BlackRidge TECHNOLOGY

Zero Trust Reference Architecture Operation

October 2019, version 10

Executive Summary

This zero trust reference architecture supports the six pillars of a zero trust security model[1] from the network to the users and applications, with each component and interaction being individually authenticated and providing continuous authorization during operations. This architecture is designed to be deployed in cloud environments with resilience[2] to anticipate, withstand, recover and evolve from attacks. The described architecture includes components and functions described by their respective capabilities.

Introduction

Zero trust is an approach to address deficiencies in cyber security in our evolving cyber environments. Zero trust has the following minimum characteristics:

1. Segregate users, devices, data, and services, within a trust framework, to ensure every access request is validated and deliberately permitted or disapproved;
2. Be resistant and resilient to attack without a large administrative burden; and
3. Be able to easily and rapidly (if not automatically) adjust to an ever-changing service environment also without a large administrative burden.

Meeting these requirements requires that zero trust services operate on a shared network and compute infrastructure that has entities of varying authorities and trust requirements. This shared network and compute environment are the various cloud architectures that are being adopted. One of the biggest challenges facing cloud adoption is security. A zero trust architecture enables secure cloud deployments by enforcing different authorities based on the confidence of trust in the requestor. Enforcing different authorities in a cloud environment has new isolation requirements. A zero trust environment enforces isolation between different entities based on their identity and assigned authorities. This isolation is specifically necessary in environments operating with shared network and compute infrastructure such as clouds.

This isolation requirement further requires that identity be authenticated, and requests authorized on a per network session basis. The challenge with existing network infrastructure and security devices is that there are no elements that can be authenticated at the network layers today. Identity is mostly deployed at the application layer and is thus only available for authentication after network sessions have been established, TLS being a good example. The identity available at the network must be able to be authenticated and must uniquely identify a single network client.

The audience for this document is the Zero Trust team including vendor partners and the deploying customer.

[1] "Zero Trust Cybersecurity, Current Trends", April 18, 2019
[2] "Cyber Resiliency Engineering Framework", MITRE, September 2011

Zero Trust Requires New Solutions for Network Isolation

Many traditional isolation technologies use groups, both implicit and explicit. Implicit group technologies, such as 802.1x, authenticate a network client and then allow that client to access an assigned VLAN or subnet. The group is implicit because it is not communicated during network communications, the network communication is allowed as an authorized consequence of the initial authentication. Explicit group technologies, such as group tags, authenticate a network client and then attach a group tag to all the client's communications. This allows different groups to share a common network. In both cases, the group determination is performed at network ingress. Downstream, the identity used to grant access to the group is not available, only the group when explicit. Thus, different policies cannot be implemented for different group members. To allow access to a new resource, the policy for the entire group must be modified or a new group must be created. A network session cannot be attributed to a specific identity, user or device by a downstream asset.

Building a zero trust environment for different operating networks within a cloud's network infrastructure requires that multiple independent clients with varying authorities operate on the same infrastructure. This requires new isolation technologies, as group based solutions cannot provide the necessary isolation in a zero trust environment. The limiting factor in implementing a zero trust environment using group isolation technology is that different groups of different authorities are prohibited from communicating with each other by design. Allowing groups of different authorities to communicate causes the merging of authorities and removes isolation between the communicating groups. This merging of groups violates the principle of least privilege. After merging groups and their associated authorities, all members the group have access to all resources of both original groups, irrespective of their identity and privilege requirements.

A new solution is needed to provide isolation while still allowing authorized communication between participants. A per flow identity, authority and privilege access control technology provides the ability to deliver on the principle of least privilege, which, unlike group based technologies, enables downstream asset access controls with end-to-end identity.

To provide the necessary isolation in a zero trust environment, individual sessions need to be isolated. Flows (sessions) are defined by a 5-tuple including the source and destination address, the source and destination port and the protocol. Each client establishes multiple concurrent sessions to multiple servers and each server supports multiple concurrent sessions from multiple clients. All sessions are individually authenticated before session establishment is allowed.

Zero Trust Architecture

The following system architecture has been developed to implement the zero trust principles. Figure 1 shows a reference zero trust architecture, including the data plane and the control plane.

