

Withdrawn Draft

Warning Notice

The attached draft document has been withdrawn, and is provided solely for historical purposes. It has been superseded by the document identified below.

Withdrawal Date August 11, 2020

Original Release Date February 13, 2020

Superseding Document

Status Final

Series/Number NIST Special Publication 800-207

Title Zero Trust Architecture

Publication Date August 2020

DOI <https://doi.org/10.6028/NIST.SP.800-207>

CSRC URL <https://csrc.nist.gov/publications/detail/sp/800-207/final>

Additional Information

2

3 **Zero Trust Architecture**

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

Scott Rose
Oliver Borchert
Stu Mitchell
Sean Connelly

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.800-207-draft2>

C O M P U T E R S E C U R I T Y

20 **Draft (2nd) NIST Special Publication 800-207**

21
22
23
24

Zero Trust Architecture

25
26
27
28
29
30
31
32
33
34
35
36

Scott Rose
Oliver Borchert
Advanced Network Technologies Division
Information Technology Laboratory

Stu Mitchell
Stu2Labs
Stafford, VA

Sean Connelly
Cybersecurity & Infrastructure Security Agency
Department of Homeland Security

37
38
39
40
41
42
43
44

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.800-207-draft2>

February 2020



45
46
47
48
49
50
51
52

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Under Secretary of Commerce for Standards and Technology

53

Authority

54 This publication has been developed by NIST in accordance with its statutory responsibilities under the
55 Federal Information Security Modernization Act (FISMA) of 2014, 44 U.S.C. § 3551 *et seq.*, Public Law
56 (P.L.) 113-283. NIST is responsible for developing information security standards and guidelines, including
57 minimum requirements for federal information systems, but such standards and guidelines shall not apply
58 to national security systems without the express approval of appropriate federal officials exercising policy
59 authority over such systems. This guideline is consistent with the requirements of the Office of Management
60 and Budget (OMB) Circular A-130.

61 Nothing in this publication should be taken to contradict the standards and guidelines made mandatory and
62 binding on federal agencies by the Secretary of Commerce under statutory authority. Nor should these
63 guidelines be interpreted as altering or superseding the existing authorities of the Secretary of Commerce,
64 Director of the OMB, or any other federal official. This publication may be used by nongovernmental
65 organizations on a voluntary basis and is not subject to copyright in the United States. Attribution would,
66 however, be appreciated by NIST.

67 National Institute of Standards and Technology Special Publication 800-207
68 Natl. Inst. Stand. Technol. Spec. Publ. 800-207, 58 pages (February 2020)
69 CODEN: NSPUE2

70 This publication is available free of charge from:
71 <https://doi.org/10.6028/NIST.SP.800-207-draft2>

72 Certain commercial entities, equipment, or materials may be identified in this document in order to describe an
73 experimental procedure or concept adequately. Such identification is not intended to imply recommendation or
74 endorsement by NIST, nor is it intended to imply that the entities, materials, or equipment are necessarily the best
75 available for the purpose.

76 There may be references in this publication to other publications currently under development by NIST in accordance
77 with its assigned statutory responsibilities. The information in this publication, including concepts and methodologies,
78 may be used by federal agencies even before the completion of such companion publications. Thus, until each
79 publication is completed, current requirements, guidelines, and procedures, where they exist, remain operative. For
80 planning and transition purposes, federal agencies may wish to closely follow the development of these new
81 publications by NIST.

82 Organizations are encouraged to review all draft publications during public comment periods and provide feedback to
83 NIST. Many NIST cybersecurity publications, other than the ones noted above, are available at
84 <https://csrc.nist.gov/publications>.

85

86 **Public comment period: February 13, 2020 through March 13, 2020**

87 National Institute of Standards and Technology
88 Attn: Advanced Network Technologies Division, Information Technology Laboratory
89 100 Bureau Drive (Mail Stop 8920) Gaithersburg, MD 20899-8920
90 Email: zerotrust-arch@nist.gov

91 All comments are subject to release under the Freedom of Information Act (FOIA).

92

Reports on Computer Systems Technology

93 The Information Technology Laboratory (ITL) at the National Institute of Standards and
94 Technology (NIST) promotes the U.S. economy and public welfare by providing technical
95 leadership for the Nation’s measurement and standards infrastructure. ITL develops tests, test
96 methods, reference data, proof of concept implementations, and technical analyses to advance the
97 development and productive use of information technology. ITL’s responsibilities include the
98 development of management, administrative, technical, and physical standards and guidelines for
99 the cost-effective security and privacy of other than national security-related information in federal
100 information systems. The Special Publication 800-series reports on ITL’s research, guidelines, and
101 outreach efforts in information system security, and its collaborative activities with industry,
102 government, and academic organizations.

103

Abstract

104 Zero trust (ZT) is the term for an evolving set of cybersecurity paradigms that move network
105 defenses from static, network-based perimeters to focus on users, assets, and resources. A zero
106 trust architecture (ZTA) uses zero trust principles to plan enterprise infrastructure and
107 workflows. Zero trust assumes there is no implicit trust granted to assets or user accounts based
108 solely on their physical or network location (i.e., local area networks versus the internet).
109 Authentication and authorization (both user and device) are discrete functions performed before
110 a session to an enterprise resource is established. Zero trust is a response to enterprise network
111 trends that include remote users and cloud-based assets that are not located within an enterprise-
112 owned network boundary. Zero trust focus on protecting resources, not network segments, as the
113 network location is no longer seen as the prime component to the security posture of the
114 resource. This document contains an abstract definition of zero trust architecture (ZTA) and
115 gives general deployment models and use cases where zero trust could improve an enterprise’s
116 overall information technology security posture.

117

Keywords

118 architecture; cybersecurity; enterprise; network security; zero trust.

119

120

Acknowledgments

121 This document is the product of a collaboration between multiple federal agencies and is
122 overseen by the Federal CIO Council. The architecture subgroup is responsible for development
123 of this document, but there are specific individuals who deserve recognition. These include Greg
124 Holden, project manager of the Federal CIO Council ZTA project; Alper Kerman, project
125 manager for the NIST/National Cybersecurity Center of Excellence ZTA effort; and Douglas
126 Montgomery.

127

Audience

128 This document is intended to describe zero trust for enterprise security architects. It is meant to
129 aid understanding of zero trust for civilian unclassified systems and provide a road map to
130 migrate and deploy zero trust security concepts to an enterprise environment. Agency
131 cybersecurity managers, network administrators, and managers may also gain insight into zero
132 trust and ZTA from this document. It is not intended to be a single deployment plan for ZTA as
133 an enterprise will have unique business use cases and data assets that require protection. Starting
134 with a solid understanding of the organization's business and data will result in a strong
135 approach to zero trust.

136

Note to Reviewers

137 The purpose of this Special Publication is to develop a technology-neutral set of terms,
138 definitions, and logical architectural components to develop and support a ZTA. This document
139 does not give specific guidance or recommendations on how to deploy zero trust components in
140 an enterprise. Reviewers are asked to tailor their comments based on the stated purpose of the
141 document.

142

Trademark Information

143 All registered trademarks or trademarks belong to their respective organizations.
144

145

Call for Patent Claims

146 This public review includes a call for information on essential patent claims (claims whose use
147 would be required for compliance with the guidance or requirements in this Information
148 Technology Laboratory (ITL) draft publication). Such guidance and/or requirements may be
149 directly stated in this ITL Publication or by reference to another publication. This call also
150 includes disclosure, where known, of the existence of pending U.S. or foreign patent applications
151 relating to this ITL draft publication and of any relevant unexpired U.S. or foreign patents.

152

153 ITL may require from the patent holder, or a party authorized to make assurances on its behalf,
154 in written or electronic form, either:

155

156 a) assurance in the form of a general disclaimer to the effect that such party does not hold and
157 does not currently intend holding any essential patent claim(s); or

158

159 b) assurance that a license to such essential patent claim(s) will be made available to applicants
160 desiring to utilize the license for the purpose of complying with the guidance or requirements
161 in this ITL draft publication either:

162

163 i. under reasonable terms and conditions that are demonstrably free of any unfair
164 discrimination; or

165

166 ii. without compensation and under reasonable terms and conditions that are
167 demonstrably free of any unfair discrimination.

168

169 Such assurance shall indicate that the patent holder (or third party authorized to make assurances
170 on its behalf) will include in any documents transferring ownership of patents subject to the
171 assurance, provisions sufficient to ensure that the commitments in the assurance are binding on
172 the transferee, and that the transferee will similarly include appropriate provisions in the event of
173 future transfers with the goal of binding each successor-in-interest.

174

175 The assurance shall also indicate that it is intended to be binding on successors-in-interest
176 regardless of whether such provisions are included in the relevant transfer documents.

