Credentialing-enabled Zero Trust Architecture for supply chains

Authors: Carsten Stöcker (Spherity) and Christiane Wirrig (Spherity), inspired by Sam Smith

Abstract

Traditional network security focuses on perimeter defenses, but many organisations, systems and processes no longer have a clearly defined network perimeter.

To protect a modern digital enterprise, companies need a comprehensive strategy for securely accessing their IT resources (e.g. applications, physical access control systems, portals, data resources, and devices) wherever they are located.

Zero Trust Architecture (ZTA) refers to security concepts and threat models that no longer assume that actors, systems or services operating within the security perimeter are automatically trusted, but instead must verify everything and everyone who attempts to connect via an API to their systems resources before granting access.

Hence, ZTA is an important design philosophy to establish security mechanisms at the API layer of each individual IT resource for increasing API Endpoint Security in both, corporate infrastructures and open systems.

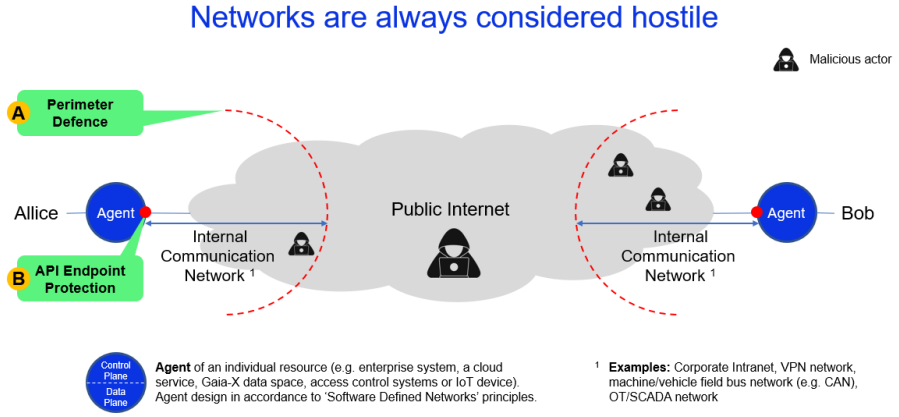

Figure 1: Moving from Perimeter Defense to API Endpoint Protections

The following draft RWOT topic paper introduces the concept of ZTA and describes how the combination of an enhanced identity governance and trust framework with using DIDs/VCs for authentication and authorisation can help digital value chain actors to move towards the implementation of ZTA principles.

1. Motivation of the Paper

APIs in supply chains and cyber-physical systems (CPS) are proliferating exponentially across the technology landscape, creating a huge attack surface that security teams struggle to understand and defend.

APIs are not inherently insecure, but the vast amount being deployed across the technology landscape creates significant challenges for enterprise security teams. In addition, interconnected (cyber-physical) processes are evolving across multiple organisations, requiring the development of a common security architecture.

Malicious actors in the digital world are professionalising at an insane rate, launching sophisticated, AI-powered cyber attacks to compromise endpoints and sell stolen credentials through Initial Access Brokers (IABs) to others who then launch ransomware attacks or steal and sell internal information.

Moving from theory into practice with ZTA principles is extremely challenging unless we start with a key component of ZTA: effective and secure endpoint management within a defined trust framework.

The biggest problem with API security lies in the building blocks of authentication and authorisation. This has become more important as APIs expose more and more critical data and services.

Endpoint visibility, controlling and knowing who is accessing what with which device and the security posture of that device, is a useful entry point and key building block of Zero Trust.

2. Introduction to ZTA

The zero trust architecture model, also known as perimeterless security, describes an approach to the design and implementation of IT systems.

The main concept behind zero trust is that devices and IT-systems should not be trusted by default, even if they are connected to a managed corporate network and even if they were previously verified. This main concept is directly applicable to cross-company ERP processes: IT-systems of counterparty supply chain actors cannot be trusted by default for accessing or processing pharmaceutical supply chain data.

The zero trust approach advocates threshold security such as mutual authentication, including checking the identity, authorization status credentials, and integrity of devices, and providing access to applications and services based on the confidence of device identity and device health in combination with user authentication.

ZTA principles were originally developed when working with users and devices in an enterprise context. Supply chain use cases, however, are working with enterprise ERP systems in an inter-organisation context.

