Zero Trust Philosophy versus Architecture

William R. Simpson

***Abstract* —. Zero trust assumes all points of trust will be questioned and mitigated, the individual resources are protected and there is no reliance on the network for protection. This has the goal of limiting threat mobility and containing damage. The presentation of rules for multifactor authentication and micro-segmentation are often cited as a Zero Trust Architecture (ZTA), but what is often missing in these so call architectures is what to do about major points of trust in the system. Zero trust is not achievable soley with theses approaches, and only minimal trust can be cultivated. Certain trust points are inevitable including Certificate Authorities, Policy evaluation and decision points and others. The more general Zero Trust Philosophy (ZTP) covers not only those architectural issues, but also the philosophical ones. The ZTP allows the network architect to examine each trust point and make a decision about verification and validation.**

***Index Terms* — Zero Trust, Minimal Trust, Network Defense, Networking, Security Architectures**

I. INTRODUCTION

Implementations of zero trust architecture are only the beginning of zero trust. It is of the items that are not listed in the zero trust architecture that form the greatest risk to information security. In this paper we will discuss just a few of those IT mechanisms that are usually taken for granted and applied in a zero trust approach.

II. WHY ZERO TRUST, WHY NOW?

Zero Trust (ZT) is a new way to structure security defenses to better defend our digital resources against attackers. It is not a product or a security tool, but a way to organize the resources and the tools we use to protect them. Instead of a network-based defense, which places protections at the network boundary, ZT is a resource-based defense that places protections at each valuable resource. This provides a better match to current threats by directly protecting what is being attacked, and it provides a more resilient defense against lateral movement within an organization. For the Department of Defense (DoD) at this time, the current defense builds upon a clear concept of the fortress approach. Many of the requirements are based on inspection and reporting prior to delivery of the communication to the intended target. The inspection and reporting requires numerous of software tools to preclude malicious entities from exfiltration of data, theft of credentials, blocking of services, and other nefarious activities.

Manuscript received 30 Oct. 2021; revised 21 Jan 2022. This work was supported by the Institute for Defense Analyses. Such support does not constitute an endorsement by either the Institute for Defense Analyses or the U. S. Department of Defense.

William R. Simpson, corresponding author, Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, VA 22311 USA

These inspections require decryption of packets, which implies that the defensive suite either impersonates the requestor or has access to the private cryptographic keys of the servers that are the target of communication. Advanced persistent threats repeatedly bypass and defeat this approach. The network-based approach has been repeatedly broken, which shows that it has not been working for some time. ZT offers a new approach to defend our networks and digital resources.

A. The Current Approach

The current approach to security creates clusters of resources within network boundaries. All resources within a network segment receive protection from a set of security tools located at the boundary (or front door) of that network segment. Computer network defense is defined as "Actions taken through the use of computer networks to protect, monitor, analyze, detect and respond to unauthorized activity within the enterprise information systems and computer networks" [1]. The current defense package assumes that the threat can be stopped at the front door, as shown in Figure 1. All traffic in the enterprise, both coming and going, is routed through this front door. The front door is often onerous enough that administrator back doors are made available [2] to bypass many of the security checks. These backdoors, in addition to crating credential theft and threat stack vulnerabilities, are often the target of exploits. One example is the recent SolarWinds attack [3].

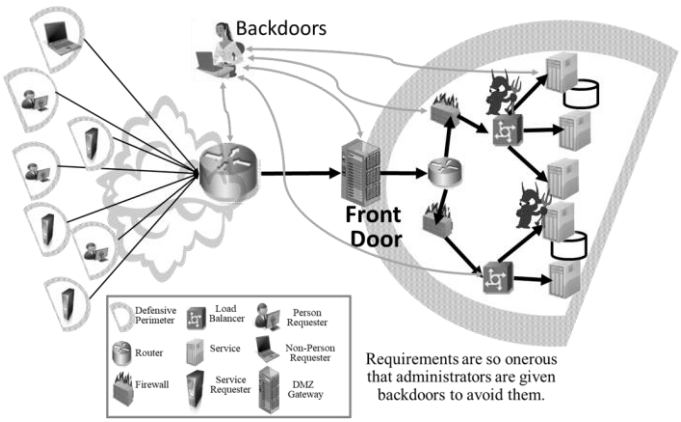

Figure 1. Fortress Protected Enterprises

The elements involved in implementing network and application defense are numerous and complicated. A wide range of appliances provides functionality. This functionality may be for quality of service to the user or quality of protection to network resources and servers. These appliances are often placed in-line, and some require access to content to provide their service. The literature is confusing because offerings include multiple services under