177

178 Such statements should be addressed to: zerotrust-arch@nist.gov

179

180

181 **Table of Contents**

182 **1 Introduction 1**

183 1.1 History of Zero Trust Efforts Related to Federal Agencies 2

184 1.2 Structure of This Document 2

185 **2 Zero Trust Basics 4**

186 2.1 Tenets of Zero Trust 6

187 2.2 A Zero Trust View of a Network 7

188 2.2.1 Assumptions for Enterprise-Owned Network Infrastructure 8

189 2.2.2 Assumptions for Nonenterprise-Owned Network Infrastructure 8

190 **3 Logical Components of Zero Trust Architecture 9**

191 3.1 Variations of Zero Trust Architecture Approaches 11

192 3.1.1 ZTA Using Enhanced Identity Governance 11

193 3.1.2 ZTA Using Micro-Segmentation 12

194 3.1.3 ZTA Using Network Infrastructure and Software Defined Perimeters. 12

195 3.2 Deployed Variations of the Abstract Architecture 12

196 3.2.1 Device Agent/Gateway-Based Deployment 13

197 3.2.2 Enclave-Based Deployment 14

198 3.2.3 Resource Portal-Based Deployment 14

199 3.2.4 Device Application Sandboxing 15

200 3.3 Trust Algorithm 16

201 3.3.1 Trust Algorithm Variations 18

202 3.4 Network/Environment Components 20

203 3.4.1 Network Requirements to Support ZTA 20

204 **4 Deployment Scenarios/Use Cases 22**

205 4.1 Enterprise with Satellite Facilities 22

206 4.2 Multicloud Enterprise 23

207 4.3 Enterprise with Contracted Services and/or Nonemployee Access 24

208 4.4 Collaboration Across Enterprise Boundaries 25

209 4.5 Enterprise with Public- or Customer-Facing Services 26

210 **5 Threats Associated with Zero Trust Architecture 27**

211 5.1 Subversion of ZTA Decision Process 27

212 5.2 Denial-of-Service or Network Disruption 27

213 5.3 Stolen Credentials/Insider Threat..... 28

214 5.4 Visibility on the Network..... 28

215 5.5 Storage of Network Information 29

216 5.6 Reliance on Proprietary Data Formats..... 29

217 5.7 Use of Nonperson Entities (NPE) in ZTA Administration 29

218 **6 Zero Trust Architecture and Possible Interactions with Existing Federal**

219 **Guidance..... 31**

220 6.1 ZTA and NIST Risk Management Framework 31

221 6.2 ZT and NIST Privacy Framework..... 31

222 6.3 ZTA and Federal Identity, Credential, and Access Management Architecture

223 32

224 6.4 ZTA and Trusted Internet Connections 3.0 32

225 6.5 ZTA and EINSTEIN (NCPS – National Cybersecurity Protection System) ... 33

226 6.6 ZTA and DHS Continuous Diagnostics and Mitigations (CDM) Program..... 33

227 6.7 ZTA, Cloud Smart, and the Federal Data Strategy 34

228 **7 Migrating to a Zero Trust Architecture..... 35**

229 7.1 Pure Zero Trust Architecture..... 35

230 7.2 Hybrid ZTA and Perimeter-Based Architecture 35

231 7.3 Steps to Introducing ZTA to a Perimeter-Based Architected Network..... 36

232 7.3.1 Identify Actors on the Enterprise 37

233 7.3.2 Identify Assets Owned by the Enterprise..... 37

234 7.3.3 Identify Key Processes and Evaluate Risks Associated with Executing

235 Process 38

236 7.3.4 Formulating Policies for the ZTA Candidate 38

237 7.3.5 Identifying Candidate Solutions 38

238 7.3.6 Initial Deployment and Monitoring 39

239 7.3.7 Expanding the ZTA..... 39

240 **References..... 41**

241

242 **List of Appendices**

243 **Appendix A— Acronyms 44**

244 **Appendix B— Identified Gaps in the Current State-of-the-Art in ZTA 45**

245 B.1 Technology Survey 45

246 B.2 Gaps that Prevent Immediate Move to ZTA..... 46

247 B.2.1 Lack of Common Terms for ZTA Design, Planning, and Procurement 46

248 B.2.2 Perception that ZTA Conflicts with Existing Federal Cybersecurity

249 Policies..... 46

250 B.3 Systemic Gaps that Impact ZTA 46

251 B.3.3 Standardization of Interfaces Between Components..... 46

252 B.3.4 Emerging Standards that Address Overreliance on Proprietary APIs 47

253 B.4 Knowledge Gaps in ZTA and Future Areas of Research 47

254 B.4.5 Attacker Response to ZTA 48

255 B.4.6 User Experience in a ZTA Environment 48

256 B.4.7 Resilience of ZTA to Enterprise and Network Disruption..... 48

257 B.5 ZTA Test Environment 49

258 B.6 References 49

259

260

List of Figures

261 Figure 1: Zero Trust Access 5

262 Figure 2: Core Zero Trust Logical Components 9

263 Figure 3: Device Agent/Gateway Model 13

264 Figure 4: Enclave Gateway Model 14

265 Figure 5: Resource Portal Model..... 15

266 Figure 6: Application Sandboxes..... 16

267 Figure 7: Trust Algorithm Input..... 17

268 Figure 8: Enterprise with Remote Employees 23

269 Figure 9: Multicloud Use Case 23

270 Figure 10: Enterprise with Nonemployee Access 24

271 Figure 11: Cross-Enterprise Collaboration 25

272 Figure 12: ZTA Deployment Cycle 36

273

274

List of Tables

275 Table B-1: Summary of Identified Deployment Gaps 45

276

1 Introduction

278 A typical enterprise’s infrastructure has grown increasingly complex. A single enterprise may
279 operate several internal networks, remote offices with their own local infrastructure, remote
280 and/or mobile individuals, and cloud services. This complexity has outstripped traditional
281 methods of perimeter-based network security as there is no single, easily identified perimeter for
282 the enterprise. Perimeter-based network security has also been shown to be insufficient since
283 once attackers breach the perimeter, further lateral movement is unhindered.

284 This complex enterprise has led to the development of a new model for cybersecurity principles
285 and network security known as “zero trust” (ZT). A ZT approach is primarily focused on data
286 protection but can be expanded to include all enterprise assets, such as devices, infrastructure,
287 and users. Zero trust security models assume that an attacker is present on the network and that
288 an enterprise-owned network infrastructure is no different—or no more trustworthy—than any
289 nonenterprise-owned network. In this new paradigm, an enterprise must continually analyze and
290 evaluate the risks to its internal assets and business functions and then enact protections to
291 mitigate these risks. In zero trust, these protections usually involve minimizing access to
292 resources (such as data and compute resources and applications) to only those users and assets
293 identified as needing access as well as continually authenticating and authorizing the identity and
294 security posture of each access request.

295 A zero trust architecture (ZTA) is an enterprise cybersecurity strategy that is based on zero trust
296 principles and designed to prevent data breaches and limit internal lateral movement. This
297 publication discusses ZTA, its logical components, possible deployment scenarios, and threats. It
298 also presents a general road map for organizations wishing to migrate to a zero trust design
299 approach to network infrastructure and discusses relevant federal policies that may impact or
300 influence a zero trust architecture strategy.

301 ZT is not a single-network architecture but a set of guiding principles in network infrastructure
302 and system design and operation that can be used to improve the security posture of any
303 classification or sensitivity level [FIPS199]. Transitioning to ZTA is a journey concerning how
304 an organization evaluates risk in its mission and cannot simply be accomplished with a wholesale
305 replacement of technology. That said, many organizations already have elements of a ZTA in
306 their enterprise infrastructure today. Organizations should seek to incrementally implement zero
307 trust principles, process changes, and technology solutions that protect their data assets and
308 business functions by use case. Most enterprise infrastructures will operate in a hybrid zero
309 trust/perimeter-based mode while continuing to invest in IT modernization initiatives and
310 improve organization business processes.

311 Organizations need to implement comprehensive information security and resiliency practices
312 for zero trust to be effective. When balanced with existing cybersecurity policies and guidance,
313 identity and access management, continuous monitoring, and best practices, a ZTA strategy can
314 protect against common threats and improve an organization’s security posture by using a
315 managed risk approach.

316 1.1 History of Zero Trust Efforts Related to Federal Agencies

317 The concept of zero trust has been present in cybersecurity since before the term “zero trust” was
318 coined. The Defense Information Systems Agency (DISA) and the Department of Defense
319 published their work on a more secure enterprise strategy dubbed “black core” [BCORE]. Black
320 core involved moving from a perimeter-based security model to one that focused on the security
321 of individual transactions. The work of the Jericho Forum in 1994 publicized the idea of de-
322 perimeterization—limiting implicit trust based on network location and the limitations of relying
323 on single, static defenses over a large network segment [JERICHO]. The concepts of de-
324 perimeterization evolved and improved into the larger concept of zero trust, which was later
325 coined by John Kindervag¹ while at Forrester.² Zero trust then became the term used to describe
326 various cybersecurity solutions that moved security away from implied trust based on network
327 location and instead focused on evaluating trust on a per-transaction basis. Both private industry
328 and higher education have also undergone this evolution from perimeter-based security to a
329 security strategy based on zero trust principles.

330 Federal agencies have been urged to move to security based on zero trust principles for more
331 than a decade, building capabilities and policies such as the Federal Information Security
332 Modernization Act (FISMA) followed by the Risk Management Framework (RMF); Federal
333 Identity, Credential, and Access Management (FICAM); Trusted Internet Connections (TIC);
334 and Continuous Diagnostics and Mitigation (CDM) programs. All of these programs aim to
335 restrict data and resource access to authorized parties. When these programs were started, they
336 were limited by the technical capabilities of information systems. Security policies were largely
337 static and were enforced at large “choke points” that an enterprise could control to get the largest
338 effect for the effort. As technology matures, it is becoming possible to continually analyze and
339 evaluate access requests in a dynamic and granular fashion to a “need to access” basis to mitigate
340 data exposure due to compromised accounts, attackers monitoring a network, and other threats.