The originalNIST ZTA paper (NIST Special Publication 800-207, Zero Trust Architecture, Zero Trust Architecture) was focussed on enterprise-centric use cases. In the NIST paper inter-organizational collaboration and cross-enterprise scenarios are barely mentioned.

The requirements for secure authentication and authorisation for providing access to data in an inter-organisational supply chain context are very similar to those of the ZTA context, e.g. providing access to a supply chain resource service based on identity and authorization credentials. Therefore some ZTA principles and implementation approaches could be transferred to the supply chain solutions. It shall be noted that projects such as Gaia-X are following the same principles by integrating trust frameworks, DIDs and VCs for API endpoint authentication and authorisation e.g., for providing access to membership portals, enterprise resources and so-called data spaces.

2.1 ZTA for Supply Chain Systems

The following paper explains the ZTA implementation patterns using DIDs/VCs by analyzing paragraphs from the original NIST paper and blending ZTA approaches with a supply chain credentialing solution. Supply chain implementation can even derive conformance criteria for such an implementation from this analysis.

We will also summarize how blended ZTA principles can be implemented through the integration of policy agents with identity wallets, credentialing and a verifiable credential issuance for identity and membership credentials.

This integration of credential-enabled policy agents with existing supply chain systems improves the security and compliance of data exchange capabilitiesand therefore supply chain security as a whole.

2.2 Benefits of Zero Trust for Enterprise Infrastructures

Typical enterprise infrastructures have grown increasingly complex. A single enterprise may operate several internal networks, remote offices with their own local infrastructure, remote and/or mobile individuals, cloud services and may need to establish trust relationships with counterparty- enterprises while providing access to authorized enterprise services and data.

This complex system has led to the development of a new model for cybersecurity known as "zero trust" (ZT). A ZT approach is primarily focused on data and service protection but can and should be expanded to include all enterprise assets (devices, infrastructure components, applications, endpoint security, virtual and cloud components) and subjects.

Subjects are end users, applications and other nonhuman entities (such as enterprise systems of manufacturers, wholesalers and dispensers) that request information from resources (e.g. the PI verify service).

NIST defines a zero trust architecture (ZTA) as an enterprise cybersecurity architecture that is based on zero trust principles and designed to prevent data breaches.

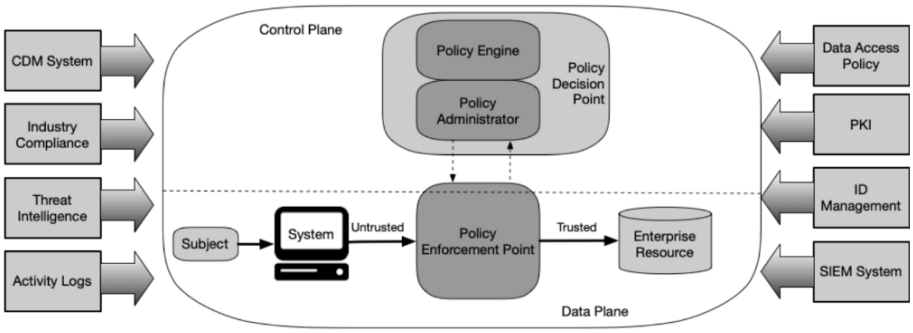

Figure 2: Core Zero Trust Logical Components (Source NIST)

When there is data exchange among supply chain actors, **zero trust principles** need to be adopted beyond a single enterprise. The principles need to be adopted for the entire supply chain ecosystem in order to enable **secure authentication and authorisation** as well as to prevent access and data breaches across supply chain actors. Hence, adoption requires a strong ecosystem innovation and deployment approach to successfully proliferate the standards and technology solutions to an early ecosystem majority.

2.3 Variations of Zero Trust Architecture Approaches

There are several ways that an enterprise ecosystem can enact a ZTA for their data exchange workflows. These approaches vary in the components used and in the main source of policy rules for an organization.

These **different ZTA approaches** include:

1. ZTA Using Enhanced Identity Governance,
2. ZTA Using Micro-Segmentation,
3. and ZTA Using Network Infrastructure and Software Defined Perimeters.

Typical supply chain ecosystems apply single enterprise ZTA best practices and combine these with a ZTA design approach for (1) Enhanced Identity Governance and (3) Network Infrastructure and Software Defined Perimeters for cross-enterprise authorisation.

(1) Enhanced Identity Governance

The enhanced identity governance approach to developing a ZTA uses the identity of actors as the key component of access policy creation. If it were not for subjects requesting access to enterprise resources, there would be no need to create access polices. It shall be understood that policies are applicable for a given context.