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

various titles such as multi-function firewalls or advanced defense systems. The fortress defense has spectacularly failed with breaches occurring daily. The appliances in the package do stop the current threats for a short period, but new threats materialize very shortly and once again defeat the fortress approach. Even with detection and mitigation, we have continued threat presence over long periods. The advanced approaches described here assume that the threat is present and in the enterprise at all times. Although this may not be true at any given time, it is certainly true at various times during operations.

B. Zero Trust

To fix the problems associated with network defense at the border, a new approach is needed. ZT is better suited to combating the current attack methods while preserving existing end-to-end security measures. ZT changes the one-size-fits-all security approach of a boundary defense to a custom-tailored approach for each resource within that boundary. The defenses are implemented at the resource, so there is no gap between the security and the resource it protects. ZT is an endpoint-based solution. It does not break the end-to-end secure communication channel between requester and resource. It scans at the endpoints and reports findings to a central monitoring facility. This allows requester and provider to authenticate each other directly and perform encryption and integrity from end to end. By focusing on the endpoints, ZT eliminates the man-in-the-middle (MITM) that boundary security introduces.

Many of the new security techniques have moved to a distributed security approach. The ZT-framework is a distributed security system that eliminates or mitigates many of the primary vulnerability points inherent with the fortress system, as shown in Figure 2. Each entity needs assurance that the entity and device they are engaged with are known entities and, specifically, the ones to whom the communication should be allowed. However, it is this distributed approach and the requirement for content inspection and reporting that causes the conflict between this approach and the traditional fortress representation. All active entities and devices in ZT systems have public key infrastructure (PKI) certificates. Identity may be bolstered by using multi-factor techniques, and temporary credentials may be issued when necessary. Communication between active entities requires bilateral, PKI, end-to-end authentication of both the participants and their hardware.

ZT represents a change from current security practice. Instead of protecting resources by blocking outsiders, the protections are placed at the resources themselves. This approach is a better match to the current threats, which are consistently breaking through firewalls and other boundary protections. ZT provides defense against outsiders and malicious insiders, and it blocks attacker lateral movement within an enterprise.

Many of the new security techniques have moved to a distributed security approach. The ZT framework is a distributed security system that eliminates or mitigates many of the primary vulnerability points inherent with the fortress system, as shown in Figure 2. Each entity needs assurance that the entity and device they are engaged with are known entities and, specifically, the ones to whom the communication should be allowed. However, it is this distributed approach and the requirement for content inspection and reporting that causes the conflict between this approach and the traditional fortress representation. All active entities and devices in ZT systems have public key infrastructure (PKI) certificates. Identity may be bolstered by using multi-factor techniques, and temporary credentials may be issued when necessary. Communication between active entities requires bilateral, PKI, end-to-end authentication of both the participants and their hardware.

ZT represents a change from current security practice. Instead of protecting resources by blocking outsiders, the protections are placed at the resources themselves. This approach is a better match to the current threats, which are consistently breaking through firewalls and other boundary protections. ZT provides defense against outsiders and malicious insiders, and it blocks attacker lateral movement within an enterprise.