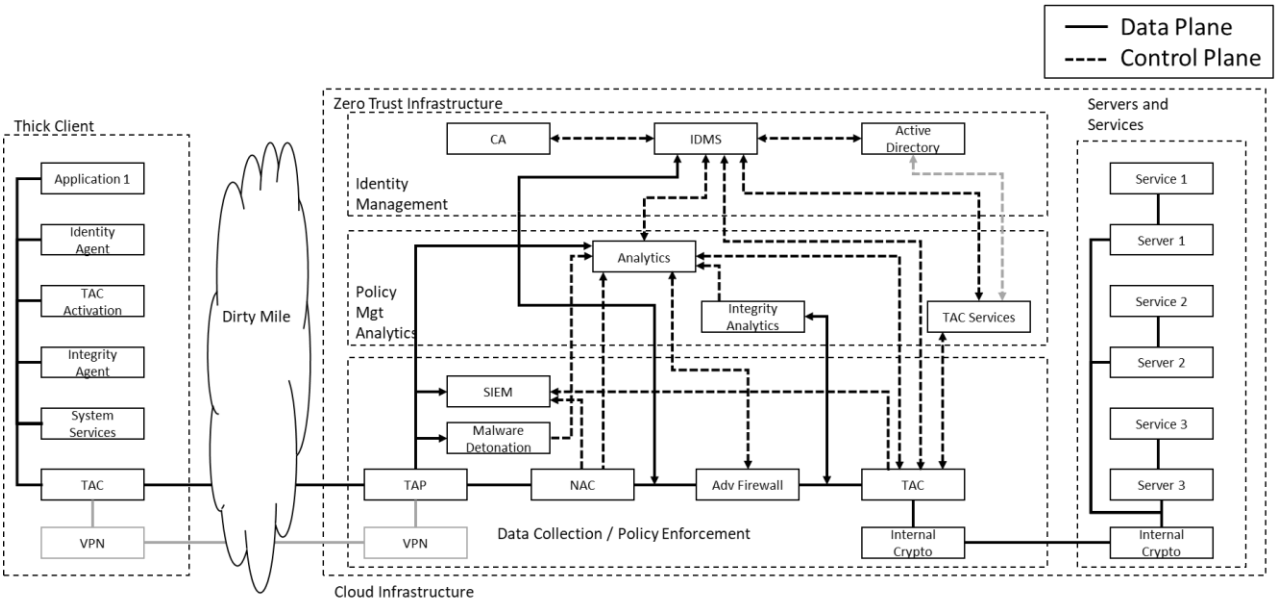

Figure 1

Components

The following section describes the functions of the various components of the reference zero trust architecture.

Advanced Firewall

The firewall uses traditional stateful inspection coupled with content inspection to enforce security policies.

Analytics

An analysis system that uses information from multiple sources to detect activities and trends that cannot be seen with single element tools. Analytics may use behavioral, statistical and other means to detect anomalies. Advanced analytics systems may also be able to instruct policy enforcement components to take constraining actions against an entity exhibiting anomalous behavior.

Client

The client is the operating system and operating platform (physical or virtual) that hosts the applications, the endpoint integrity software, the identity agent and the TAC agent and client.

Endpoint Integrity

An integrity agent on the remote client initializes and performs its endpoint evaluation and communicates with an integrity analytics service. The integrity analytics service validates the software and applications on the remote client. The endpoint integrity system provides continuous reporting and protection against endpoint resident malware.

Identity Management

An identity management agent on the remote client authenticates the provided identity credentials with an identity management service. Some identity management systems determine policy group by querying an AD server. The identity management system communicates with a certificate authority (CA) to authenticate PKI certificates.

Malware Detonation

A malware detonation device operates by attempting to cause suspected malware to perform prohibited behaviors, thereby demonstrating malicious properties.

NAC

A NAC function quarantines all traffic from a remote client until a determination is made that the remote client is in a compliant security state. The quarantine may be enforced by assigning traffic to a remediation or quarantine VLAN until all compliance requirements are met. The NAC has the capability to interrogate the remote client to determine the remote client's security state. If any required aspect is not in compliance, the NAC function places the remote client into a remediation VLAN until compliance is (re)established.

SDN Environment

A software defined network (SDN) is virtualized environment in which network devices and security policy enforcement infrastructure are instantiated. SDN is the next step beyond network function virtualization (NFV) in that NFV components are orchestrated to operate together.

Server

The server is the operating system and operating platform (physical or virtual) that hosts various services.