This means that APIs must make sure the context of an API request is defined so that the relevant policy can be identified and applied.

For this approach, **enterprise resource access policies are based on identity and assigned attributes**. The primary requirement for resource access is based on the access privileges granted to the given subject. Decentralized systems can express access policy attributes in the form of W3C verifiable credentials.

To provide verifiable credentials a use case domain establishes a Trust Domain and a **Verifiable Credential Issuer** (VCI) role. The VCI verifies the enterprise identity and authorisation status of a supply chain actor such as suppliers, manufacturers, wholesalers or retailers and issues verifiable identity and authorisation credentials for supply chain actors.

These credentials can be blended with each data request or response message so that any supply chain actor can check the identity and the authorisation status of an incoming request of a given counterparty. In case the identity or authorisation status of a supply chain actor changes, the VCI revokes the respective credential and eventually issues a new credential. Therefore, authorization is behavioral.

When a supply chain actor receives a data request (or request response) the ERP system of the supply chain actor interacts with a credential **verification module** (e.g. module of the identity wallet) that checks the validity of an authorisation credential. In case the authorisation credential is valid, the access management capability of the ERP processes the request and returns a response. In case the authorisation credential is not valid, access is permitted.

Note: It shall be understood that such an approach requires a challenge response protocol to exchange so-called verifiable presentations. For the sake of simplicity the elaboration of verifiable presentation flows is not in scope of this paper.

The credential verification module of the identity wallet is the core feature to establish a *policy engine* for the control plane in accordance with ZTA principles.

Verification of an authorisation credential of incoming request messages could be **done every-time for every incoming message**. With this check a credential-based ZTA implementation verifies the authenticity and integrity of incoming request messages every time. When the request message body hash is included into the verifiable presentation of authorisation credential, a verification module of the API call handler can check both authenticity and integrity of the incoming message and the authorisation status of the message requestor.

Within an open supply chain network API service endpoints are exposed to any bothe type of actors, honest and malicious actors. API endpoint access is initially granted to all supply chain actors and potentially to malicious actors but access to the API endpoint business logic is restricted to requestors with the appropriate access privileges (e.g. authorisation credential).

An ERP system uses the identity of requesters (or responder) to form and enforce policy for processing PI verify requests (or response) on its

solution platform. The identity-driven approach works well for supply chains industry wide policy definitions for authorised eligible parties since counterparty system identity and authorisation status provide support data to access decisions.

There is a downside in granting basic network connectivity as malicious actors could still attempt network reconnaissance and/or use the network to launch denial of service attacks either internally or against a third party. Thread modeling for these attack vectors need to be conducted.

(2) Micro-Segmentation

An enterprise ecosystem may choose to implement a ZTA based on placing individual or groups of resources on a unique network segment protected by a gateway security component. In this approach, the enterprise places infrastructure devices such as intelligent switches (or routers) or next generation firewalls (NGFWs) or special purpose gateway devices to act as policy enforcement point protecting each resource or small group of related resources. Alternatively (or additionally), the enterprise may choose to implement host-based micro-segmentation using software agents or firewalls on the endpoint asset(s). These gateway devices dynamically grant access to individual requests from a client, asset or service.

The micro-segmentation approach is NOT applicable for open supply chain ecosystem infrastructure as it needs to be implemented on a lower OSI layer. Such an implementation across multiple supply chain actors would be very difficult to deploy and maintain, because of inhomogeneous cloud and network infrastructures and dynamic development of micro-segmentation attributes. Therefore we recommend abstracting from underlying low-level networking devices and segmentation settings..

(3) Network Infrastructure and Software Defined Perimeters

The last approach uses the network infrastructure to implement a ZTA. The ZTA implementation could be achieved by using an overlay network (i.e., OSI layer 7 Application). This approach is sometimes referred to as software defined perimeter (SDP) approach. In this approach, the Policy Administrator (PA) acts as the network controller that sets up and reconfigures the network based on the decisions made by the Policy Engine (PE). The requester system continues to request access via Policy Enforcement Points (PEP), which are managed by the Policy Administrator component.