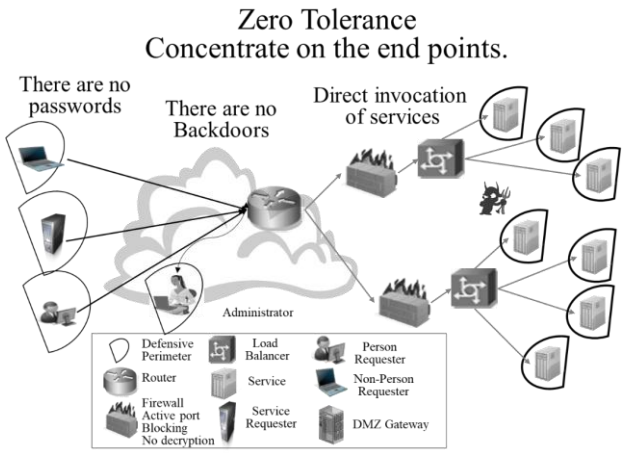

Figure 2. ZT Enterprise

III. ZT Architecture

To achieve this vision, most architectural representations provide five foundational concepts for a ZT approach:

1. Two-way strongly authenticated communication, often with multi-factor authentication processes.
2. Endpoint device management
3. End-to-end encryption and integrity, not always specified as unbroken, but for true zero trust processes encryption should be unbroken between two communicating endpoints.
4. Policy-based authorization. This may include role based and claims based access control.
5. Accountability for actions

In the DoD, these techniques have been fully developed, tested, and verified on the National Cyber Range and are described in the Air Force Consolidated Enterprise IT Baseline [4-6].

ZTA was designed to address lateral threat movement within the network. ZTA embraces the principle of never trust, always verify. ZTA is a paradigm that moves defenses from network-based perimeters to focus on users, assets, and resources. More information on ZTA is provided by NIST SP 800-207 [7]. However, these simplified

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

architectuaral elements only partially address the zero trust approach.

IV. THE OTHER PART OF ZT

There are many examples of current IT approaches that are counter to the ZTP. A few examples are presented here.

A. SSO and ZTA

Single sign-on (SSO) is a convenience for users to avoid multiple authentication instances in computer-based sessions. It is a way to centralize authentication for a collection of related resources. It simplifies the process of authentication by providing users a single place to establish their identity, and a single method for resources to authenticate requesters. Zero Trust (ZT) Architecture (ZTA) is a security approach that moves protections away from network borders and to the resources themselves. It removes the ability and need to trust networks and requires each requester to prove access based on their credentials at the time of a request. The question is whether these two can work together. The short answer is "No," but the full answer is more nuanced because the term SSO is used somewhat loosely. We look at the concepts of SSO and ZTA and show how the most common use of SSO does not work with ZTA.

SSO transfers authentication information between endpoints. The SSO server creates an SSO token after a requester authenticates to the SSO server [8]. This authentication may be tailored to the resource the user is requesting, with multi-factor or other methods to provide different strengths of authentication. In addition, the SSO server may provide many different options to accommodate users with different credentials, locations, and devices. The primary motivation to adopt SSO is often ease of use. This applies to both the users and the enterprise. The users have a single portal for authentication that accommodates all users, and the enterprise implements one authentication server and simply implements token processors at the resources. It is centralized, efficient, and easy to use.

However, SSO is typically not secure. Any authentication token that can be reused or transferred between users allows impersonation, a fundamental violation of basic security. SSO tokens are often implemented as "bearer tokens," meaning that the bearer (whether a proper user or attacker) can use the token to authenticate as the associated requester. SSO tokens are protected by Hypertext Transfer Protocol Secure (HTTPS) from SSO server to requester and again from requester to resource, but this piecemeal security leaves a gaping hole at the requester. Tokens that are implemented as a URL parameter or a cookie in the HTTP header can be easily copied and shared among users. The SSO approach is better than no security, but it falls short of the Department of Defense's (DoD) needs, and the complexity of proper implementation means a one-size-fits-all approach will cater to the lowest security level of the systems it supports.

SSO authenticates on one connection and provides resources on another connection. This violates the zero trust assumptions. Although SSO authentication to the SSO server is dynamic and may be strictly enforced, the access is being granted at the resource, and the resource only receives a static SSO token, not a dynamic, interactive authentication. This also violates zero trust assumptions..

The problem is that the SSO token provides no guarantee that the holder of the token is the entity named in the token. It is a bearer token. Thus, security relies on externally trusted entities, policies and practices. This is not the ZT approach.