SIEM

A SIEM data collector maintains a record of all network traffic. This traffic information provides a history of all traffic and the SIEM data repository can be used as a source for correlation and for forensic purposes.

System Services

The services provided by the client's operating environment.

TAC

A TAC (Transport Access Control) function performs non-interactive authentication of each and every TCP session. The TAC function behavior is dynamically managed by the analytics function using the Confidence Level API.

VPN

A VPN function provides confidentiality and integrity to the network traffic passing through it.

System Operation

Using Figure 1, we show how application 1 on the remote client to the left of the drawing gets authenticated and authorized to access service 1 on server 1 on right of the drawing.

It is assumed that the remote client has been allocated an IP address via DHCP locally and been provided a route through the dirty mile that can access server 1. The remote client is connecting across the "dirty mile". The "dirty mile" is an unsecured infrastructure of legacy networks and equipment. The "dirty mile" is be assumed to be hostile. To protect against Man-in-the-Middle attacks and data interception, transit across the "dirty mile" is protected by a VPN. The VPN encrypts all traffic between the RC and the zero trust infrastructure. The traffic is decrypted at the zero trust infrastructure boundary, allowing the zero trust components to perform their tasks. Additional encryption can be used on the server facing side of the zero trust infrastructure, as is show in the diagram.

Upon entering the zero trust infrastructure, all traffic first passes through a network tap. All traffic passing through the network tap is mirrored to a SIEM data collector, a malware detonation device and an analytics system. The SIEM data collector maintains a record of all network traffic. This traffic information provides a history of all traffic and is used as a source for correlation and for forensic purposes. The malware detonation device operates out of band with respect to the data streams. If the malware detonation device detects malware, this is indicated to the analytics system.

After passing through the network tap, the traffic hits the NAC device. The NAC device quarantines all traffic from the remote client until a determination is made that the remote client is in a compliant security state.

Once the remote client is allowed access from the NAC device, the identity agent on the remote client authenticates the provided identity credentials. The identity management system may determine policy group by querying the AD server.

Once the remote client's credentials have been authenticated, the endpoint integrity agent on the remote client initializes and performs its endpoint evaluation and communicates with the endpoint integrity service. To communicate with the endpoint integrity service, the traffic passes through the advanced firewall. The advanced firewall uses content inspection to monitor all traffic passing through it and enforce security policies. The advanced firewall is continuously updated by the analytics system. The analytics system, in addition to the information provided to it by the advanced firewall, receives telemetry information from the IDMS, AD, the malware detonation device, the NAC system, the endpoint integrity service and TAC gateways among other sensor sources. After passing through the advanced firewall, the endpoint integrity agent communicates with the endpoint integrity service, validating the software and applications on the remote client. The endpoint integrity agent provides continuous reporting and protection against endpoint resident malware.

Once the endpoint integrity agent has initialized, the TAC Activation agent communicates with the TAC services. During the activation process, the remote client's identity credentials are again authenticated using the IDMS, and the authorized identity is activated within the TAC system. The TAC Activation agent is initialized to enable the generation of TAC identity tokens. Within the TAC system, the authorized identity and its associated authorities and enforcement policies are communicated to TAC enforcement points.

At this point, the zero trust aspects of this system have been initialized and are ready for application communications.

On the remote client, application 1 is started. Application 1 operates because it is authorized by the endpoint integrity agent. An application not authorized by the endpoint integrity agent will be blocked from operating, denying access to network communications. Application 1 initiates a TCP/IP session to service 1 on server 1. On the first packet of this TCP/IP session, the TAC client inserts a single use identity token into the TCP SYN packet. The TCP syn packet is communicated across the "dirty mile" to the zero trust infrastructure. The network tap copies the packet and sends it to the SIEM and the malware detonation device. To prevent replay attacks, TAC Tokens expire after 4 seconds if it is not received and is immediately invalidated upon reception. A SIEM recording TCP sessions with TAC tokens does not enable replay attacks.

The original packet is then forwarded on to the NAC device. The NAC recognizes the source address as being an address that has previously been authorized and forwards the packet to the advanced firewall. The advanced firewall performs its policy enforcement. Assuming that the source IP address of remote client is authorized to access server 1, a session table entry is made in the advanced firewall and the packet is forwarded to the TAC Gateway.