When this approach is implemented at the application network layer (i.e., OSI layer 7), the most common deployment model is the agent/gateway:

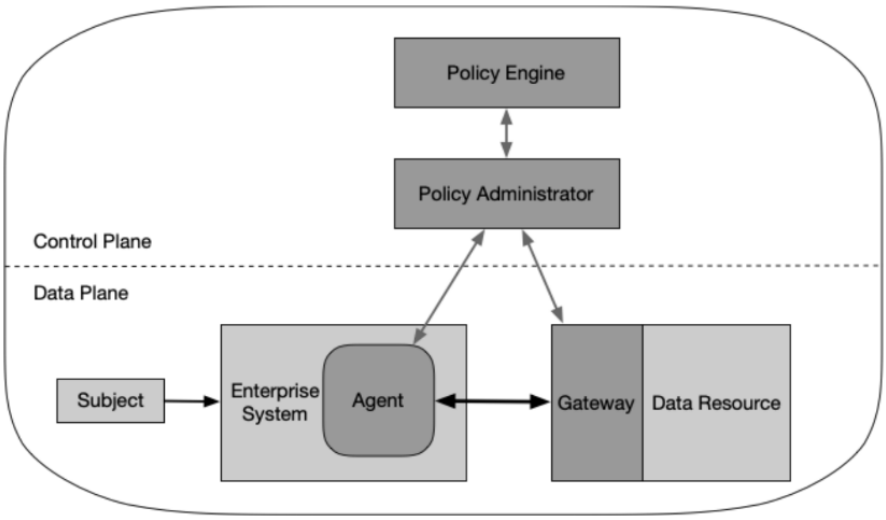

Figure 3: Agent/Gateway Model in a single Enterprise

In this implementation, the agent and resource gateway (acting as the single PEP and configured by the PA) establish a secure channel used for communication between the client and resource.

We propose the idea of authorized identity subjects, authorizations are conveyed via signed statements in the form of verifiable credentials for a cross-enterprise supply chain scenario. These statements must not merely be signed in motion (verifiable presentations) but must be signed at rest as well (verifiable credentials). This means that no intervening infrastructure can tamper with them, neither at rest nor in motion.

The Agent/Gateway Model is directly not applicable for inter-organizational collaboration scenarios as there are no shared PE and PA capabilities in open cross-enterprise systems. The control plane is distributed. Therefore we propose to extend the ZTA principles for cross-enterprise usage by establishing dedicated PA, PE and PEP resources within both enterprise trust domains (as shown in the chart below) as well as diffuse trust perimeter-less security principles (as described further below).

3. ZTA Implementation Approach for Cross-Enterprise Authorization

The PEP includes both capabilities, agent and gateway, depending on the direction of the communication flow, request or response. The PEP agent interacts with a PA feature in the VRS software to configure communication channels with authorisation credentials, with a PE for evaluating authorisation status and with a (pre-configured) gateway feature to connect to the respective enterprise data resources.

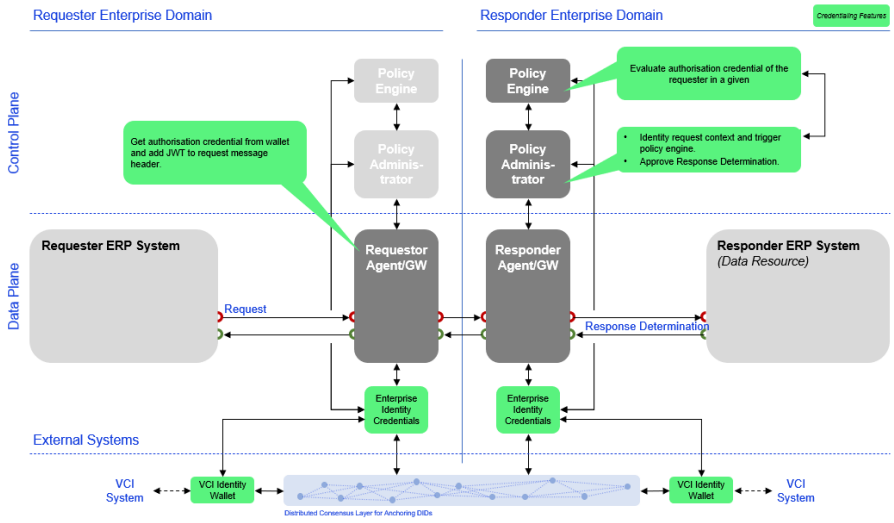

Figure 4: Cross-enterprise model, with distributed control plane

In this scenario, a requester with a requester ERP system wishes to connect to the ERP system of a service of the responder in accordance to the following approach:

