Zero Trust: The Magic Bullet or Devil's Advocate?

Helvi Salminen
Thales DIS Finland Oy, Vantaa, Finland
helviksalminen@gmail.com

Abstract: The concept of Zero trust was first introduced in mid 1990's, and has gradually attracted increasing attention. This approach to building organizations' information system infrastructures has been developed as response to increasing interaction and interconnection of information systems. Along with organizational boundaries have become less clear with the new business models where a business process exceeds the organizational boundaries, also the boundaries of information systems are no longer clear. In this interconnected world the purely perimeter-based security model defining zones of trusted entities inside the perimeter and the untrusted external world outside the perimeter no longer serves the needs of new business models. And the combination of complex technology and sophisticated attack methods it is no longer possible to be sure that all system components and actors inside the perimeter can be trusted. The Zero trust approach brings the sophisticated controls from the perimeter to the entire system. The core idea can be expressed with the four words "never trust, always verify". No system component is by default trusted , and one-time verification is not sufficient – access to a resource must be verified at each connection attempt. Mutual authentication of the communicating parties is in the core of the approach. But does the zero trust approach have unwanted side-effects? The complexity of the system increases when new control layers are built, and system complexity can increase the possibility of configuration errors. Can there be other side-effects as well? The need for trust does not disappear even when the systems are built on the zero trust principles. When studying the zero trust approach the author started thinking what would happen in human interaction and organizational co-operation if they are based on or partly apply the zero trust approach. And the scenarios were quite gloomy. But is this only a nightmare or already at least partly present in our reality? This article describes the zero trust approach and its applicability to technical environments. The second part present scenarios of the impacts which application of zero trust principles could have – or maybe already has - in human communication and organizational relationships.

Keywords: Zero Trust, Perimeter-Based Security, Organizational Culture

1. Zero trust architecture

1.1 Introduction

The objective of the zero trust approach is to protect an organization's valuable resources which are data, systems and applications. Already the name expresses what is the core idea often expressed as *never trust, always verify*. No subject trying to access a resource is by default trusted, and the access is granted with the least privilege per request. Netskope (2021) defines the primary goal of the zero trust approach to be shifting from *trust to verify* to *verify then trust*.

Rose et al (2020) describe in the NIST special publication (later referred as NIST) an abstract definition of zero trust architecture (ZTA). The following picture presents an abstract model of access. A subject requests access to a resource. The policy decision point (PDP) makes the access decision which is enforced in the policy enforcement point (PEP).

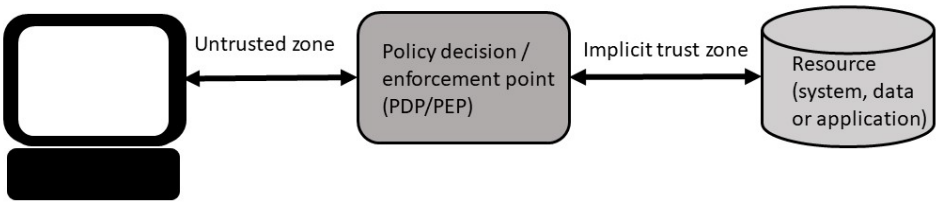

Figure 1: Zero trust access in NIST model

The subjects requesting access are in the untrusted zone, When making the decision the PDP/PEP engine uses a complex set of criteria to evaluate if the requester can be trusted and given access to the implicit trust zone where all entities are considered trusted . The size of the implicit trust zone should be restricted to minimum.

The basic seven tenets of zero trust according to NIST are:

1. All data sources and computing services are considered resources.
2. All communication is secured regardless of network location.

3. Access to individual enterprise resources is granted on per-session basis.
4. Access to resources is determined by dynamic policy—including the observable state of client identity, application/service, and the requesting asset—and may include other behavioural and environmental attributes.
5. The enterprise monitors and measures the integrity and security posture of all owned and associated assets.
6. All resource authentication and authorization are dynamic and strictly enforced before access is allowed.
7. The enterprise collects as much information as possible about the current state of assets, network infrastructure and communications and uses it to improve its security posture.

These tenets define an target situation, and all of them may not be fully implemented in a given zero trust solution. The core components of the zero trust architecture and the trust decision-making mechanisms are described in more detail in section 1.4.