341 1.2 Structure of This Document

342 The rest of the document is organized as follows:

- 343 • **Section 2** defines ZT and ZTA and lists some assumptions when designing a ZTA for an
344 enterprise. This section also includes a list of the tenets of ZT design.
- 345 • **Section 3** documents the logical components, or building blocks, of ZT. It is possible that
346 unique implementations make up ZTA components differently yet serve the same logical
347 functionality.
- 348 • **Section 4** lists some possible use cases where a ZTA may make enterprise environments
349 more secure and less prone to successful exploitation. These include enterprises with
350 remote employees, cloud services, and guest networks.

¹ <https://go.forrester.com/blogs/next-generation-access-and-zero-trust/>

² Any mention of commercial products or services within NIST documents is for information only; it does not imply a recommendation or endorsement by NIST.

- 351
- 352
- 353
- 354
- 355
- 356
- 357
- 358
- **Section 5** discusses threats to an enterprise using a ZTA. Many of these threats are similar to more traditionally architected networks but may require different mitigation techniques.
 - **Section 6** discusses how ZTA tenets fit into and/or complement existing guidance for federal agencies.
 - **Section 7** presents the starting point for transitioning an enterprise (such as a federal agency) to a ZTA. This includes a description of the general steps needed to plan and deploy applications and enterprise infrastructure that are guided by ZT tenets.
- 359

2 Zero Trust Basics

361 Zero trust is a cybersecurity paradigm focused on resource protection and the premise that trust
362 is never granted implicitly but must be continually evaluated. Zero trust architecture is an end-to-
363 end approach to enterprise resource and data security that encompasses identity (person and non-
364 person entities), credentials, access management, operations, endpoints, hosting environments,
365 and the interconnecting infrastructure. The initial focus should be on restricting resources to
366 those with a need to access and grant only the minimum privileges (e.g., read, write, delete)
367 needed to perform the mission. Traditionally, agencies (and enterprise networks in general) have
368 focused on perimeter defense, and authenticated users are given authorized access to a broad
369 collection of resources. As a result, unauthorized lateral movement within a network has been
370 one of the biggest challenges for federal agencies.

371 The TIC and agency perimeter firewalls provide strong internet gateways. This helps block
372 attackers from the internet, but the TICs and perimeter firewalls are less useful for detecting and
373 blocking attacks from inside the network and cannot protect users outside of the perimeter (e.g.,
374 remote workers, cloud-based services).

375 An operative definition of zero trust and zero trust architecture is as follows:

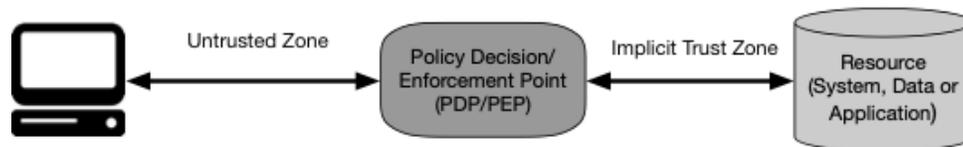
376 *Zero trust (ZT)* provides a collection of concepts and ideas designed to reduce the
377 uncertainty in enforcing accurate, per-request access decisions in information systems
378 and services in the face of a network viewed as compromised. *Zero trust architecture*
379 *(ZTA)* is an enterprise's cybersecurity plan that utilizes zero trust concepts and
380 encompasses component relationships, workflow planning, and access policies.
381 Therefore, a zero trust enterprise is the network infrastructure (physical and virtual) and
382 operational policies that are in place for an enterprise as a product of a zero trust
383 architecture plan.

384 An enterprise decides to adopt zero trust as its cybersecurity foundation and generate a zero trust
385 architecture as a plan developed with zero trust principles in mind. This plan is then deployed to
386 produce a zero trust environment for use in the enterprise.

387 This definition focuses on the crux of the issue, which is the goal to *prevent unauthorized access*
388 *to data and services* coupled with making the *access control enforcement as granular as*
389 *possible*. That is, authorized and approved subjects (combination of user, application, and
390 device) can access the data to the exclusion of all other subjects (i.e., attackers). To take this one
391 step further, the word "resource" can be substituted for "data" so that ZT and ZTA are about
392 resource access (e.g., printers, compute resources, Internet of Things [IoT] actuators) and not just
393 data access.

394 To lessen uncertainties (as they cannot be eliminated), the focus is on authentication,
395 authorization, and shrinking implicit trust zones while minimizing temporal delays in
396 authentication mechanisms. Access rules are restricted to least privilege and made as granular as
397 possible.

398 In the abstract model of access shown in Figure 1, a user or machine needs access to an
 399 enterprise resource. Access is granted through a policy decision point (PDP) and corresponding
 400 policy enforcement point (PEP).³



401

402

Figure 1: Zero Trust Access

403 The system must ensure that the user is authentic and the request is valid. The PDP/PEP passes
 404 proper judgment to allow the subject to access the resource. This implies that zero trust applies to
 405 two basic areas: authentication and authorization. What is the level of confidence about the
 406 user’s identity for this unique request? Is access to the resource allowable given the level of
 407 confidence in the user’s identity? Does the device used for the request have the proper security
 408 posture? Are there other factors that should be considered and that change the confidence level
 409 (e.g., time, location of subject, subject’s security posture)? Overall, enterprises need to develop
 410 and maintain dynamic risk-based policies for resource access and set up a system to ensure that
 411 these policies are enforced correctly and consistently. This means that an enterprise should not
 412 rely on implied trustworthiness wherein if the user has met a base authentication level (e.g.,
 413 logging into an asset), all resource requests are assumed to be equally valid.

414 The “implicit trust zone” represents an area where all the entities are trusted to at least the level
 415 of the last PDP/PEP gateway. For example, consider the passenger screening model in an airport.
 416 All passengers pass through the airport security checkpoint (PDP/PEP) to access the boarding
 417 gates. The passengers mill about in the terminal area, and all the cleared passengers are
 418 considered trusted. In this model, the implicit trust zone is the boarding area.

419 The PDP/PEP applies a set of controls so that all traffic beyond the PEP has a common level of
 420 trust. The PDP/PEP cannot apply additional policies beyond its location in the flow of traffic. To
 421 allow the PDP/PEP to be as specific as possible, the implicit trust zone must be as small as
 422 possible.

423 Zero trust provides a set of principles and concepts around moving the PDP/PEPs closer to the
 424 resource. The idea is to explicitly authenticate and authorize all users, devices, applications, and
 425 workflows that make up the enterprise.

³ Part of the concepts defined in OASIS XACML 2.0 https://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

426 2.1 Tenets of Zero Trust

427 Many definitions and discussions of ZT stress the concept of removing wide-area perimeter
428 defenses (e.g., enterprise firewalls) as a factor. However, most of these definitions continue to
429 define themselves in relation to perimeters in some way (such as micro-segmentation or micro-
430 perimeters; see Section 3.1) as part of the functional capabilities of a ZTA. The following is an
431 attempt to define ZT and ZTA in terms of basic tenets that should be involved rather than what is
432 excluded. These tenets are the ideal goal, though it must be acknowledged that not all tenets may
433 be fully implemented in their purest form for a given strategy.

434 A zero trust architecture is designed and deployed with adherence to the following zero trust
435 basic tenets:

- 436 1. **All data sources and computing services are considered resources.** A network may be
437 composed of several different classes of devices. A network may also have small
438 footprint devices that send data to aggregators/storage, software as a service (SaaS),
439 systems sending instructions to actuators, and other functions. Also, an enterprise may
440 decide to classify personally owned devices as resources if they can access enterprise-
441 owned resources.
- 442 2. **All communication is secured regardless of network location.** Network location does
443 not imply trust. Access requests from assets located on enterprise-owned network
444 infrastructure (e.g., inside a traditional network perimeter) must meet the same security
445 requirements as access requests and communication from any other nonenterprise-owned
446 network. In other words, trust should not be automatically granted based on the device
447 being on enterprise network infrastructure. All communication should be done in the
448 most secure manner available, protect confidentiality and integrity, and provide source
449 authentication.
- 450 3. **Access to individual enterprise resources is granted on a per-session basis.** Trust in
451 the requester is evaluated before the access is granted. This could mean only “sometime
452 previously” for this particular transaction and may not occur directly before initiating a
453 session or performing a transaction with a resource. However, authentication and
454 authorization to one resource will not automatically grant access to a different resource.
- 455 4. **Access to resources is determined by dynamic policy—including the observable state
456 of client identity, application, and the requesting asset—and may include other
457 behavioral attributes.** An organization protects resources by defining what resources it
458 has, who its members are (or ability to authenticate users from a federated community),
459 and what access to resources those members need. For zero trust, client identity includes
460 the user account and any associated attributes assigned by the enterprise to that account
461 or artifacts to authenticate automated tasks. Requesting asset state includes device
462 characteristics such as software versions installed, network location, time/date of request,
463 previously observed behavior, and installed credentials. Behavioral attributes include
464 automated user analytics, device analytics, and measured deviations from observed usage
465 patterns. Policy is the set of access rules based on attributes that an organization assigns
466 to a user, data asset, or application. These rules and attributes are based on the needs of
467 the business process and acceptable level of risk. Resource access and action permission

468 policies can vary based on the sensitivity of the resource/data. Least privilege principles
469 are applied to restrict both visibility and accessibility.

