CCDCOE
NATO COOPERATIVE
CYBER DEFENCE
CENTRE OF EXCELLENCE

No 13 / November 2021

Recent Cyber Events:

Considerations for Military and National Security Decision Makers

Zero Trust:

- → What is it?
- → Best Practice
- → Architecture and Technology
- → Legal Aspects

Published by CCDCOE

What is the Zero Trust security model?

Today, when companies and organisations set out to modernise and improve their cybersecurity posture, chances are that they will base their approach on the Zero Trust security model —and for good reasons.

Computer security professionals love to say that there is no such thing as 100% security. Despite that, many security strategies seem to be founded on the assumption that setting up border protection and establishing a secure perimeter around the enterprise network can be done well enough. Zero Trust does not make that assumption. Zero Trust is a security model built around the idea that no user or device should be trusted just because it is operating in a 'private' network.

Although it may not have been the first use of the term Zero Trust, the introduction of the model by John Kindervag at the research company Forrester in 2010 is often considered a starting point of the current trend.[1] Since then, commercial support for the model has grown and it is gaining in popularity. Many major corporations and organisations are now using it, with Google being one of the early adopters. In the last year, government security agencies have been pushing the model more and more. There has, however, been a lot of market hype connected with the model, and different vendors will claim that their particular offering is the true path for implementing Zero Trust.

It is therefore important to be familiar with the Zero Trust model and to understand its fundamentals because the model is based on sound principles that should always have been applied and may now become mandated, and also because it is important to be able to see through the hype.

Even the name Zero Trust may lead people astray. Zero Trust does not mean that you do not have to trust something, or that you should not trust anything. It is simply the concepts that follow from that basic principle of not implicitly trusting entities on the network, meaning Zero *explicit* trust. In this, the Zero Trust model is closely related to the concepts of assumed breach and defence-in-depth.[2]

The basic components of the Zero Trust model are described in different ways, but it is often presented with these three underlying principles:[3]

1. Don't trust – verify explicitly
2. Use least privileged access
3. Assume breach

The first is the principle that has given the model its name. The identity and rights of the requestor should always be verified with every access request, with no distinction between requests from the internet or the 'internal' network. Authentication is thus needed not just for accessing the network, but for every session and access to data or other protected resources.

The second principle states that access should not automatically be granted to every user in the organisation that can be authenticated. Access decisions should be taken based on the need-to-know. It should be granted only for the bare minimum of information needed to perform the work duties.

The final principle is the underlying reason for the first. Since complete security cannot constantly be achieved, we must assume that breaches will happen or that they have already happened. This requires constant vigilance and a watchful stance with measures taken to detect and mitigate breaches.

These concepts are in no way new. They have been applied separately and in combinations for a long time. What the Zero Trust model does is to bring them to the fore and build a coherent security paradigm around them. The model is probably at its best when viewed as a set of guiding principles to be followed as closely as reasonably possible when evolving the cybersecurity of an organisation rather than a fixed formula or an end state to be achieved.

Some of the principles of the model may seem contrary to how military networks and classified systems have traditionally been protected. Highly classified systems and networks have traditionally been air-gapped whenever possible. When transmission over wide area networks, radio links or other long-distance connections has been needed, high-assurance cryptographic solutions have been adopted. These measures give a high degree of confidence that an adversary will not be able to breach the system remotely and creates what may be seen as a 'trusted' network.

Zero Trust teaches us that we should always assume a breach and verify every access even from within the internal network, even with air-gapped networks. There is always an insider threat and the risk of physical intrusion or failure of security mechanisms and the same Zero Trust principles should be applied for these systems. The outer shell may be more secure, but the damage, if it is breached, will be much greater.

Applying these principles also means using end-to-end encryption to protect information, even on the internal network. There should not be any implicit trust in the network or its users just because it is internal.

The Zero Trust principle of lowest privileged access builds on the need-to-know principle that has traditionally been

1 John Kindervag, Stephanie Balaouras and Lindsey Coit, "No More Chewy Centers: Introducing The Zero Trust Model of Information Security", Forrester, 2010.

2 Defence-in-depth refers to the principle of not trusting just one layer of defence or protection but adding more layers that can stop an intruder if the outer defences are breached.