SSO is a broad term that can mean many things, and some implementations are better than others. However, the key problem for ZTA is the reliance on trust of external elements. One is the user. A user can easily extract, copy, and share the SSO token received from the SSO server. If a user can do it, an attacker can do it too. Often, the attackers are better at this than most users, and stopping these attacks can be difficult due to the contrasting requirements for security and maximum functionality in browsers and web protocols.

B. Identity Proxies

Identity proxies are provided at various points within the network, often associated with load balancers, but not exclusively. These proxies assume the identity of the requester and are a classic man-in-the-middle in all transactions. They act on the behalf of the user consolidating web service activities and other features. They may claim to provide security services and are often provided access to hardware storage modules, and/or private keys of users and servers. A considerable trust is given to these proxies which are not free of vulnerabilities and offer attack vector opportunities for advesaries. They violate several aspects of zero trust and should be avoided if at all possible.

C. Security Scanners in a ZT Architecture

Current security best practice includes the use of security scanners. These look for patterns in data, behavior, or other aspects of the network or its traffic in order to automatically identify, document, and stop potentially malicious activity. The most capable scanners operate at the application layer and understand the protocols and data formats in use. This requires access to encrypted content, which requires breaking end-to-end secure connections.

Security scanners typically operate on network traffic. In some cases, the scanner resides where network traffic naturally converges, such as a firewall or gateway router. Often these are the points of maximum traffic, and require hardware4 accelerators and high-speed interfaces. In other cases, traffic is explicitly routed to the scanner before it is sent onward to its destination. This routing may be load balanced to reduce throughput requirements. The first approach works well for traffic between enclaves, and the second works better for traffic within an enclave. The scanner breaks the end-to-end encryption of the communication, scans traffic that has already been decrypted, or simply scans unencrypted traffic as-is. A notional setup of traffic scanners is provided in Figure 3.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

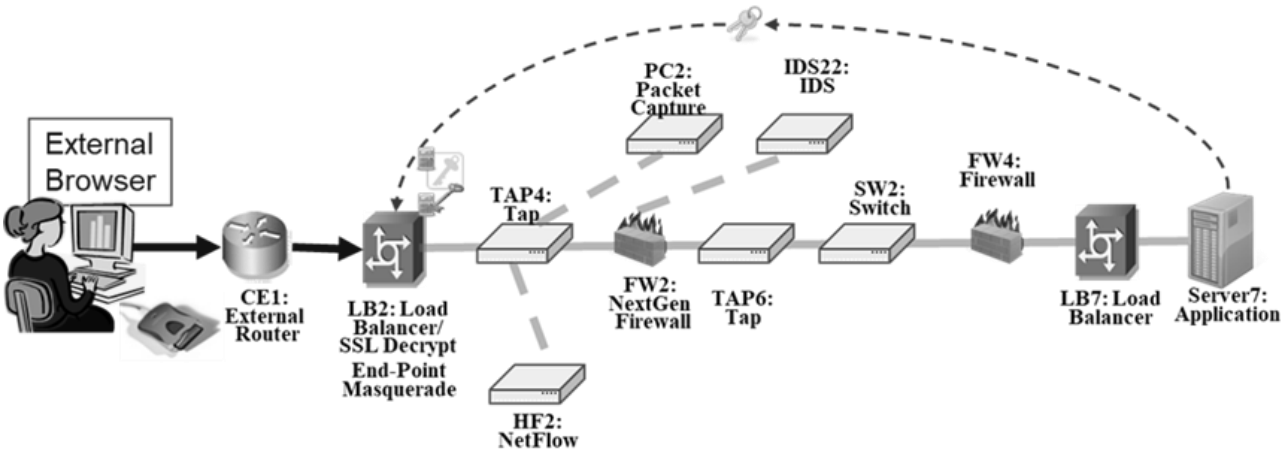

Figure 1. Notional Setup of a Scan Architecture

In a secure enterprise, traffic is encrypted, so scanning requires breaking this encryption. This can be accomplished by another entity, such as a load balancer (as shown in Figure 1) or a firewall, which integrates with the security scanner. In this case, the scanner is positioned within or near the other entity so that it can operate on the unencrypted data. Scanners can also sometimes break encryption themselves, which removes the requirement for other entities and allows more flexibility of placement. In either case, as the traffic is decrypted between the two endpoints, the two endpoints must trust this man-in-the-middle (MITM).