470 5. **The enterprise ensures that all owned and associated devices are in the most secure**
471 **state possible and monitors assets to ensure that they remain in the most secure state**
472 **possible.** No device is inherently trusted. Here, “most secure state possible” means that
473 the device is in the most practicable secure state and still performs the actions required
474 for the mission. An enterprise implementing a ZTA should establish a CDM or similar
475 system to monitor the state of devices and applications and should apply patches/fixes as
476 needed. Devices that are discovered to be subverted, have known vulnerabilities, and/or
477 are not managed by the enterprise may be treated differently (including denial of all
478 connections to enterprise resources) than devices owned by or associated with the
479 enterprise that are deemed to be in their most secure state. This may also apply to
480 associated devices (e.g., personally owned devices) that may be allowed to access some
481 resources but not others. This, too, requires a robust monitoring and reporting system in
482 place to provide actionable data about the current state of enterprise resources.

483 6. **All resource authentication and authorization are dynamic and strictly enforced**
484 **before access is allowed.** This is a constant cycle of obtaining access, scanning and
485 assessing threats, adapting, and continually reevaluating trust in ongoing communication.
486 An enterprise implementing a ZTA would be expected to have Identity, Credential, and
487 Access Management (ICAM) and asset management systems in place. This includes the
488 use of multifactor authentication (MFA) for access to some or all enterprise resources.
489 Continuous monitoring with possible reauthentication and reauthorization occurs
490 throughout user interaction, as defined and enforced by policy (e.g., time-based, new
491 resource requested, resource modification, anomalous user activity detected) that strives
492 to achieve a balance of security, availability, usability, and cost-efficiency.

493 7. **The enterprise collects as much information as possible about the current state of**
494 **network infrastructure and communications and uses it to improve its security**
495 **posture.** An enterprise should collect data about network traffic and access requests,
496 which is then used to improve policy creation and enforcement. This data can also be
497 used to provide context for access requests from subjects (see Section 3.3.1).
498

499 The above tenets attempt to be technology agnostic. For example, “user identification (ID)”
500 could include several factors such as username/password, certificates, and onetime password.
501 These tenets apply to work done within an organization or in collaboration with one or more
502 partner organizations and not to public or consumer-facing business processes. An organization
503 cannot impose internal policies on external actors (e.g., customers or general internet users).

504 2.2 A Zero Trust View of a Network

505 There are some basic assumptions for network connectivity for any organization that utilizes
506 ZTA in network planning and deployment. Some of these assumptions apply to enterprise-owned
507 network infrastructure, and some apply to enterprise-owned resources used on nonenterprise-
508 owned network infrastructure (e.g., public Wi-Fi). The network in an enterprise implementing a
509 ZTA should be developed with the ZTA tenets outlined above and with the following
510 assumptions.

511 2.2.1 Assumptions for Enterprise-Owned Network Infrastructure

- 512 1. **The entire enterprise private network is not considered an implicit trust zone.** Assets
513 should always act as if an attacker is present on the enterprise network, and
514 communication should be done in the most secure manner available (see tenet 2 above).
515 This entails actions such as authenticating all connections and encrypting all traffic.
- 516 2. **Devices on the network may not be owned or configurable by the enterprise.** Visitors
517 and/or contracted services may include nonenterprise-owned assets that need network
518 access to perform their role. This includes bring-your-own-device (BYOD) policies that
519 allow enterprise users to use nonenterprise-owned devices to access enterprise resources.
- 520 3. **No resource is inherently trusted.** Every asset must have its security posture evaluated
521 via a PEP before connecting to an enterprise-owned resource (similar to tenet 6 above for
522 assets as well as users). Enterprise-owned devices may have artifacts that enable
523 authentication and provide a confidence level higher than the same request coming from
524 nonenterprise-owned devices. User credentials alone are insufficient for device
525 authentication to an enterprise resource.

526 2.2.2 Assumptions for Nonenterprise-Owned Network Infrastructure

- 527 1. **Not all enterprise resources are on enterprise-owned infrastructure.** Resources
528 include remote enterprise users as well as cloud services. Enterprise-owned or -managed
529 assets may need to utilize the local (i.e., nonenterprise) network for basic connectivity
530 and network services (e.g., DNS resolution).
 - 531 2. **Remote enterprise users cannot fully trust the local network connection.** Remote
532 users should assume that the local (i.e., nonenterprise-owned) network is hostile. Assets
533 should assume that all traffic is being monitored and potentially modified. All connection
534 requests should be authenticated and authorized, and all communications should be done
535 in the most secure manner possible (i.e., provide confidentiality, integrity protection, and
536 source authentication). See the tenets of ZTA above.
- 537

3 Logical Components of Zero Trust Architecture

There are numerous logical components that make up a ZTA deployment in an enterprise. These components may be operated as an on-premises service or through a cloud-based service. The conceptual framework model in Figure 2 shows the basic relationship between the components and their interactions. Note that this is an ideal model showing logical components and their interactions. From Figure 1, the policy decision point (PDP) is broken down into two logical components: the policy engine and policy administrator (defined below). The ZTA logical components use a separate control plane to communicate, while application data is communicated on a data plane (see Section 3.4).

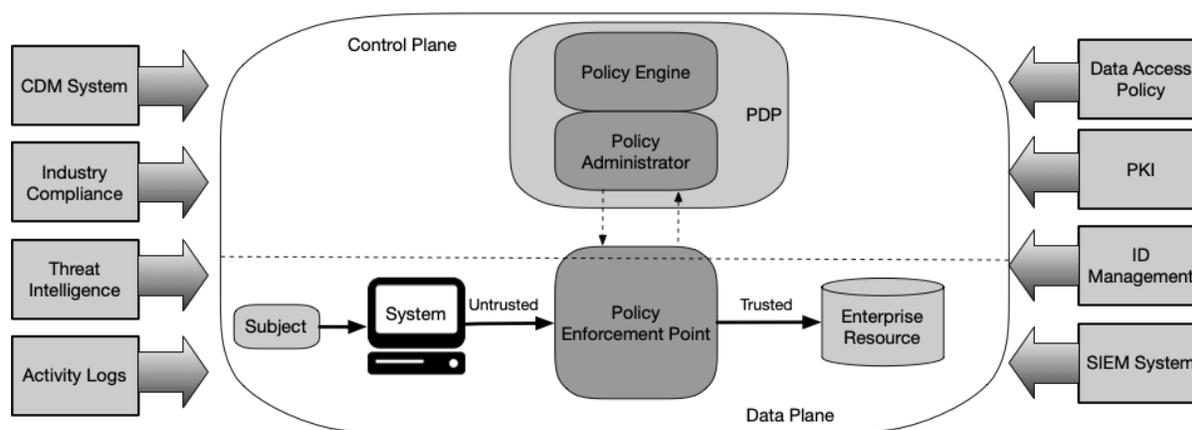


Figure 2: Core Zero Trust Logical Components

The component descriptions:

- Policy engine (PE):** This component is responsible for the ultimate decision to grant access to a resource for a given subject. The PE uses enterprise policy as well as input from external sources (e.g., CDM systems, threat intelligence services described below) as input to a trust algorithm (see Section 3.3 for more details) to grant, deny, or revoke access to the resource. The PE is paired with the policy administrator component. The policy engine makes and logs the decision, and the policy administrator executes the decision.
- Policy administrator (PA):** This component is responsible for establishing and/or shutting down the communication path between a subject and a resource. It would generate any authentication and authentication token or credential used by a client to access an enterprise resource. It is closely tied to the PE and relies on its decision to ultimately allow or deny a session. Some implementations may treat the PE and PA as a single service; here, it is divided into its two logical components. The PA communicates with the PEP when creating the communication path. This communication is done via the control plane.
- Policy enforcement point (PEP):** This system is responsible for enabling, monitoring, and eventually terminating connections between a subject and an enterprise resource.

568 This is a single logical component in ZTA but may be broken into two different
569 components: the client (e.g., agent on user's laptop) and resource side (e.g., gateway
570 component in front of resource that controls access) or a single portal component that acts
571 as a gatekeeper for communication paths. Beyond the PEP is the implicit trust zone (see
572 Section 2) hosting the enterprise resource.