3 See for example Microsoft or Focus-IT.

strong in military and national security. However, need-to-know has been complemented if not supplanted by the responsibility-to-share doctrine. This does not mean that all information needs to be shared, but it does recognise the importance of sharing the information that others may need for the tasks, and that it may be difficult to know in advance what information will be important to whom.

The principle of least privilege tells us to balance the need-to-know with the responsibility-to-share, making it possible to share all information that is needed by others while limiting it to the information that may be needed and revoking access rights once there is no longer a need. The dynamic and granular access control and continuous monitoring of the Zero Trust model will support this.

Zero Trust guidance and best practices

The persistence of large-scale cyber incidents has yielded a series of guidelines from the government and the private sector on how to implement Zero Trust models to replace older approaches based on perimeter protection. The primary challenge for such guidelines is that the concept is still evolving, and there is no out-of-the-box solution that fits every situation. Rather, Zero Trust implementation requires organisation-specific system integration.

The value of Zero Trust can be illustrated by four, common challenges:[4] credential theft, remote exploitation, insider threat and compromised supply chain. Countering credential theft can be achieved by ensuring that access is not possible by just presenting fixed information like username and password. In Zero Trust, other factors such as device identity are also considered when granting access. Guidance for implementing the Zero Trust model also usually prescribe multi-factor authentication (MFA), making credential theft more difficult.

If a malicious actor has already gained access to a system, Zero Trust decrees implementing network segmentation and other measures to prevent lateral movement, essentially building defence-in-depth and limiting the possibility of remote exploitation. In a Zero Trust Architecture, there is no implicit trust in any hardware or software components in the internal network. By denying any compromised device or application connections to remote addresses for command and control by default, a Zero Trust compliant system can mitigate a supply chain attack.

One of the core issues in introducing Zero Trust Architecture is that there is no standard design that will ensure the desired outcome. This puts a lot of responsibility on the policymakers and network administrators. Implementation may take years, and procurement of software, hardware and training of personnel all come with costs for the organisation.

The core mechanism of Zero Trust is ensuring valid access. This is achieved by a centralised management system known as brokered access, enforcing a policy for access. The mechanisms implementing this policy continually ensure the validity of access granted by assessing the confidence in the request based on factors such as type of authentication, time of access, the status of the device used and its location. Essential for the correct function of these brokers is the use of reliable protocols and algorithms, and that it is properly implemented and configured. Configuration mistakes may easily defeat the purpose of the policy. To mitigate the risk of misconfiguration, the US National Institute of Standards and Technology (NIST) has stipulated ten network requirements to assist organisations transitioning to Zero Trust in its Special Publication 800-207. The challenge in implementing such mechanisms is that they must be available at all times as they are the sole route for access to protected resources. This means that should these mechanisms be targeted by a Denial of Service attack, it may disrupt the entire network.

With all these challenges and costs, the perspective of introducing Zero Trust models may seem daunting. To help, several government institutions have issued guidelines.[5] The common denominator amongst these is to ensure that the security protocol covers all aspects of communication including the user and any device, regardless of location.

Zero Trust is the new security paradigm,[6] and organisations need to adapt. However, this will take time and significant changes to infrastructures and workflow. To meet this challenge, organisations must conduct risk assessments and prioritise systems for security upgrades.