Fernandez (2023) lists a set of fundamental principles that should be applied when building secure systems. These principles which were proposed by Saltzer and Schroeder (1975) include

- *Closed system* – anything not explicitly allowed is forbidden.
- *Least privilege (need to know)* – access is granted only to resources required to perform a task.
- *Complete mediation* – every access request must be validated.
- *Minimal trust* – the parts of a system which must be trusted should be minimal.
- *Holistic security* – all architectural layers must be secure, and security applied during development.
- *Defence in depth* – more than one line of defence is needed.
- *Submarine principle* – some units can be lost, but essential units must be preserved.
- *Compromise recording* – in some case reliable recording of occurred compromise can be used instead of mechanisms which prevent loss.

Many of these principles are a part of zero trust solutions, the exact set is solution dependent. According to Fernandez (2023) all ZTA solutions seem to include closed systems, least privilege and complete mediation.

1.2 A brief history of zero trust

Like many other architectural concepts, zero trust is not a single invention. Instead, today's zero trust is a result of several innovations built on preceding ideas. The following picture presents a timeline.

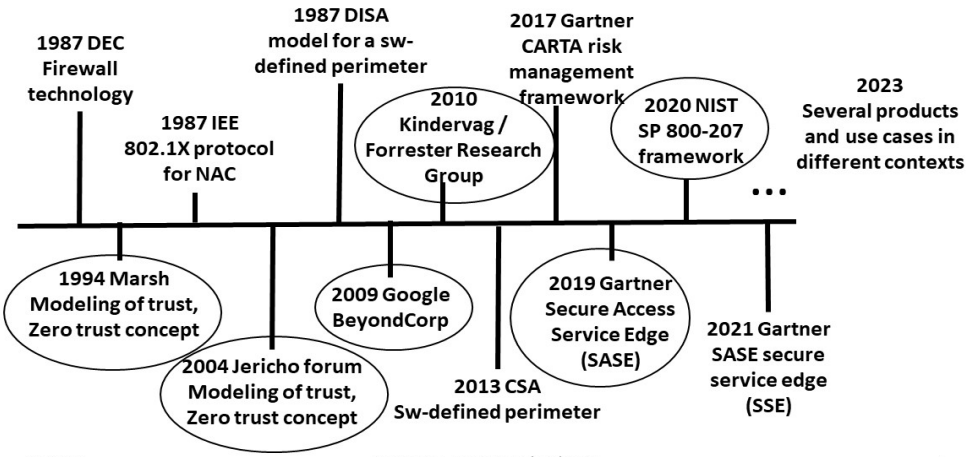

Figure 2: Zero trust timeline – information from several sources

The milestones which are most frequently mentioned in the publicly available zero trust articles and presentations are the following:

- 1987 – Engineers from the Digital Equipment Corporation (DEC) publish the first paper on firewall technology
- 1994 - Marsh (1994) is mentioned in many sources as the first publication to use the term *zero trust*. Marsh's work extends the concept from relationships of humans to artificial agents.
- 2004 - Jericho forum was founded. The mission of the forum was to define and promote de-perimeterization. Core ideas of the forum are presented in the Jericho Forum commandments. The conclusion of the commandments is that the de-perimeterization is inevitable and an organization needs to plan for it and should build a roadmap how to manage the development.
- 2009 – Google BeyondCorp introduces by Ward and Beyer (2009).
- 2010 – Analyst John Kindervag uses the term *zero trust* in a Forrester Research Group paper
- 2019 – Gartner Secure Access Service Edge (SASE) is introduced.
- 2020 – NIST SP 800-207 framework is published.

1.3 Zero trust vs. other architectures

The zero trust approach is often compared with perimeter-based architecture model. The focus of zero trust is in protecting the resources– assets, services, workflows, accounts etc. – regardless of their location. Perimeter-based architecture, instead, protects network segments separated by perimeters with security mechanisms in place. The following picture illustrates the main difference of the two approaches.