Each entity in a communication must have assurance that the party they are engaged with is a known entity and, specifically, the one to whom the communication is intended. Access and privilege should only be granted to an authenticated identity if credentials for access and privilege are presented, verified, and validated. Finally, all communications should be encrypted and provided with integrity protections that allow the recipient of communications to verify that what was received was actually sent. References [9] and [10] provide extensive descriptions of these processes.

Security scanners, as currently implemented, violate zero trust assumptions. They prevent end-to-end secure communication by explicitly breaking these connections and scanning the contents. The security scanners act as the MITM, which is not permitted by ZT. The MITM scanners also prevent dynamic authentication. The endpoints can only authenticate to the MITM and can only authenticate the MITM. As a requester, there is no way to know whether content from the MITM accurately reflects the data that the actual source provided, or whether the MITM even retrieved the data from the intended source or just generated the data itself.

Although most scanners are benevolent, they are not attack-proof. Every piece of hardware and software has vulnerabilities, and security scanners are no exception. A compromised security scanner acting as the MITM is particularly dangerous because an attacker can potentially view or modify any traffic in the enterprise in arbitrary ways. This is why the ZT approach necessitates not trusting the network. The assumption is that attackers are in the network already, which may include the network security scanners.

A new approach to security scanning is required for ZT. This approach must be consistent with endpoint-based security, because ZT protects the resources, which are located at the endpoints. This new approach has three key features:

- Implement security scanners in software instead of hardware.
- Move scanners from the network to the endpoints.
- Tailor the scanners to the resources they protect.

The first feature enables the other two. Hardware boxes must be physically placed in a location on the network. They are expensive, high-maintenance, and difficult to duplicate or otherwise scale up or down. Hardware-based protection leads to a centralized approach due to the characteristics of the hardware itself, which is not consistent with ZT.

The second feature uses the software-based scanners to provide protection where it is needed: at the resource endpoints. These endpoints integrate the scanner software into their existing software at a point where the content to be scanned is available. If multiple layers are to be scanned, the scanner software can access multiple points of the processing pipeline, from raw network traffic to internal application data. No extra decryption or authentication is needed, because the scanner is now part of the endpoint itself instead of a separate entity. Although this does not eliminate the threat of compromise of this code, such a compromise now only affects one resource rather than the entire enclave.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