TAC shields the protected resources from unauthorized scanning, discovery and access at the network layer. TAC protects the protected resources, not the zero trust infrastructure.

The TAC gateway extracts the TAC identity token and tests its authenticity by looking in its token cache for a matching token. Assuming a matching token is found, the token entry in the token cache is invalidated to prevent replay attacks and the associated security policy is enforced. Assuming that the remote client is authorized to access server 1, a session table entry is made in the TAC Gateway, a log record with the session attribution information is sent to the SIEM, and the packet is forwarded to server 1. In the future, the TAC attribution logs can be sent directly to the relevant analytics to create a faster closed loop system. On server 1, the TCP SYN packet is responded to by a TCP SYN/ACK packet to establish the TCP/IP session. The TCP SYN/ACK it sent back to remote client along the same path. Once the session is established, application 1 communicates application data with server 1.

There are three points of continuous monitoring in this architecture; endpoint integrity monitoring the remote client behavior, advanced firewall monitoring content and TAC monitoring the identity used for session establishment.

It is assumed that applications may establish a TLS encryption session on top of the just established TCP session. This will not affect BlackRidge operation. A TLS session may affect firewall operation if the TLS session is not terminated at the firewall to allow content inspection.

There are several options for providing confidence feedback using TAC's confidence level API:

1. The advanced firewall or the analytics system detects an anomaly that causes a change in confidence level. The analytics system indicates this change to TAC using the confidence level API.
2. The endpoint integrity agent detects malware and an indication of this detection is sent to the endpoint integrity service. The endpoint integrity service sends an indication of this to the analytics system. The analytics system determines that this will cause a change to the confidence level. The analytics system indicates this change to TAC using the confidence level API.

3. The IDMS detects a condition that causes a change in confidence. The IDMS indicates this change to TAC using the confidence level API.

It is important to recognized that any identity information obtained by monitoring session content, as is performed by the advanced firewall, occurs after TCP/IP session establishment. The TCP/IP session establishment process, by default, does not provide per session identity and is unable to be authenticated. Using TAC adds the ability to authenticate each TCP/IP session on the first packet, blocking network scanning, mapping and discovery by unidentified and unauthorized users and devices. TAC provides session identity attribution for each TCP/IP session on the first packet to SIEM and analytics systems.

TAC Capabilities and Enabling Technologies

Within the described zero trust architecture, TAC provides the following resilient capabilities applied to network traffic:

- Adaptive Response
- Deception
- Dynamic Representation
- Privilege Restriction
- Segmentation
- Substantiated Integrity

Adaptive Response is to take actions in response to indications that an attack is underway based on attack characteristics. TAC can adaptively change policies for an identity based on an authenticated identity's assigned confidence.

Deception is to use obfuscation and misdirection (e.g., disinformation) to confuse an adversary. Deception can take the form of dissimulation ("hiding the real") or simulation ("showing the false"). TAC blocks scanning, mapping and discovery of network resources by unauthorized users and devices.

Dynamic Representation is to construct and maintain dynamic representations of components, systems, services, mission dependencies, adversary activities, and effects of alternative cyber courses of action. TAC operates independently of network addresses and topology.

Privilege Restriction is to restrict privileges required to use cyber resources, and privileges assigned to users and cyber entities, based on the type(s) and degree(s) of criticality and trust respectively, to minimize the potential consequences of adversary activities. TAC provides the ability to access network resources based on the privilege assigned to the authenticated identity.

Segmentation is to separate (logically or physically) components based on pedigree and/or criticality, to limit the spread of or damage from successful exploits. Segmentation reduces the attack surface and enables more cost-effective placement of defenses based on resource criticality. TAC provides network segmentation based on authenticated identity.

Substantiated Integrity is to ascertain that critical services, information stores, information streams, and components have not been corrupted by an adversary. TAC activation uses authenticated identity credentials to enable a system to provide authenticated session establishment.