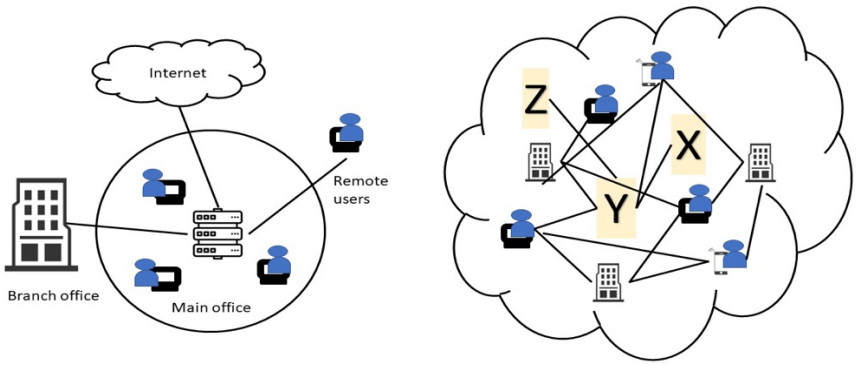

Figure 3: Perimeter-based vs. zero trust architecture

This division is simplified, as also in the perimeter-protected network segments many of the security mechanisms listed in Saltzer and Schroeder (1975) are used.

1.4 Core components of zero trust architecture

The NIST model defines core components which are needed in a zero trust implementation. The following picture presents the components and their roles in the access to a resource.

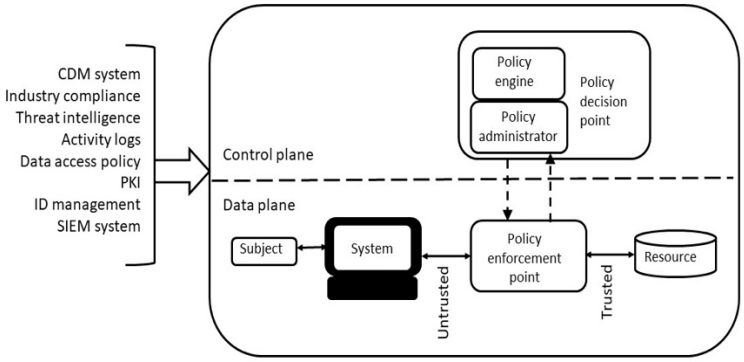

Figure 4: Core components of zero trust architecture in NIST model

- Policy engine (PE) makes the access decision. In the decision it uses the information from the external sources as input to the trust algorithm which makes the decision based on pre-established criteria.
- Policy administrator (PA) allows or denies access based on PE's decision.
- Policy enforcement point (PEP) enables, monitors and in the end terminates the subject's access to the resource. It stands between the untrusted and trusted zones.

Trust algorithm used by the PE is the heart or brain of the architecture. It uses several sources of information when evaluating if the subject can be trusted and given access to the requested resource.

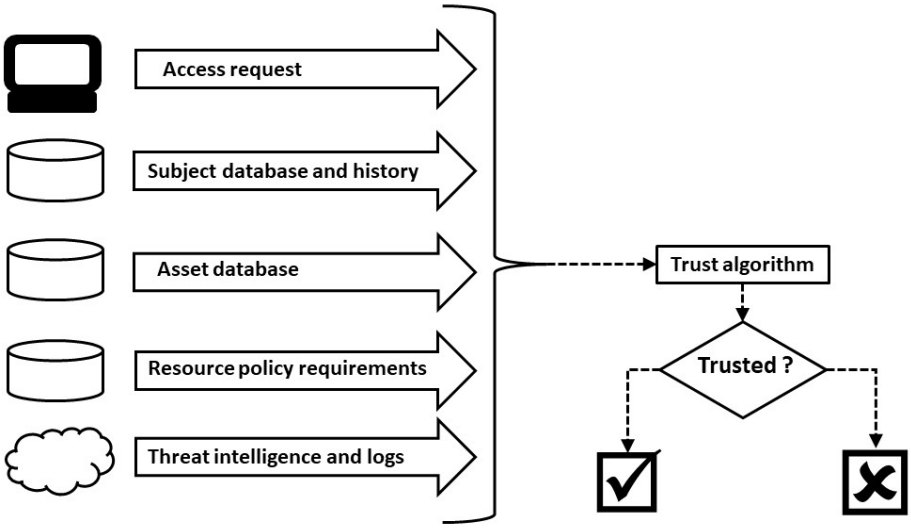

Figure 5: Trust algorithm input in NIST model

There can be different variations of the trust algorithm (TA).

- Criteria-based TA requires all predefined criteria to be met before granting access.
- Score-based TA accepts request which have the trust level above the defined threshold.
- Singular TA evaluates each access request individually and does not use the subject history in the evaluation.
- Contextual TA uses also the subject recent history when evaluating the request.