- The request is taken by the requester agent, and the request is forwarded to the agent of the responder together with a context definition.
- The request includes an authorization credential in the header of the request (e.g. in JWT format).
- The policy administrator of the responder receives the request via its agent and triggers the respective policy engine evaluation for the given context.
- In case the policy engine of the responder returns a valid evaluation result for the authorisation credential the policy administrator configures a communication with the responder ERP systems for the given context and a response message is being created and send to the requester
- It shall be understood that that responder might be obliged to add an authorization credential to its response so that the requester can determine if the response is coming from an authorized responder.
- The connection between requester agent and responder agent is terminated when the workflow is completed or when triggered by the policy administrator due to a security event (e.g., session time-out, failure to authenticate, failure to verify authorisation credentials).

As outlined under *(1) Enhanced Identity Governance*, the ERP/agent uses the identity of a given requester (or responder) to form and enforce policy for processing requests (or responses) among ERP systems.

4. Industry-wide Consensus

In order to implement such a credential-enabled ZTA architecture successfully a use case domain must refine the NIST zero trust networking principles and leverage an industry-wide consensus to increase the security of the (distributed) control plane which provides the policies for authentication and authorization. Basically authorization is the control place.

This consensus includes industry-wide endorsement, acceptance and implementation of conformance criteria for

- "Root of Trust" mechanism and authorisation credential chaining
- Industry compliance (e.g. integration with industry specific messaging protocols)
- Technology conformance criteria for
  - Identity and verifiable credential standards
  - Wallet infrastructure and security
  - Identity and authorisation verification and credential issuance
  - Verifiable presentation creation and verifiable presentation verification for access policy enforcement
- Verifiable Data Registry (VDR) for storing information about a cryptographic identifier (i.e., signing keys and service end-points stored W3C DID documents).
- Semantic schemas
- Revocation mechanisms

Using such a consensus allows an industry ecosystem to use the idea of what is called threshold structures in security.

Where a threshold of components acting in concert provides more security than any of the components acting singly could do. Threshold structures are applied in multi-signature signing schemes, multi-factor authentications, multiple credentials for attribute based access management (ABAC), and distributed consensus algorithms.

Distributed consensus algorithms are used to protect the DID documents of the supply chain actors by storing DID documents on a blockchain (e.g. public Ethereum blockchain).

5. Credentialing-enabled ZTA Principles and ZT Computing

Together with Sam Smith augmented principles were developed in order to support a generic ZT Computing across multiple ecosystem actors (Include Reference to "diffuse trust perimeter-less security principles")

#	Principle	Description
P1	Network Hostility	The network is always hostile, internally & externally; Locality is not trustworthy. Solutions must provide means to mitigate network layer security vulnerabilities (man-in-the-middle, DNS hijacking, BGP attacks).
P2	E2E Security	Inter-host communication must be end-to-end signed/encrypted and data must be stored signed/encrypted. Data is signed/encrypted in motion and at rest.
P3	E2E Provenance	Data flow transformations must be end-to-end provenanced using verifiable data items (verifiable data chains or VCs). Every change shall be provenanced.
P4	Verify every-time for every-thing	Every network interaction or data flow must be authenticated and authorized using best practice cryptography.
P5	Authorization is behavioral and context-dependent	Policies for authentication and authorization must be dynamically modified based on behavior (reputation).

#	Principle	Description
P6	No single point of trust	Policies for authentication and authorization must be governed by end-verified diffuse-trust distributed consensus. Policy is protected by diffuse trust.
P7	Agents locked down & Agent/Wallet Security Feature Detection	Agents or agent components executing any of the logic mentioned above must be locked down. Any changes to their execution logic or behavior must be fully security tested and validated on the respective possible combinations of hardware and software platform. Special security measures must be implemented regarding secure execution of the logic (e.g. code injection, insecure object references, cross-site/service request forgery, cross-service scripting, secure key management). Security features of agent and wallets shall be detectable by a counterparty.

6. Outlook

There are multiple industry and R&D projects that are evaluating options to implement a credential-based ZTA approach and attribute-based access control mechanisms that are linked with an identity and trust framework.

One of these projects is the Gaia-X sovereign cloud infrastructure project of the European Union.

We expect that these projects evaluate and test different architecture options while developing a complte ecosystem to roll out the technology for large scale field testing within the next 1-2 years.