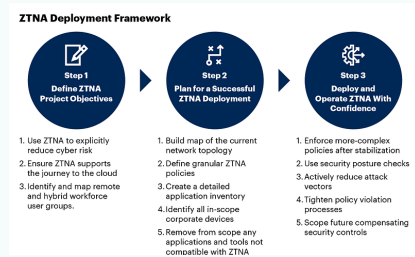
RESEARCH ▪ 13 July 2023



Your Peers Also Viewed

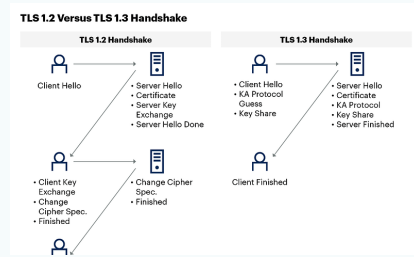
Overcoming the Challenges of Implementing Zero Trust Network Access

RESEARCH ▪ 24 November 2022



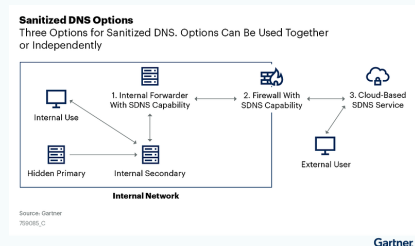
Overcome the Challenges of Implementing and Supporting TLS 1.3

RESEARCH ▪ 22 December 2021



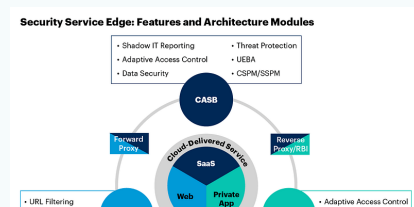
Guide to Network Security Concepts

RESEARCH ▪ 13 July 2023



Adopt Security Service Edge (SSE) to Replace Stand-Alone SWG, CASB and ZTNA Products

RESEARCH ▪ 17 January 2023



How to Choose an EPP/EDR Solution That Fits Your Organization

RESEARCH ▪ 10 May 2023

Guidance Framework for Selecting the Best EPP/EDR Solution for Your Organization

	1	2	3	4	5	6	
Prework	Identify Success Criteria	Assemble Requirements	Identify Candidates	Conduct Deep Research	Perform Proof of Concept/ Lab Tests	Execute Pilot	Follow-up: Implementation
Phase Deliverables	Success Criteria	Requirements	Longlist	Shortlist	Final Prioritized Shortlist, Including Candidate Product	Implementation Decision	
Duration	1-2 Weeks	3-4 Weeks	1-2 Weeks	3-4 Weeks	4-6 Weeks	8-10 Weeks	

Source: Gartner
782018, G

Gartner

Supporting Initiatives



Security Technology and Infrastructure for Technical Professionals



© 2024 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

[POLICIES](#) [PRIVACY POLICY](#) [TERMS OF USE](#) [OMBUDS](#)

[CONTACT US](#)

© 2024 Gartner, Inc. and/or its affiliates. All rights reserved.

Get The App