1.5 Benefits of zero trust approach

Zero trust is not a specific architecture, but a strategy or philosophy to build secure information systems. The main driving force is the change in business models. Organizations concentrate increasingly in their core functions and producing a service requires involvement of several organizations. So, a business process extends beyond the organizational boundaries.

These complex value chains and increasing use of cloud services make it challenging, in many cases impossible, to use purely perimeter-based architectures. Zero trust approach enables protection of the resources regardless of their location. However, physical and network perimeters have not lost their importance in all contexts, which is discussed in section 1.8.

1.6 Potential problems in zero trust approach

As described above, the zero trust approach to information systems architecture embeds security to the complex distributed environment where perimeters no longer can guarantee that all connecting entities can be trusted. However, there are some threats which require attention.
Fernandez (2023) lists potential weaknesses in the ZTA approach. The list includes the following concerns:

- The decision process can be subverted and unauthorized access enabled by attacking the PE or PA.
- Fake component can be introduced to the infrastructure as result of successful attack against the Certification Authority. Fake PE or PA would give the attacker full power to make access decisions.

- Denial of service attacks against the interfaces which handle access requests can block authorized users' access to the system.
- Manipulation or coercion of automated functions, e.g. trust calculations, can cause the system disruption or unauthorized access.

Fernandez et al (2023) also emphasize the importance of governance procedures. This includes the definition of roles, assignment of rights to roles and role and rights management. ZT approach has been criticized for not taking into account the existing risk management governance frameworks or applicable regulations.

In a security architecture the risks of overkill, incompleteness and excess overhead need attention. When evaluating the applicability of ZTA there are some important questions to be asked. Fernandez et al (2023) ask four questions and provide also preliminary answers.

Q1. Does ZTA apply many unnecessary checks? All data items or resources do not need the same level of protection. Equal protection regardless of the sensitivity may be an overkill.

Q2. Does ZTA introduce an intolerable overhead? Dynamic access control and trust evaluation can cause significant overhead, especially if trust must be evaluated separately for each access request.

Q3. Is ZTA implementable for most institutions? ZTA is feasible, but significant support from vendors may be needed in ZTA implementations. Well-planned gradual implementations could be the best strategy, but it must be noted that intermediate stages may not meet the security target of the organization.

Q4. Can ZTA control more, fewer, or similar number of threats as compared with other approaches? ZTA concentrates on the attack surface of the system and makes it more difficult for the attacker to achieve the target. However, ZTA methodology does not ensure that all system levels are covered by the security mechanisms. In this sense ZTA shares the weaknesses of perimeter-based controls.

Townley (2023) points out a misunderstanding about the interpretation what zero trust means – Kindervag's original premise is *trust no packet* instead of *trust no one* as often understood. He also emphasizes the importance of integrating the ZT objectives to the enterprise architecture.

1.7 What remains to be trusted in ZT environment

In the zero trust architecture all access requests of a subject to resources are evaluated based on pre-established criteria. Many implementations also use adaptable criteria which take into account information about recent events. By default the subjects are untrusted.

However, there is at least one essential building block of a zero trust architecture which needs to be trusted: the trust algorithm used by the Policy Engine. The correct configuration and functioning of the trust algorithm is essential to both the resource owner and the requester who needs access to the resource. It is in the interests of both to block unauthorized access and approve legitimate access. Verifying the correct functioning of the trust algorithm is not possible for a basic user, and difficult also for the resource owner in a complex environment.

For a subject accessing a system it is important that the information stored and processed in the system is accurate and up to date. However, often the subject cannot verify this, but has to trust the correct functioning of the system and accuracy and timeliness of the data.

1.8 Physical and network perimeters are still needed

There are many security frameworks which still require information system assets to be placed in secure areas protected with strong physical perimeters – in some cases several nested physical protection layers are required. The same frameworks also include a network architecture with several security layers and strong restrictions to the traffic passing through the perimeters. PCI Council's security standards for card production (PCI-CP) and Finnish National Security Authority's auditing criteria (Katakri) are examples of such frameworks. PCI-CP defines security rules for environments where payment card production data is processed, and Katakri defines security measures of governmental classified data processing.