Zero Trust Architecture and technology

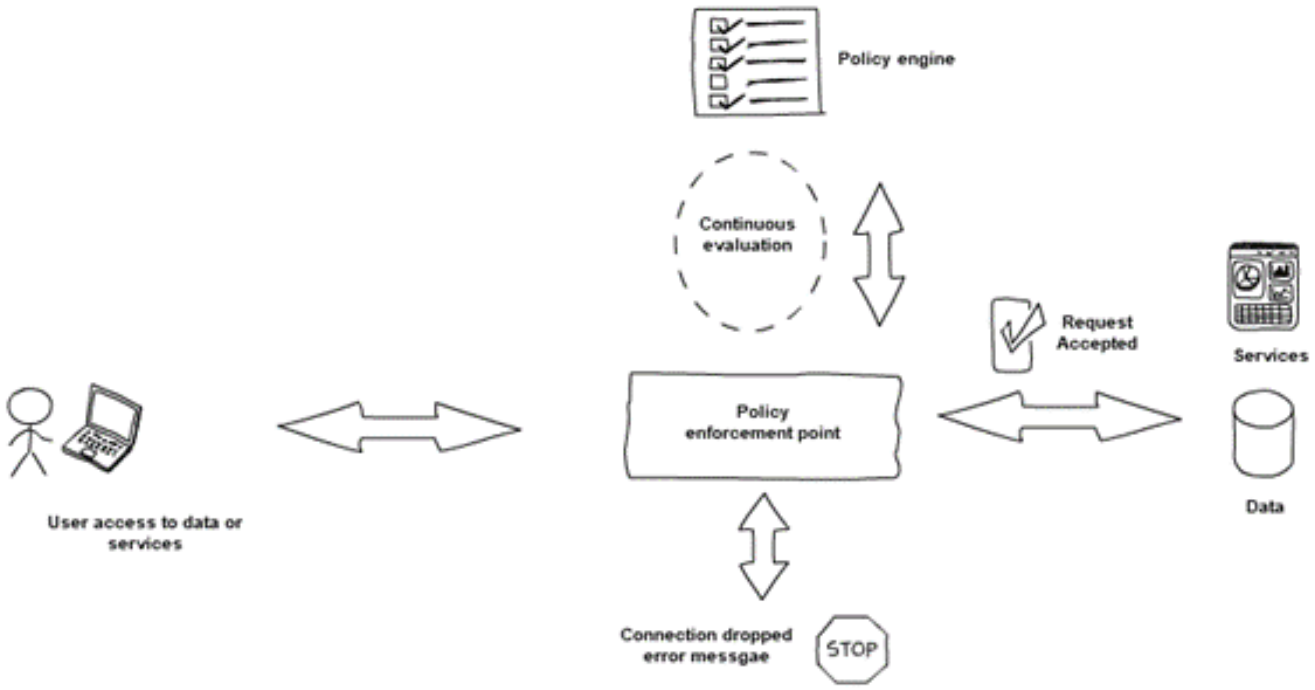
Implementing Zero Trust requires fundamental changes to a system's infrastructure to create what is known as Zero Trust Architecture (ZTA). This cannot be achieved with a single security appliance or software package, but rather using mechanisms implemented throughout the architecture such as:

1. Secure authentication
2. Policy management
3. Event monitoring
4. Encryption

4 National Security Agency: Embracing a Zero Trust Security Model

5 For example from the National Institute of Standards and Technology (NIST) and the National Security Agency (NSA) in the US and the National Cyber Security Centre (NCSC) in the UK.

6 According to Microsoft's survey carried out in US, Germany, Japan, and Australia/New Zealand with 900 participants, 76% of organisations have at least started implementing Zero Trust.



Overview of the principal parts of a Zero Trust Architecture. Source: NCSC

Since the model is based on strict identity verification for each user and device, a core function of Zero Trust needs to be strong authentication, usually MFA. An enterprise Public Key Infrastructure (PKI) will typically be an essential part of this.

As the name suggests, the central concept in Zero Trust is to always verify access to protected resources, hence the brokered access mechanism described in the previous section. This mechanism consists of a centralised Policy Engine, or Policy Decision Point (PDP) and one or more Policy Enforcement Points (PEP).

ZTA is based on all access to protected resources going through the PEP, with the Policy Engine considering not only the identity of the user or device but also things like the method of authentication, device location and status, and even threat intelligence. Collecting and assessing this information builds confidence, thereby creating a basis for allowing access. The higher the value of the resource, the higher the confidence required to gain access.

There are several architectural options for implementing the Policy Engine and PEP. The US NIST's Special Publication 800-207 presents several ZTA variations and implementation scenarios.

Another essential part of a ZTA is continuous monitoring, logging and analysis. This part of the architecture will typically use several products for collecting, storing and analysing signals from the network, services and applications. Essential components are sensors, activity logs, and a security information and event management system (SIEM).

Finally, Zero Trust requires the protection of data in motion through encryption. Traditional VPNs can be part of the solution but since trust should not rely on just presence on a network, not even a virtual one, protection is often based on communicating directly with services using secure protocols. The protection must also allow the inspection of encrypted traffic for continuous monitoring, placing additional requirements on the architecture.