573 In addition to the core components in an enterprise implementing a ZTA, several data sources
574 provide input and policy rules used by the policy engine when making access decisions. These
575 include local data sources as well as external (i.e., nonenterprise-controlled or -created) data
576 sources. These include:

- 577 • **Continuous diagnostics and mitigation (CDM) system:** This gathers information about
578 the enterprise asset's current state and applies updates to configuration and software
579 components. An enterprise CDM system provides the policy engine with the information
580 about the asset making an access request, such as whether it is running the appropriate
581 patched operating system (OS) and applications or whether the asset has any known
582 vulnerabilities.
- 583 • **Industry compliance system:** This ensures that the enterprise remains compliant with
584 any regulatory regime that it may fall under (e.g., FISMA, healthcare or financial
585 industry information security requirements). This includes all the policy rules that an
586 enterprise develops to ensure compliance.
- 587 • **Threat intelligence feed(s):** This provides information from internal or external sources
588 that help the policy engine make access decisions. These could be multiple services that
589 take data from internal and/or multiple external sources and provide information about
590 newly discovered attacks or vulnerabilities. This also includes blacklists, newly identified
591 malware, and reported attacks to other assets that the policy engine will want to deny
592 access to from enterprise assets.
- 593 • **Data access policies:** These are the attributes, rules, and policies about access to
594 enterprise resources. This set of rules could be encoded in or dynamically generated by
595 the policy engine. These policies are the starting point for authorizing access to a
596 resource as they provide the basic access privileges for accounts and applications in the
597 enterprise. These policies should be based on the defined mission roles and needs of the
598 organization.
- 599 • **Enterprise public key infrastructure (PKI):** This system is responsible for generating
600 and logging certificates issued by the enterprise to resources, subjects, and applications.
601 This also includes the global certificate authority ecosystem and the Federal PKI,⁴ which
602 may or may not be integrated with the enterprise PKI. This could also be a PKI that is not
603 built upon X.509 certificates.
- 604 • **ID management system:** This is responsible for creating, storing, and managing
605 enterprise user accounts and identity records (e.g., lightweight directory access protocol

⁴ <https://www.idmanagement.gov/topics/fpki/>

606 (LDAP) server). This system contains the necessary user information (e.g., name, email
607 address, certificates) and other enterprise characteristics such as role, access attributes,
608 and assigned assets. This system often utilizes other systems (such as a PKI) for artifacts
609 associated with user accounts. This system may be part of a larger federated community
610 and may include nonenterprise employees or links to nonenterprise assets for
611 collaboration.

- 612 • **Network and system activity logs:** This is the enterprise system that aggregates asset
613 logs, network traffic, resource access actions, and other events that provide real-time (or
614 near-real-time) feedback on the security posture of enterprise information systems.
- 615 • **Security information and event management (SIEM) system:** This collects security-
616 centric information for later analysis. This data is then used to refine policies and warn of
617 possible attacks against enterprise assets.

618 3.1 Variations of Zero Trust Architecture Approaches

619 There are several ways that an enterprise can enact a ZTA for workflows. These approaches vary
620 in the components used and in the main source of policy rules for an organization. Each
621 approach implements all the tenets of ZT (see Section 2.1) but may use one or two (or one
622 component) as the main driver of policies. The approaches include enhanced identity
623 governance–driven, logical micro-segmentation via next-generation firewalls (NGFWs), and
624 network-based segmentation.

625 Certain approaches lend themselves to some use cases more than others. An organization looking
626 to develop a ZTA for its enterprise may find that its chosen use case and existing policies point
627 to one approach over others. That does not mean the other approaches would not work but rather
628 that other approaches may be more difficult to implement and may require more fundamental
629 changes to how the enterprise currently conducts business flows.

630 3.1.1 ZTA Using Enhanced Identity Governance

631 The enhanced identity governance approach to developing a ZTA uses the identity of actors as
632 the key component of policy creation. If it were not for subjects requesting access to enterprise
633 resources, there would be no need to create access policies. For this approach, enterprise resource
634 access policies are based on identity and assigned attributes. The primary requirement for
635 resource access is based on the access privileges granted to the given subject. Other factors such
636 as device used, asset status, and environmental factors may alter the final confidence level
637 calculation (and ultimate access authorization) or tailor the result in some way, such as granting
638 only partial access to a given data source based on network location. Individual resources or PEP
639 components protecting the resource must have a way to forward requests to a policy engine
640 service or authenticate the subject and approve the request before granting access.

641 Enhanced identity governance-based approaches for enterprises are often found using an open
642 network model or an enterprise network with visitor access or frequent nonenterprise devices on
643 the network (such as with the use case in Section 4.3 below). Network access is initially granted
644 to all assets with access to resources that are restricted to identities with the appropriate access
645 privileges. The identity-driven approach works well with the resource portal model since device

646 identity and status provide secondary support data to access decisions. Other models work as
647 well, depending on policies in place.

648 **3.1.2 ZTA Using Micro-Segmentation**

649 An enterprise may choose to implement a ZTA based on placing individual or groups of
650 resources on its own network segment protected by a gateway security component. In this
651 approach, the enterprise places NGFWs or gateway devices to act as PEPs protecting each
652 resource or group of resources. These gateway devices dynamically grant access to individual
653 requests from a client asset. Depending on the model, the gateway may be the sole PEP
654 component or part of a multipart PEP consisting of the gateway and client-side agent (see
655 Section 3.2.1).

656 This approach applies to a variety of use cases and deployment models as the protecting device
657 acts as the PEP, with management of said devices acting as the PE/PA component. This
658 approach requires an identity governance program to fully function but relies on the gateway
659 components to act as the PEP that shields resources from unauthorized access and/or discovery.

660 The key necessity to this approach is that the PEP components are managed and should be able
661 to react and reconfigure as needed to respond to threats or change in the workflow. It is possible
662 to implement some features of a micro-segmented enterprise by using less advanced gateway
663 devices and even stateless firewalls, but the administration cost and difficulty to quickly adapt to
664 changes make this a very poor choice.

665 **3.1.3 ZTA Using Network Infrastructure and Software Defined Perimeters**

666 The third approach uses the network infrastructure to implement a ZTA. The ZT implementation
667 could be achieved by using an overlay network (i.e., layer 7 but also could be set up lower of the
668 ISO network stack). These approaches are sometimes referred to as software defined perimeter
669 (SDP) approaches and frequently include concepts from SDN [SDNBOOK] and intent-based
670 networking (IBN) [IBNVN]. In this approach, the PA acts as the network controller that sets up
671 and reconfigures the network based on the decisions made by the PE. The clients continue to
672 request access via PEPs, which are managed by the PA component.

673 When the approach is implemented at the application network layer (i.e., layer 7), the most
674 common deployment model is the agent/gateway (see Section 3.2.1). In this implementation, the
675 agent and resource gateway (acting as the single PEP and configured by the PA) establish a
676 secure channel used for communication between the client and resource.

677 **3.2 Deployed Variations of the Abstract Architecture**

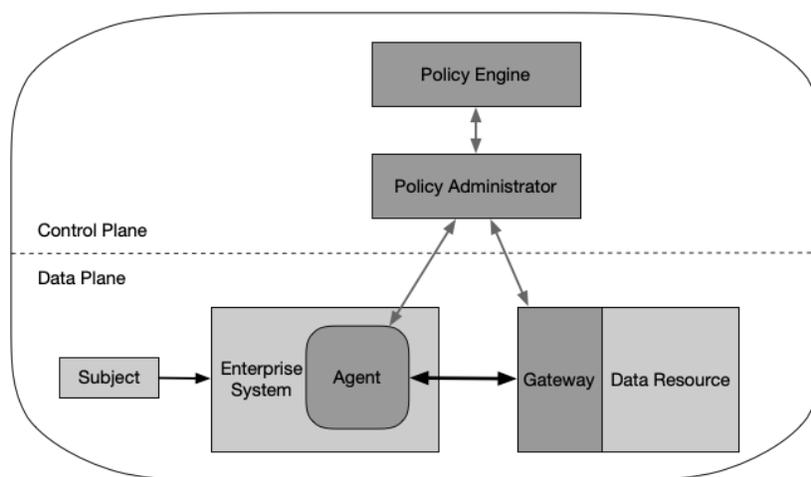
678 All of the above components are logical components. They do not necessarily need to be unique
679 systems. A single asset may perform the duties of multiple logical components, and likewise, a
680 logical component may consist of multiple hardware or software elements to perform the tasks.
681 For example, an enterprise-managed PKI may consist of one component responsible for issuing
682 certificates for devices and another used for issuing certificates to end users, but both use
683 intermediate certificates issued from the same enterprise root certificate authority. In some ZT
684 product offerings currently available on the market, the PE and PA components are combined in

685 a single service.

686 There are several variations on deployment of selected components of the architecture that are
 687 outlined in the sections below. Depending on how an enterprise network is set up, multiple ZTA
 688 deployment models may be in use for different business processes in one enterprise.

689 3.2.1 Device Agent/Gateway-Based Deployment

690 In this deployment model, the PEP is divided into two components that reside on the resource or
 691 as a component directly in front of a resource. For example, each enterprise-issued asset has an
 692 installed device agent that coordinates connections, and each resource has a component (i.e.,
 693 gateway) that is placed directly in front so that the resource communicates only with the
 694 gateway, essentially serving as a proxy for the resource. The gateway is responsible for
 695 connecting to the policy administrator and allows only approved communication paths
 696 configured by the policy administrator (see Figure 3).