Physical access to a computer or a network device offers an attacker many interesting possibilities. The attacker can add to a computer a keylogger or USB device which collects data, connect a traffic recording device to a network switch, steal a hard disk from the storage system … and many other things which cannot be done

remotely. Naturally, the attacker can also damage the equipment. Physical security arrangements with several nested perimeters, well managed physical access rights and good physical intrusion detection capabilities give the organization time to react before the attacker has reached the target.

Like in physical security, the *defence in depth* model is in use also in networks where security critical data is stored and processed. In order to reach the target – the valuable data – an external attacker must pass several control points which alert the organization if they detect suspicious activity. This again enables the organization to react to the attack before the "crown jewels" are compromised. Controls must be in place when transferring data between the protected zone and external entities – for instance encryption of data, communication only with pre-approved parties. Naturally controls must be applied also to the entities located inside the perimeter.

Organization's decision on security architecture should be risk-based. The risks to be assessed can be regulatory (compliance with mandatory security criteria), operational, or any type of risk which is relevant for the context.

In a zero trust environment the most severe risk could be compromising of the core components, which make the access decisions. These trusted components should be protected accordingly – which means well controlled physical and logical perimeters.

2. Human aspects of trust

2.1 Trust and distrust

Cambridge dictionary (2023) defines the verb trust "to believe that someone is good and honest and will not harm you, or that something is safe and reliable".

McKnight and Chervany (2000) analyse the challenges in defining trust. They also present a summary of various trust definitions. The following characteristics are included in several definitions of trust: competence, predictability, benevolence, integrity.

The trustor cannot, however, be sure or verify that the trustee meets these expectations. She/he has to interpret the trustee's often indirect signals to make the trusting decision. Bachrarach and Gambetta (2001) describe a trust decision-making model in the form of a trust game. In this game the trustor is in a vulnerable position.

People expect reciprocity in their relationship with others. Hart and Regner (2017) state that lack of reciprocity can create a spiral of distrust in which can be hard to break. Sagarin et al (1998) shed light on the deceiver's position in the trust game. The person who betrays other, assumes that also others are economical with the truth, which gives to the deceiver a justification for not feeling guilty for being dishonest.

Castelfranchi describes trust as a three-level model. First, there must be the disposition to trust. Second, trusting requires trustor's decision to trust, and the final step is the act of trusting.

2.2 Trust in organizational culture

Organizations are what people belonging to the context make them. Klynn (2021) describes the necessity and positive impact of trust in an organization. In a trusting atmosphere we can expect people to be honest and act in a way which is beneficial for both parties. Communication and cooperation are easier when we don't need to doubt the others' words and good intentions. So the mental energy of people can be addressed more to useful and productive work instead of repeated and heavy controls.

Reina (2006) describes a three-dimensional model of trust – transactional trust. The key components of the model are:

- *Contractual trust* implies that there is mutual understanding that people do what they say they will do. Managing mutual expectations, keeping agreements and consistency are essential for contractual trust.
- *Communication trust* main characteristics are the willingness to share information, telling the truth, mutual feedback and speaking with good purpose. Trust and communication are strongly linked with each other.
- *Competence trust* means acknowledging people's skills, allowing people to make decisions.

Lack of trust can have many negative impacts on the organization. If we imagine an organization which has a low level of trust, what could be the impacts of the low trust level in the three components of transactional trust:

- *Low contractual trust* – If you cannot trust that people do what they say and keep the agreements, getting things done and maintaining situational awareness requires significant and detailed monitoring which increases administrative overhead. As people need trusted relationships, low contractual trust could increase the appearance of informal groups. These groups can promote their own agendas, the agendas of different groups may be in conflict with each other and with the organization's objectives. A shadow organization is born.
- *Low communication trust* – When there is low level of trust people tend to withhold information to ensure their own position. On the other hand, having insufficient and inaccurate information is demotivating.
- *Low competence trust* – Competence trust could be described as the *fair play rule*. When people's skills are recognised and they are empowered and can make decisions, the organisational overhead decreases and people see themselves as valuable players in the team. But if the skills are not properly valued and even small decisions must be taken to high level in the organisation people are not committed to the organisation's objectives and can also have a sense of being meaningless and easily replaceable.

3. The impact of technology on humans

3.1 The persuasive power of information technology

A lot of research has been made about human-system interaction (HSI) from different points of view. HSI issues specific to the zero trust environment have so far been addressed mainly as a part of zero trust technical articles – especially the impact of a rigorously tuned trust algorithm which obliges the users to present their credentials frequently, which has a negative impact on the users evaluation of the system.