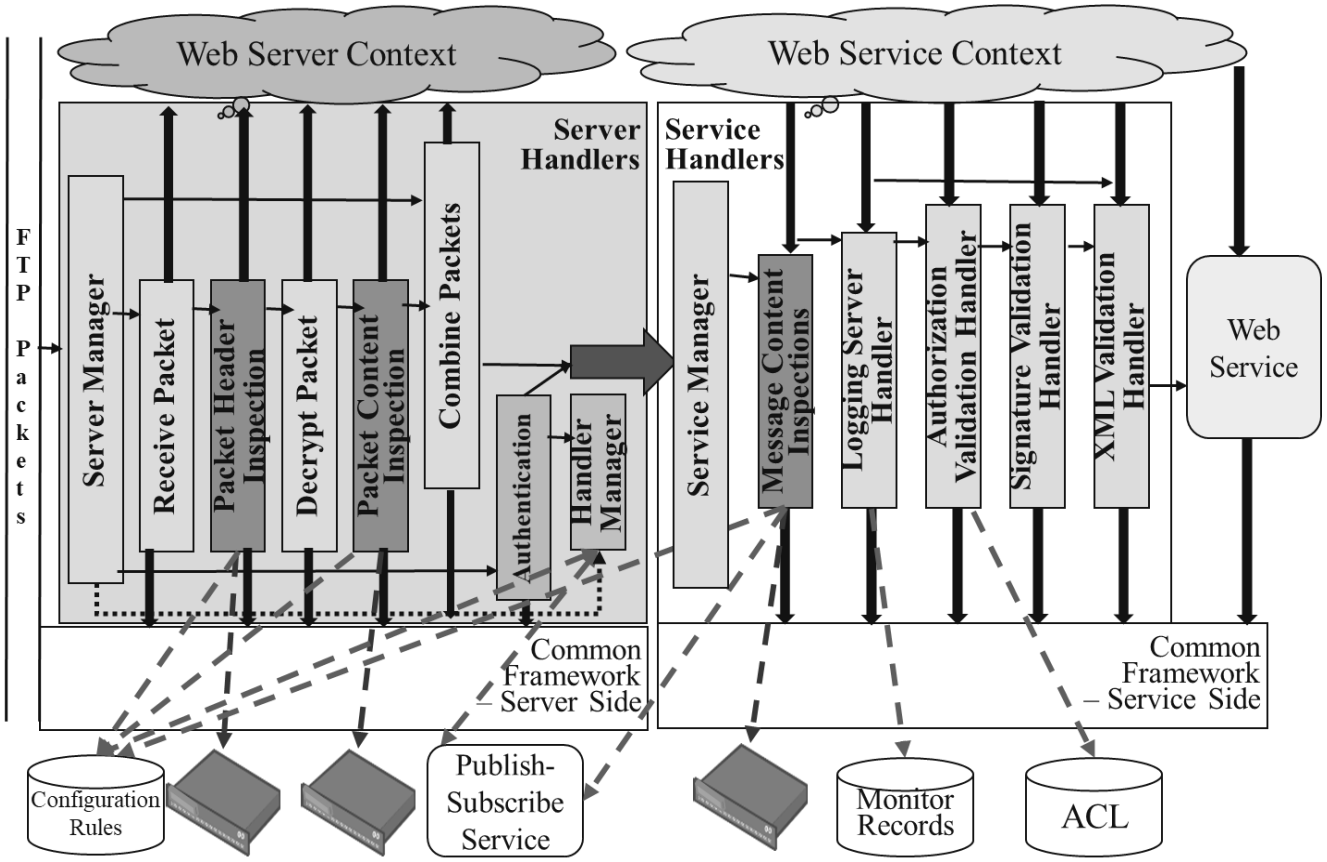

Figure 3. Notional Request Processing at the Server

The third feature uses the modular nature of individual software scanners to tailor the protections at any particular endpoint. Instead of implementing an identical stack of security scanners for all traffic, the scanners for a particular endpoint can be selectively implemented. An email resource uses email scanners, and a web server uses web traffic scanners. This reduces the performance requirements for the scanners because they are only scanning relevant traffic instead of all network traffic. Also, higher security resources could utilize a full security scanner stack, whereas lower security resources could selectively utilize a smaller set.

A notional server setup, providing end-point scanning is depicted in Figure 4. In the figure, the use of hardware appliances may be invoked by the server when software only versions of inspection prorocols are not available to be included in the handler chain. The figure also shows, as a part of the framework, the use of Access Contol Lists (ACL)s, and monitor record storage in the web service context. The handler management may be configured within the server, or with a publish subscribe service for greater flexibility. The publish-subscribe system amy also be used for required reporting under CNSSI and others. The migration path from the current approach to the ZT approach is fairly simple, but the benefits are only realized with a full transition. The initial transition can move scanners one-by-one from a central position to the endpoints. However, until all scanners are at the endpoints, the central MITM must remain, which negates ZT principles. Other benefits, such as performance, scalability, and tailoring of protections to resources can be achieved with a partial approach, but the core ZT ideas are only realized when all scanners are moved to the endpoints and the central MITM is removed.

V. OTHER CONSIDERATIONS

There are many other instances of zero trust violations including a partial list as:

- cloud-based applications,
- partial segmentation,
- server-side authentication versus bi-lateral authentication,
- load balancing,
- SSL consolidation,
- web accelerators,
- and separate policy evaluation and policy enforcement points,

The Philosophical part of zero trust would urge the examination of each contruct in the network defense and evaluation the trust implications and what it will do to the verification and validation required by a zero trust philosophy.