TAC provides these capabilities using three technologies; Transport Access Control, TAC Activation and Confidence Level API. The individual capabilities of each technology are shown in the table below:

	Adaptive Response	Deception	Dynamic Representation	Privilege Restriction	Segmentation	Substantiated Integrity
TAC		X	X	X	X	X
TAC Activation				X		X
Confidence API	X		X	X	X	

Transport Access Control (TAC)

TAC works by inserting identity into the first packet of a TCP/IP session to communicate that identity across the network. This allows TAC protected services to determine and authenticate the identity of the requester before responding, much like Caller-Id is used in the telephone system. Using TAC, protected network resources only see and respond to session requests authorized by the requesting identity. Requests from unauthorized identities produce no response, effectively disconnecting that network resource from unidentified and unauthorized requests. TAC enables organizations to identify requests and manage discovery and access to network resources with precision previously unachievable by traditional approaches. TAC conceals network assets from scanning, discovery, and access in physical, virtual, cloud and SDE environments.

TAC makes network resources unresponsive to unauthorized users and devices. Unauthorized users can't see the network, can't map the network and can't mount an attack through the network. Network resources protected by TAC act like they have been unplugged when an unauthorized user attempts to scan, probe or access them- those network resources are not there. Authorized users have full access to protected resources. Making the network invisible greatly reduces the cyber-attack surface, increasing overall security. TAC can also prevent unauthorized outbound session establishment by requiring authorization prior to session establishment.

TCP/IP, the protocol of the Internet, always responds to network connection requests. This is a primary mechanism that is used for network scanning, mapping and reconnaissance. Before TAC, there was no effective way to block scanning and reconnaissance without blocking access to the network and cloud resources being scanned. TAC blocks network scanning and reconnaissance from unauthorized users and devices, while allowing authorized scanning. This prevents unauthorized discovery and awareness of network and cloud resources. The ability to selectively block scanning is a feature unique to TAC.

TAC operates by combining network identity with non-interactive authentication. This applies identity to a more foundational location than is commonly used today; applying identity to the network in addition to being used at the application. Using non-interactive authentication allows authentication to occur with a single packet, eliminating unauthorized network scanning, mapping, reconnaissance and discovery. Together, network identity and non-interactive authentication make the network invisible to unauthorized users and devices.

Activation

The TAC activation service allows a provisioned TAC client (software or TAC-ID) to securely activate into a TAC system. The process is initiated by an activation agent resident on the TAC client. The activation agent provides identity credentials to the activation service. The activation service authenticates the provided credentials and determines the associated identity group. The activation service then provides the activation agent with a session key and configuration information, enabling TAC token generation. The activation service causes the session key to be distributed to all TAC enforcement points in the associated identity group. Session keys are non-persistent, they are not stored across power cycles or reboots and session keys have a maximum lifetime of four hours. After the session key expires, it must be refreshed by the activation agent restarting the activation process.

The identity credentials used by the activation service today are simple identities; a certificate associated with a person or a device. In the future, composite identities will be supported. Composite identities are identities composed of multiple credentials such a user + device + application. Additional metadata such as geolocation can be added to composite identities. When used with composite identities, the activation service may provide multiple session keys, with each key assigned to a different combination of credentials. For example, different session keys could be provided to different applications operating on the same device by a single user.

The activation service can be enhanced to provide continuous authentication of the provided credentials. Continuous authentication is the periodic re-authentication of the credentials provided by the activation agent. Re-authentication can include requiring the activation agent to resubmit credentials, binding information and other confirmatory information. Authentication failure during continuous authentication can result in the invalidation of the session keys or a lowering of the associated confidence level.

Confidence Level API

The confidence level API allows analytics systems external to TAC with the capability to execute actions to adjust the authority of individual identities within a set of pre-configured authorities. This enables predictable and deterministic operation independent of network topology and source addresses. Most analytics systems operate as scoring engines, providing a score indicating the confidence in a given identity. Analytics systems using the confidence level API map score ranges to confidence levels. For example, a score of 85-100 is confidence level 6, a score of 50-84 is confidence level 5 and a score below 50 is confidence level 4. When an analytics system detects a change in confidence level, the new confidence level and the identity is communicated via the confidence level API to a TAC enforcement point. The TAC enforcement point enforces all traffic policies for the provided identity according to the new confidence level policy. This approach does not require that the analytics system have any implementation knowledge of a TAC system or knowledge of the network topology. TAC communicates identity confidence levels to all TAC policy enforcement points.

Summary

This zero trust architecture supports the six pillars of a zero trust security model from the network to the users and applications, with each component and interaction being individually authenticated and providing continuous authorization during operations. This architecture is designed with resilience to anticipate, withstand, recover and evolve from attacks.