Information systems are created to serve the business in which they are used. Business application vendors create standard solutions which can be adapted to different use cases. However, a new information system may also oblige to make changes to the business process. So, information systems change the business – not only by increasing efficiency but also by changing business process of the target environment.

But do computers and systems also change the people who use them? And if they do, how?

Ndudule et al (2022) have studied personality-targeted persuasive gamified systems and how effective the persuasive strategies of these systems are in promoting behaviour change. The results of the research show that the strategies can be very effective, but the effectiveness depends on the domain and the personalities of the people involved.

Turkle (2004) describes based on a long term experience the impact that information technology has on people's thinking and mutual relationships. Based on some events she began to study not only what computers do *for* us, but also what they do *to* us. The impact described in the article include the following:

- Transition from slide rules to calculators weakened ability deal with scale.
- People tend to project thoughts and feelings onto their machines, and computational objects are increasingly designed to have emotional and cognitive effects.
- Silent approval of the use of surveillance technology and as consequence more limited (or loss of) privacy.
- Presentation tools like PowerPoint not only provide a way to transmitting content, they also present a model of thinking and aesthetic – and presentation itself becomes its own powerful idea. It is also mentioned that PowerPoint equates bulleting with clear thinking.
- Complex simulations can accustom us to manipulating systems whose core assumptions are hard to understand and may not be true.

Turkle suggests that it is important to develop information technology literacy, and we should be capable of asking the journalists' traditional questions: who, what, when, where, why and how. Since 2004 when the article was written, information technology has increasingly penetrated our everyday lives. And as the tools we use changes us, being aware of these manipulating mechanisms is essential.

Fogg (2003) analyses how computers can have the role of persuasive social actors. A social actors can be persuasive by rewarding with positive feedback, modelling a target behaviour or attitude, and providing social support. And people respond to computer system as though they were social entities. The persuasion uses different types of social cues. The article includes several interesting examples, for instance the following:

- Attractive device or interface has more persuasive power than unattractive technology.
- Computers motivate and persuade more effectively if they show personality traits similar to the user.
- Use of language which indicates social presence stimulates the user to act in a certain manner – for instance to buy more products from an e-commerce site.

A frequent discussion in these days is the impact of social media. These platforms tend to promote powerful emotional expressions which increase polarisation in human relationships and in the society.

3.2 Impact of zero trust architecture on humans

As described in the previous section, the technology itself has an impact on the users. The impact is not limited to the context in which the technology is used, but can extend to the users' perception of the reality.

What specific impacts could the zero trust approach have on its users and on the organizations in which it is used? Based on the information about the impact of other types of systems and software it is possible to create some scenarios.

Let us return to the question how information systems don't only do things *for us* but also do things *to us*. As Turkle (2004) explains information systems don't only transmit content, they can also carry their own way of thinking – the ideas of the creator or publisher of the system. Using the systems can impact our thinking and understanding of the reality in a way that we aren't even aware of.

According to Turkle (2004) PowerPoint has been developed to serve the needs of corporate boardrooms, and the software is designed to convey authority. What kind of thinking could a zero trust system represent? If we ask impacts of zero trust systems from the technical specialists of zero trust could result in the conclusion that the impacts of ZT approach are beneficial – we can trust these systems because they don't trust us as users but verify our credentials by request before granting access. Creating trust by not trusting – a bit of a paradox.

How is the message of the basic assumptions – *never trust, always verify or verify, then trust* – transmitted to the users of ZT systems? Let us have a look at the impact from the point of view of an individual user. If the user does not know the logic of the trust algorithm making the access decision – which often is the case – there is clearly a setup of power distance. The sensation of being powerless is enhanced if the system keeps asking the user to frequently represent the credentials. If people have a repeated experience of being untrusted they also tend to trust others less. The spiral of distrust can fuel itself from this individual experience.

What could the impact be on organizational level? As discussed earlier in section 2.2, low level of trust in an organization has several negative impacts. Lower individual performance and higher level of controls are typical in a low trust environment, which decreases organisation's efficiency. Lack of trust is demotivating, and demotivated people are not likely to take initiative which is beneficial to the organisation.