A challenge in implementing a ZTA is that in most cases components from several vendors are needed and all must communicate with the Policy Engine, while at the same time there are no given open standards for the interfaces between the different components. Vendors will often use proprietary API forcing customers and other vendors to develop solutions for several APIs and to keep them updated as the APIs evolve. Products will have to be selected carefully for compatibility, something made difficult by the lack of standards to base interoperability requirements on. Planning implementation across more than one year is even more of a challenge given the evolving market and technology. Efforts are underway to standardise protocols that will be useful in implementing ZTA, giving some hope for an improved situation.[7]

7 See for example Special Publication 800-207, page 48 for more information.

Legal aspects of Zero Trust

When considering legal mandates to adopt the Zero Trust model, the first aspect to consider is the requirement for a correct legal basis. Usually, the domestic legal framework offers answers to mandate-related questions, as laws and regulations provide the necessary framework and guidance. Every state has laws and regulations that regulate information security issues, communications and cyber incident handling procedures.[8] Security regulations are nationally regulated, therefore the mandate for implementation must be sought domestically. Organisations must set internal rules and regulations for implementing and enforcing legal requirements.

The Biden administration broke new ground by explicitly mandating the adoption of the Zero Trust security model in its executive order 14028, 'Improving the Nation's Cybersecurity'. The Executive Order sets a timeframe for the Federal Government and several of its agencies to modernise and implement stronger cybersecurity standards by advancing towards ZTA. Agencies have been ordered to develop plans and strategies to enhance and improve communication, information sharing and compliance frameworks. A draft federal strategy and a maturity model have already been released for public feedback.

The Executive Order is the next step to advancing US national cybersecurity and it prescribes the actions to be taken to achieve that enhancement to make systems stronger and more resilient. It includes a requirement to report progress against the set timeframe. It also urges cooperation between sectors in implementing Zero Trust and is setting high ambitions towards secure cloud services.

The Directive on the security of network and information systems (the NIS Directive) provides legal measures and defines sectors within the EU to enhance cybersecurity and make the Union more resilient to cyber threats. Although it does not specifically prescribe Zero Trust, it requires member states to conduct risk assessments and ensure security in vital sectors, as detailed in the implementation regulation. Using Zero Trust as a mechanism, while not mandated by NIS or the proposed updated NIS2 Directive, is certainly not precluded and may help to maintain security levels and mitigate risks and thus facilitate compliance with NIS.

The EU is also improving its approach to secure cloud services with a forthcoming EU Cloud Rulebook and an EU-wide Cloud cybersecurity certification scheme.[9]

Legal requirements should be clear and specific enough to provide a good understanding of the requirements that are necessary for implementation and effective enforcement. Legal clarity also offers possible consequences when standards and conditions are not met. However, a more general prescription to apply appropriate measures will allow freedom of interpretation and adjustment to each specific case.

Another aspect to consider in evaluating legal regulations is the technology-neutral language that should be used in legislation.[10] If norms are written without any specific system or security feature in mind, they will be more accommodating of emerging technologies.

The answer to the question of clarity vs ambiguity, will likely be a balancing act. Referencing the Zero Trust model as a conceptual framework may be useful in striking that balance in rules and regulations. The principles of Zero Trust are likely to be more long-lived than the specific technology used to implement it today.

CONTRIBUTORS

Sungbaek Cho
Kārlis Podiņš
Damjan Štrucl
Urmet Tomp
Grete Toompere
Ingrid Winther
Jan Wünsche

PREVIOUS ISSUES

This paper is part of a series of monthly reports. This issue and all previous issues are available in the CCDCOE online library.

FEEDBACK

To continuously improve this regular report, input from readers is essential. CCDCOE encourages feedback on both how the reports are of use to you and how you think they can be made better.

Please send your comments and suggestions to feedback@ccdcoe.org

8 For example compendium of US regulations: CISA's Resources for Lawyers

9 For further reading: European Commission: Cloud and Edge Computing: a different way of using IT

10 One good example of neutral language use is the Budapest Convention on cybercrime that has been in effect and without changes for almost 20 years with 66 parties to the convention.

ABOUT THIS PAPER

This recurring report is the collaborative view of NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) researchers highlighting the potential effects of current events and developments in cyberspace on armed forces, national security and critical infrastructure, based on publicly available information. It does not set out to be exhaustive. While the authors have made every effort to describe events from a perspective relevant to NATO and partner nations, there may be national and regional differences which this paper does not address.

The authors of this paper are independent researchers at the NATO CCDCOE; they do not represent NATO, nor does this paper reflect NATO's position. The aim of the paper is not to replace information about vulnerabilities and incidents provided by CSIRTs and providers of CIS products and services.