697

698

Figure 3: Device Agent/Gateway Model

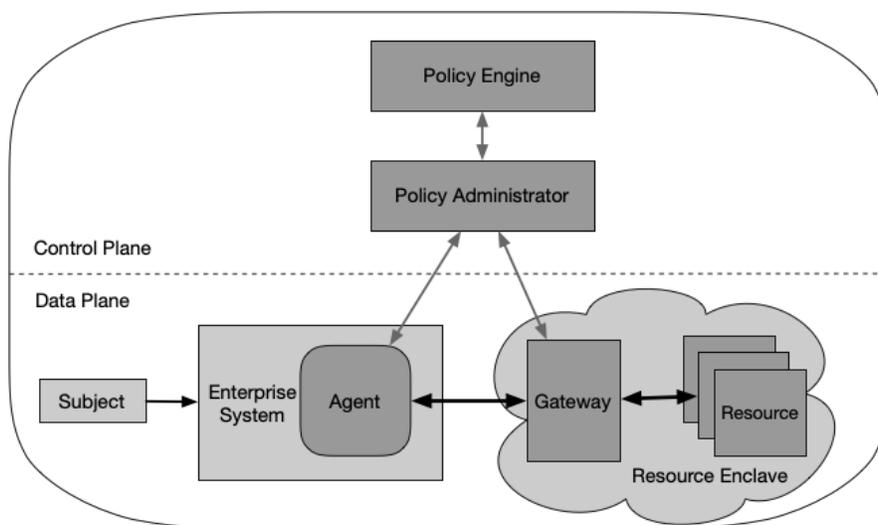
699 In a typical scenario, a user with an enterprise-issued laptop wishes to connect to an enterprise
 700 resource (e.g., human resources application/database). The access request is taken by the local
 701 agent, and the request is sent to the policy administrator. The policy administrator and policy
 702 engine could be an enterprise local asset or a cloud-hosted service. The policy administrator
 703 forwards the request to the policy engine for evaluation. If the request is authorized, the policy
 704 administrator configures a communication channel between the device agent and the relevant
 705 resource gateway via the control plane. This may include an internet protocol (IP) address, port
 706 information, session key, or similar security artifacts. The device agent and gateway then
 707 connect, and encrypted application data flows begin. The connection between the device agent
 708 and resource gateway is terminated when the workflow is completed or when triggered by the
 709 policy administrator due to a security event (e.g., session time-out, failure to reauthenticate).

710 This model is best utilized for enterprises that have a robust device management program in
 711 place as well as discrete resources that can communicate with the gateway. For enterprises that
 712 heavily utilize cloud services, this is a client-server implementation of the Cloud Security
 713 Alliance (CSA) Software Defined Perimeter (SDP) [CSA-SDP]. This model is also appropriate

714 for enterprises that do not want a BYOD policy in place. Access is possible only via the device
715 agent, which can be placed on enterprise-owned assets.

716 3.2.2 Enclave-Based Deployment

717 This deployment model is a variation of the device agent/gateway model above. In this model,
718 the gateway components may not reside on assets or in front of individual resources but instead
719 reside at the boundary of a resource enclave (e.g., on-location data center) as shown in Figure 4.
720 Usually, these resources serve a single business function or may not be able to communicate
721 directly to a gateway (e.g., legacy database system that does not have an application
722 programming interface [API] that can be used to communicate with a gateway). This deployment
723 model may also be useful for enterprises that use cloud-based micro-services for business
724 processes (e.g., user notification, database lookup, salary disbursement). In this model, the entire
725 private cloud is located behind a gateway.



726

727

Figure 4: Enclave Gateway Model

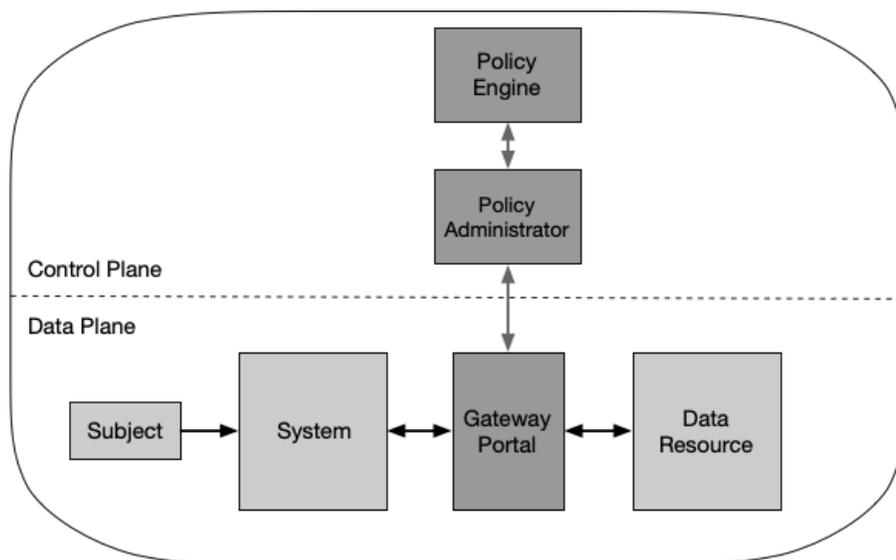
728 It is possible for this model to be a hybrid with the device agent/gateway model. In this model,
729 enterprise assets have a device agent that is used to connect to enclave gateways, but these
730 connections are created using the same process as the basic device agent/gateway model.

731 This model is useful for enterprises that have legacy applications or on-premises data centers that
732 cannot have individual gateways in place. The enterprise needs a robust asset and configuration
733 management program in place to install/configure the device agents. The downside is that the
734 gateway protects a collection of resources and may not be able to protect each resource
735 individually. This may also allow for subjects to see resources which they do not have privileges
736 to access.

737 3.2.3 Resource Portal-Based Deployment

738 In this deployment model, the PEP is a single component that acts as a gateway for user requests.
739 The gateway portal can be for an individual resource or a secure enclave for a collection of
740 resources used for a single business function. One example would be a gateway portal into a

741 private cloud or data center containing legacy applications as shown in Figure 5.



742

743

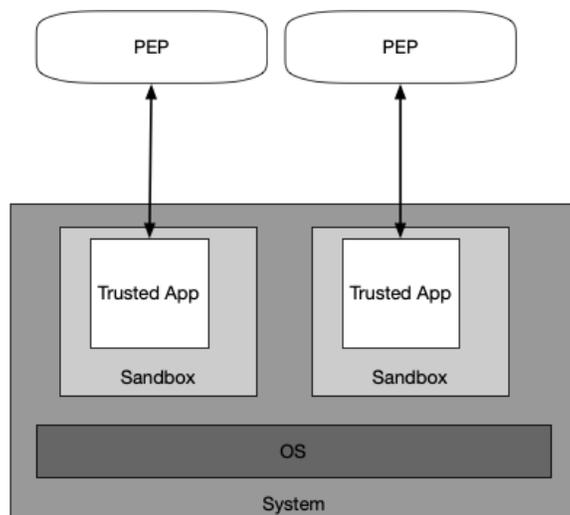
Figure 5: Resource Portal Model

744 The primary benefit of this model over the others is that a software component does not need to
 745 be installed on all client devices. This model is also more flexible for BYOD policies and inter-
 746 organizational collaboration projects. Enterprise administrators do not need to ensure that each
 747 device has the appropriate device agent before use. However, limited information can be inferred
 748 from devices requesting access. This model can only scan and analyze assets and devices once
 749 they connect to the PEP portal and may not be able to continuously monitor them for malware
 750 and appropriate configuration.

751 The main difference with this model is that there is no local agent that handles requests, and so
 752 the enterprise may not have full visibility or arbitrary control over assets as it can only see/scan
 753 them when they connect to a portal. The enterprise may be able to employ measures such as
 754 browser isolation to mitigate or compensate. These assets may be invisible to the enterprise
 755 between these sessions. This model also allows attackers to discover and attempt to access the
 756 portal or attempt a denial-of-service (DoS) attack against the portal. The portal systems should
 757 be well-provisioned to provide availability against a DoS attack or network disruption.

758 **3.2.4 Device Application Sandboxing**

759 Another variation of the agent/gateway deployment model is having vetted applications or
 760 processes run compartmentalized on assets. These compartments could be virtual machines,
 761 containers, or some other implementation, but the goal is the same: to protect the application or
 762 instances of applications from a possibly compromised host or other applications running on the
 763 asset.



764

765

Figure 6: Application Sandboxes

766 In Figure 6, the user device runs approved, vetted applications in a sandbox. The applications can
 767 communicate with the PEP to request access to resources, but the PEP will refuse requests from
 768 other applications on the asset. The PEP could be an enterprise local service or a cloud service in
 769 this model.

770 The main advantage of this model variant is that individual applications are segmented from the
 771 rest of the asset. If the asset cannot be scanned for vulnerabilities, these individual, sandboxed
 772 applications may be protected from a potential malware infection on the host asset. One of the
 773 disadvantages of this model is that enterprises must maintain these sandboxed applications for all
 774 assets and may not have full visibility into client assets. The enterprise also needs to make sure
 775 each sandboxed application is secure, which may require more effort than simply monitoring
 776 devices.

777 3.3 Trust Algorithm

778 For an enterprise with a ZTA deployment, the policy engine can be thought of as the brain and
 779 the PE's trust algorithm (TA) as its primary thought process. The TA is the process used by the
 780 policy engine to ultimately grant or deny access to a resource. The policy engine takes input
 781 from multiple sources: the policy database with information about users, user attributes and
 782 roles, historical user behavior patterns, threat intelligence sources, and other metadata sources.
 783 The process can be visualized in Figure 7.