But do ZT systems increase the level of distrust in the organization more than other types of systems? There is no clear answer to this question, as research results on this topic are not available. Experience with other types of solutions demonstrate that information systems can change the users, and the type of change is linked with the planning premise of the system. ZT systems' basic assumption is *never trust, always verify* – ZT systems could have a negative impact on the trust in the organisation using them. And this can have many harmful consequences, including a spiral of distrust which may be difficult to break. Already the term *zero trust* itself can cause negative associations in people's minds.

4. Conclusion

It is clear that the zero trust approach has been adopted for a good reason – other architectural ideas do not respond to the needs of current business models and complex information systems. When there is a business case, the transition to the new architectural model should be well managed, and risk assessment is a necessary part of the project.

Performance and service level problems are potential issues in a ZT implementation. The systems can also have negative impacts on the user experience and organizational trust. Therefore, a strategy for managing the potential negative impacts of the systems should be part of a ZT implementation project.

It seems that zero trust approach is neither a magic bullet resolving all access problems nor devil's advocate destroying the trust in the organization – it seems to have features of both.

References

Bacharach M., Gambetta D. (2001), *Trust in Signs, Trust in society p 148-182,* Russell Sage Foundation, New York

Bachrach, M., Gambetta, D. (2001) *Trust in signs*, In K. S. Cook (Ed.), Trust in society (pp. 148–184). Russell Sage Foundation

Cambridge dictionary (2023) https://dictionary.cambridge.org/dictionary/english/trust

Castelfranchi, C. (2008) *Trust and Reciprocity: Misunderstandings,* International Review of Economics

Fernandez E.B. et al (2023) *A critical analysis of Zero Trust Architecture (ZTA)* Available in Internet as preprin

Fogg, B. (2003) *Computers as Persuasive Social Actors*, Persuasive Technology: 89-120, Morgan Kaufmall, Burlington

Hart, N., Regner, T. (2017) *The spiral of distrust: (Non-)cooperation in a repeated trust game is predicted by anger and individual differences in negative reciprocity orientation*, International Journal of Psychology, 2017 Vol. 52, No. S1, 18–25

Jericho Forum (2007) *Jericho Forum Commandments*

Klynn, B. (2021) *Building a Culture of Trust,* The Ohio State University, Fisher College of Business

Marsh, S (1994) *Formalising Trust as a Computational Concept,* University of Stirling

McKnight, H., Chervany, N. (2000) *What is Trust? A Conceptual Analysis and an Interdisciplinary Model*, Americas Conference on Information Systems (AMCIS) 2000 Proceedings

National Security Authority of Finland (2020) *Katakri – Information Security Audit Tool for Authorities*

Ndudule, Chinenye et al (2022) *Personality-targeted persuasive gamified systems: exploring the impact of application domain on the effectiveness of behaviour change strategies,* User Modeling and User-Adapted Interaction (2022) 32:165–214, Springer

Netskope (2021) *White paper: Blueprint for Zero Trust in a SASE Architecture, Continuous Adaptive Trust—The Key to Adopting Zero Trust and SASE and How to Get There*

PCI Council 2022) *Payment Card Industry (PCI) Card Production and Provisioning, Logical Security*

PCI Council (2022) *Payment Card Industry (PCI) Card Production and Provisioning, Physical Security*

Reina, D., Reina, M. (2006) *Trust and Betrayal in Workplace,*Berrett-Koehler Publishers Inc, San Francisco

Rose, S. et al (2020) *NIST Special Publication 800-207, Zero Trust Architecture*, NIST Gaithersburg, MD

Sagarin, B., Rhoads, K., Cialdini, R. (1998*) Deceiver's Distrust: Denigration as a Concequence of Undiscovered Deception*, Personality and Social Psychology Bulletin 24(11):1167-1176

Saltzer, J. H.; Schroeder, M. D (1975) *The protection of information in computer systems*. Proceedings of the IEEE, 1975, 63(9).

Townley, Andrew (2023) *Zero Trust: so much noise and confusion* – Newsletter of Archistry Inc

Turkle, S. (2004) *How Computers Change the Way We Think,* The Chronicle of Higher Education

Ward, R., Beyer, B. (2014) *BeyondCorp: A New Approach to Enterprise Security*. Login 20014, Vol. 39, vol. No. 6, pp. 6-11

Zscaler (2022) *A Brief History of Zero Trust*