VI. CONCLUSION

There is a significant difference between a zero trust architecture and a zero trust philosophy. While the architecture can provide secrity elements that minimize the trust, it does not provide a way to prevent that trust from being given away at other points within the network system.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

It is only through a comprehensive application of the zero trust philosophy that this can be resolved. This work is part of a broader basedx examination of network architectures. A portion of these are covered in references [11] – [16]

REFERENCES

[1] Andress, Jason, and Steve Winterfeld. "Computer Network Defense." In *Cyber Warfare*, pp. 179–191. Rockland, MA: Syngress, 2011.

[2] TechTarget.com. "Backdoor (Computing)." https://searchsecurity.techtarget.com/definition/back-door, last accessed November 22, 2019.

[3] Datta, Pratim. "Hannibal at the Gates: Cyberwarfare and the Solarwinds Sunburst Hack." *Journal of Information Technology Teaching Cases*, March 12, 2021. https://doi.org/10.1177/2043886921993126.

[4] *Technical Profiles for the Consolidated Enterprise IT Baseline*, release 6.0. https://intelshare.intelink.gov/sites/afceit/ (CAC required).

[5] Simpson, William R., *Enterprise Level Security – Securing Information Systems in an Uncertain World*. Boca Raton, FL: Auerbach Publications, 2016.

[6] Simpson, William R., and Kevin E. Foltz. *Enterprise Level Security 2: Advanced Techniques for Information Technology in an Uncertain World* Abindgon, United Kingdom: Taylor & Francis Group, 2020.

[7] Scott Rose, Oliver Borchert, Stu Mitchell, and Sean Connelly, National Institute of Standards and Technology (NIST) SP 800-207, Zero Trust Architecture, August 2020.

[8] Teravainen, Taina. "Single Sign-on (SSO)." https://searchsecurity.techtarget.com/definition/single-sign-on, last accessed April 26, 2021.

[9] Organization for the Advancement of Structured Information Standards (OASIS) open set of Standards S. Cantor et al., "Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0," OASIS Standard, March 2005

[10] William R. Simpson, Kevin E. Foltz, Proceedings of the 9th International Conference on Software Engineering and Applications (JSE 2020), Ed: David C. Wyld, Natarajan Meghanathan, ISBN: 978-1-925953-28-2, "Network Defense in an End-to-End Paradigm," pp. 177–187, virtual presentation, Zurich, Switzerland, November 21–22, 2020.

[11] The 10th International Conference on Electronics, Communications, and Networks (CECNet 2020), "Secure Server Key Management Designs for the Public Cloud", October 25-27, 2020, Virtual Conference, Co-authored by Kevin Foltz.

[12] Proceedings of the 9th International Conference on Software Engineering and Applications (JSE 2020), David C. Wyld, Natarajan Meghanathan, (Eds): ISBN : 978-1-925953-28-2, "Network Defense in an End-to- End Paradigm", pp 177=187, virtual presentation, Zurich, Switzerland, November 21-22, 2020, DOI: 10.5121/csit.2020.101414, Co-authored by Kevin Foltz.

[13] International Journal of Network Security & Its Applications (IJNSA) Vol.13, No.1, "Resolving Network Defense Conflicts with Zero Trust Architectures and Other End-To-End Paradigms," January 2021, DOI: 10.5121/ijnsa.2021.13101, pp. 1-20, Co-authored by Kevin Foltz.

[14] Third International Conference on Internet of things, Data and Cloud Computing (ICC 2021), "Zero Trust Using Delegation of Access and Privilege," Cambridge, UK, June 2021, in process. Co-authored by Kevin Foltz.

[15] International Journal of Emerging Technology and Advanced Engineering, "Maintaining Zero Trust with Federation," Volume 11, Issue No. 3, March 2021 in process. Co-authored by Kevin Foltz.

[16] Lecture Notes in Engineering and Computer Science, Proceedings of the World Congress on Engineering (WCE) 2021, "Network Segmentation and Zero Trust Architectures," pp. 201-206, Virtual, Imperial College, London, 7-9 July 2021. Co-authored by Kevin Foltz.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)