784

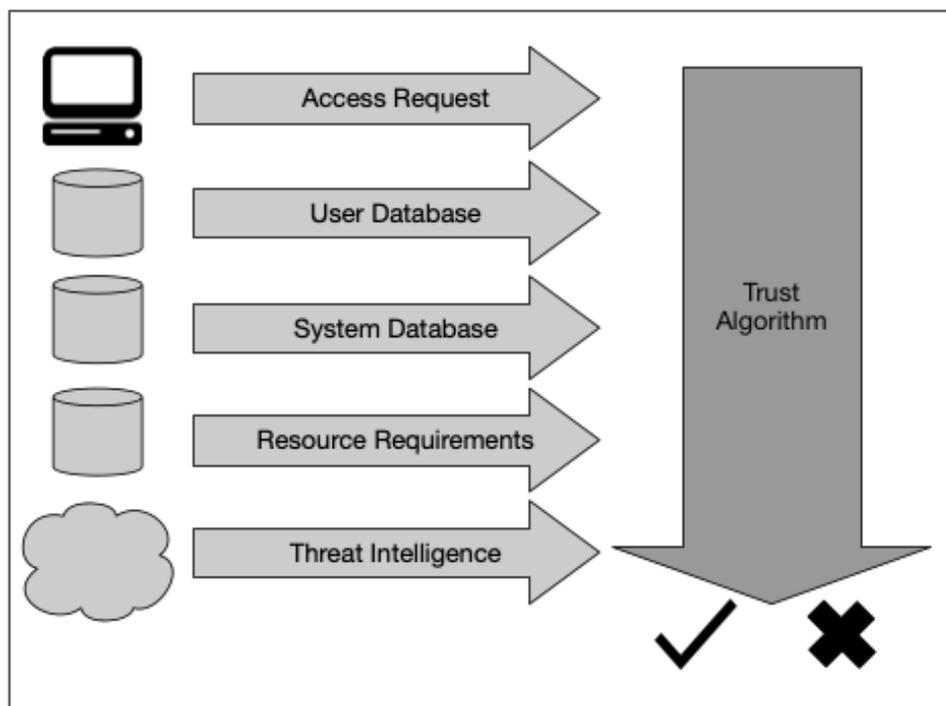


Figure 7: Trust Algorithm Input

785
786

787 In the figure, the inputs can be broken into categories based on what they provide to the trust
788 algorithm.

- 789
- 790 • **Access request:** This is the actual request from the subject. The resource requested is the
791 primary information used, but information about the requester is also used. This can
792 include OS version, application used, and patch level. Depending on these factors and the
asset security posture, access to assets might be restricted or denied.
 - 793 • **User identification, attributes, and privileges:** This is the “who” that is requesting
794 access to a resource [SP800-63-3]. This is the set of users (human and processes) of the
795 enterprise or collaborators and a collection of user attributes/privileges assigned. These
796 users and attributes form the basis of policies for resource access [SP800-162] [NISTIR
797 7987]. User identities can include a mix of logical identity (e.g., account ID) and results
798 of authentication checks performed by PEPs. Attributes of identity that can be factored
799 into deriving the confidence level include time and geolocation. A collection of privileges
800 given to multiple users could be thought of as a role, but privileges should be assigned to
801 a user on an individual basis and not simply because they may fit into a particular role.
802 This collection should be encoded and stored in an ID management system and policy
803 database. This may also include data about past observed user behavior in some (TA)
804 variants (see Section 3.3.1).
 - 805 • **Asset database and observable status:** This is the database that contains the known
806 status of each enterprise-owned asset (physical and virtual, to some extent). This is
807 compared to the observable status of the asset making the request and can include OS
808 version, application used, location (network location and geolocation), and patch level.

809 Depending on the asset state compared with this database, access to assets might be
810 restricted or denied.

811 • **Resource access requirements:** This set of policies complements the user ID and
812 attributes database [SP800-63-3] and defines the minimal requirements for access to the
813 resource. Requirements may include authenticator assurance levels, such as MFA
814 network location (e.g., deny access from overseas IP addresses), data sensitivity
815 (sometimes referred to as “data toxicity”), and requests for asset configuration. These
816 requirements should be developed by both the data custodian (i.e., those responsible for
817 the data) and those responsible for the business processes that utilize the data (i.e., those
818 responsible for the mission).

819 • **Threat intelligence:** This is an information feed or feeds about general threats and active
820 malware operating on the internet. These feeds can be external services or internal scans
821 and discoveries and can include attack signatures and mitigations. This is the only
822 component that will most likely be under the control of a service rather than the
823 enterprise.

824 The weight of importance for each data source may be a proprietary algorithm or may be
825 configured by the enterprise. These weight values can be used to reflect the importance of the
826 data source to an enterprise.

827 The final determination is then passed to the PA for execution. The PA’s job is to configure the
828 necessary PEPs to enable authorized communication. Depending on how the ZTA is deployed,
829 this may involve sending authentication results and connection configuration information to
830 gateways and agents or resource portals. PAs may also place a hold or pause on a
831 communication session to reauthenticate and reauthorize the connection in accordance with
832 policy requirements. The PA is also responsible for issuing the command to terminate the
833 connection based on policy (e.g., after a time-out, when the workflow has been completed, due to
834 a security alert).

835 3.3.1 Trust Algorithm Variations

836 There are different ways to implement a TA. Different implementers may wish to weigh the
837 above factors differently according to the factors’ perceived importance. There are two other
838 major characteristics that can be used to differentiate TAs. The first is how the factors are
839 evaluated, whether as binary decisions or weighted parts of a whole “score” or confidence level.
840 The second is how requests are evaluated in relation to other requests by the same subject,
841 application, or device.

842 • **Criteria- versus score-based:** A criteria-based TA assumes a set of qualified attributes
843 that must be met before access is granted to a resource or an action (e.g., read/write) is
844 allowed. These criteria are configured by the enterprise and should be independently
845 configured for every resource. Access is granted or an action applied to a resource only if
846 all the criteria are met. A score-based TA computes a confidence level based on values
847 for every data source and enterprise-configured weights. If the score is greater than the
848 configured threshold value for the resource, access is granted, or the action is performed.

849 Otherwise, the request is denied, or access privileges are reduced (e.g., read access is
850 granted but not write access for a file).

- 851 • **Singular versus contextual:** A singular TA treats each request individually and does not
852 take the user/application history into consideration when making its evaluation. This can
853 allow faster evaluations, but there is a risk that an attack can go undetected if it stays
854 within a user’s allowed role. A contextual TA takes a user or network agent’s recent
855 history into consideration when evaluating access requests. This means the PE must
856 maintain some state information on all users and applications but may be more likely to
857 detect an attacker using subverted credentials to access information in a pattern that is
858 atypical of what the PE sees for the given subject. Analysis of user behavior can be used
859 to provide a model of acceptable use, and deviations from this behavior could trigger
860 additional authentication checks or resource request denials.

861 The two factors are not always dependent on each other. It is possible to have a TA that assigns a
862 confidence level to every user and/or device and still considers every access request
863 independently (i.e., singular). However, contextual, score-based TAs work best, since the score
864 provides a current confidence level for the requesting account.

865 Ideally, a ZTA trust algorithm should be contextual, but this may not always be possible with the
866 infrastructure components available to the enterprise. A contextual TA can mitigate threats
867 where an attacker stays close to a “normal” set of access requests for a compromised user
868 account or insider attack. It is important to balance security, usability, and cost-effectiveness
869 when defining and implementing trust algorithms. Continually prompting a user for
870 reauthentication against behavior that is consistent with historical trends and norms for their
871 mission function and role within the organization can lead to usability issues. For example, if an
872 employee in the HR department of an agency normally accesses 20 to 30 employee records in a
873 typical workday, a contextual TA may send an alert if the access requests suddenly exceed 100
874 records in a day. A contextual TA may also send an alert if someone is making access requests
875 after normal business hours as this could be an attacker exfiltrating records by using a
876 compromised HR account. These are examples where a contextual TA can detect an attack
877 whereas a singular TA may fail to detect the new behavior. In another example, an accountant
878 who typically accesses the financial system during normal business hours is now trying to access
879 the system in the middle of the night from an unrecognizable location. A contextual TA may
880 trigger an alert and require the user to satisfy a more stringent confidence level or other criteria
881 as outlined in NIST Special Publication 800-63A [SP800-63A].

882 Developing a set of criteria or weights/threshold values for each resource requires planning and
883 testing. Enterprise administrators may encounter issues during the initial implementation of ZTA
884 where access requests that should be approved are denied due to misconfiguration. This will
885 result in an initial “tuning” phase of deployment. Criteria or scoring weights may need to be
886 adjusted to ensure that the policies are enforced while still allowing the enterprise’s business
887 processes to function. How long this tuning phase lasts depends on the enterprise-defined metrics
888 for progress and tolerance for incorrect access denials/approvals for the resources used in the
889 workflow.

890 3.4 Network/Environment Components

891 In a ZT environment, there should be a separation (logical or possibly physical) of the
892 communication flows used to control and configure the network and application communication
893 flows used to perform the actual work of the organization. This is often broken down to a *control*
894 *plane* for network control communication and a *data plane* for application communication flows
895 [Gilman].

896 The control plane is used by various infrastructure components (both enterprise-owned and from
897 service providers) to maintain assets; judge, grant, or deny access to resources; and perform any
898 necessary operations to set up communication paths between resources. The data plane is used
899 for actual communication between applications. This communication channel may not be
900 possible before the path has been established via the control plane. For example, the control
901 plane could be used by the PA and PEP to set up the communication path between the user and
902 the enterprise resource. The application workload would then use the data plane path that was
903 established.

904 3.4.1 Network Requirements to Support ZTA

- 905 1. **Enterprise assets have basic network connectivity.** The local area network (LAN),
906 enterprise controlled or not, provides basic routing and infrastructure (e.g., DNS). The
907 remote enterprise asset may not necessarily use all infrastructure services.
- 908 2. **The enterprise must be able to distinguish between what assets are owned or**
909 **managed by the enterprise and their current security posture.** This is determined by
910 enterprise-issued credentials and not unauthenticated information (e.g., network MAC
911 addresses that can be spoofed).
- 912 3. **The enterprise can capture all network traffic.** The enterprise can record packets seen
913 on the data plane but may not be able to perform application layer inspection (i.e., ISO
914 layer 7) on all packets. The enterprise can filter out metadata about the connection (e.g.,
915 destination, time, device identity) to dynamically update policies and inform the PE in
916 evaluating access requests.
- 917 4. **Enterprise resources should not be reachable without accessing a PEP.** Enterprise
918 resources do not accept arbitrary incoming connections from the internet. Resources
919 accept custom-configured connections only after a client has been authenticated and
920 authorized. These communication paths are set up by the PEP. Resources may not even
921 be discoverable without accessing a PEP. This prevents attackers from identifying targets
922 via scanning and launching DoS attacks against resources located behind PEPs. Note that
923 not all resources should be hidden this way; some network infrastructure components
924 (e.g., DNS servers) must be accessible.
- 925 5. **The data plane and control plane are logically separate.** The policy engine, policy
926 administrator, and PEPs communicate on a network that is logically separate and not
927 directly accessible by enterprise assets and resources. The data plane is used for
928 application data traffic. The policy engine, policy administrator, and PEPs use the control

- 929 plane to communicate and manage communication paths between assets. The PEPs must
930 be able to send and receive messages from both the data and control planes.
- 931 6. **Enterprise assets can reach the PEP component.** Enterprise users must be able to
932 access the PEP component to gain access to resources. This could take the form of a web
933 portal, network device, or software agent on the enterprise asset that enables the
934 connection.
- 935 7. **The PEP is the only component that accesses the policy administrator as part of a
936 business flow.** Each PEP operating on the enterprise network has a connection to the
937 policy administrator to establish communication paths from clients to resources. All
938 enterprise business process traffic passes through one or more PEPs.
- 939 8. **Remote enterprise assets should be able to access enterprise resources without
940 needing to traverse enterprise network infrastructure first.** For example, a remote
941 user should not be required to use a link back to the enterprise network (i.e., virtual
942 private network [VPN]) to access services utilized by the enterprise and hosted by a
943 public cloud provider (e.g., email).
- 944 9. **The infrastructure used to support the ZTA access decision process should be made
945 scalable to account for changes in process load.** The PE(s), PA(s), and PEPs used in a
946 ZTA become the key components in any business process. Delay or inability to reach a
947 PEP (or inability of the PEPs to reach the PA/PE) negatively impacts the ability to
948 perform the workflow. An enterprise implementing a ZTA needs to provision the
949 components for the expected workload or be able to rapidly scale the infrastructure to
950 handle increased usage when needed.
- 951 10. **Enterprise assets may not be able to reach certain PEPs due to observable factors.**
952 For example, there may be a policy stating that mobile assets may not be able to reach
953 certain resources if the requesting asset is located outside of the enterprise's home
954 country. These factors could be based on location (geolocation or network location),
955 device type, or other criteria.

956 **4 Deployment Scenarios/Use Cases**

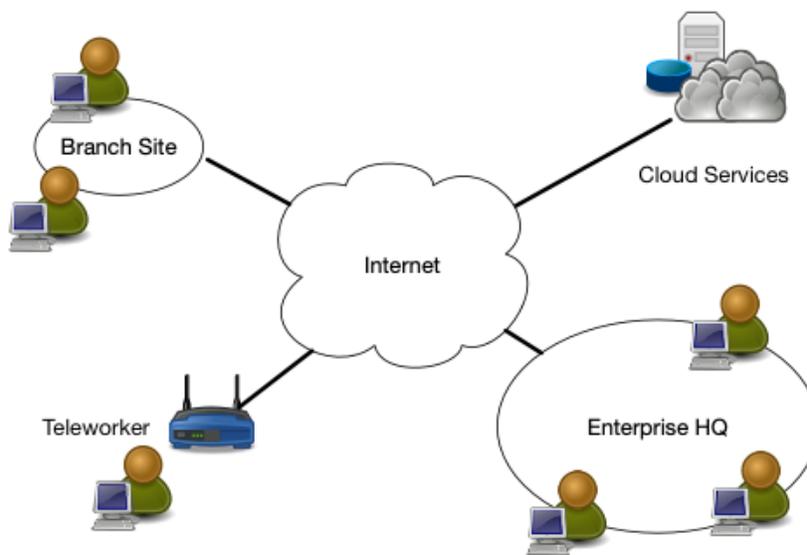
957 Any enterprise environment can be designed with zero trust tenets in mind. Most organizations
958 already have some elements of zero trust in their enterprise infrastructure or are on their way
959 through implementation of information security and resiliency policies and best practices.
960 Several deployment scenarios and use cases lend themselves readily to a zero trust architecture.
961 For instance, ZTA has its roots in organizations that are geographically distributed and/or have a
962 highly mobile workforce. That said, any organization can benefit from a zero trust architecture.

963 In the use cases below, ZTA is not explicitly indicated since the enterprise likely has both
964 perimeter-based and possibly ZTA infrastructures. As discussed in Section 7.2, there will likely
965 be a period when ZTA components and perimeter-based network infrastructure are concurrently
966 in operation in an enterprise.

967 **4.1 Enterprise with Satellite Facilities**

968 The most common scenario involves an enterprise with a single headquarters and one or more
969 geographically dispersed locations that are not joined by an enterprise-owned physical network
970 connection (see Figure 8). Employees at the remote location may not have a full enterprise-
971 owned local network but still need to access enterprise resources to perform their tasks.
972 Likewise, employees may be teleworking or in a remote location and using enterprise-owned or
973 personally-owned devices. In such cases, an enterprise may wish to grant access to some
974 resources (e.g., employee calendar, email) but deny access or restrict actions to more sensitive
975 resources (e.g., HR database).

976 In this use case, the PE/PA(s) is often hosted as a cloud service (which usually provides superior
977 availability and would not require remote workers to rely on enterprise infrastructure to access
978 cloud resources) with end assets having an installed agent (see Section 3.2.1) or accessing a
979 resource portal (see Section 3.2.3). It may not be most responsive to have the PE/PA(s) hosted on
980 the enterprise local network as remote offices and workers must send all traffic back to the
981 enterprise network to reach applications hosted by cloud services.



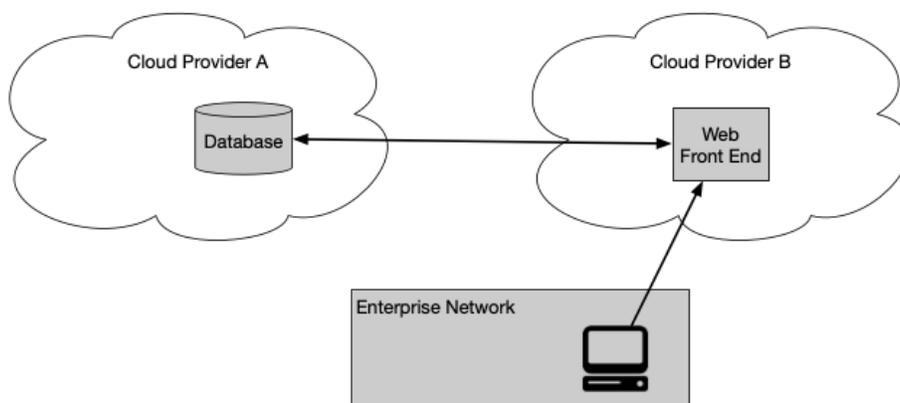
982

983

Figure 8: Enterprise with Remote Employees

984 **4.2 Multi-cloud Enterprise**

985 One increasingly common use case for deploying a ZTA is an enterprise utilizing multiple cloud
 986 providers (see Figure 9). In this use case, the enterprise has a local network but uses two or more
 987 cloud service providers to host applications and data. Sometimes, the application is hosted on a
 988 cloud service that is separate from the data source. For performance and ease of management, the
 989 application hosted in Cloud Provider A should be able to connect directly to the data source
 990 hosted in Cloud Provider B rather than force the application to tunnel back through the enterprise
 991 network.



992

993

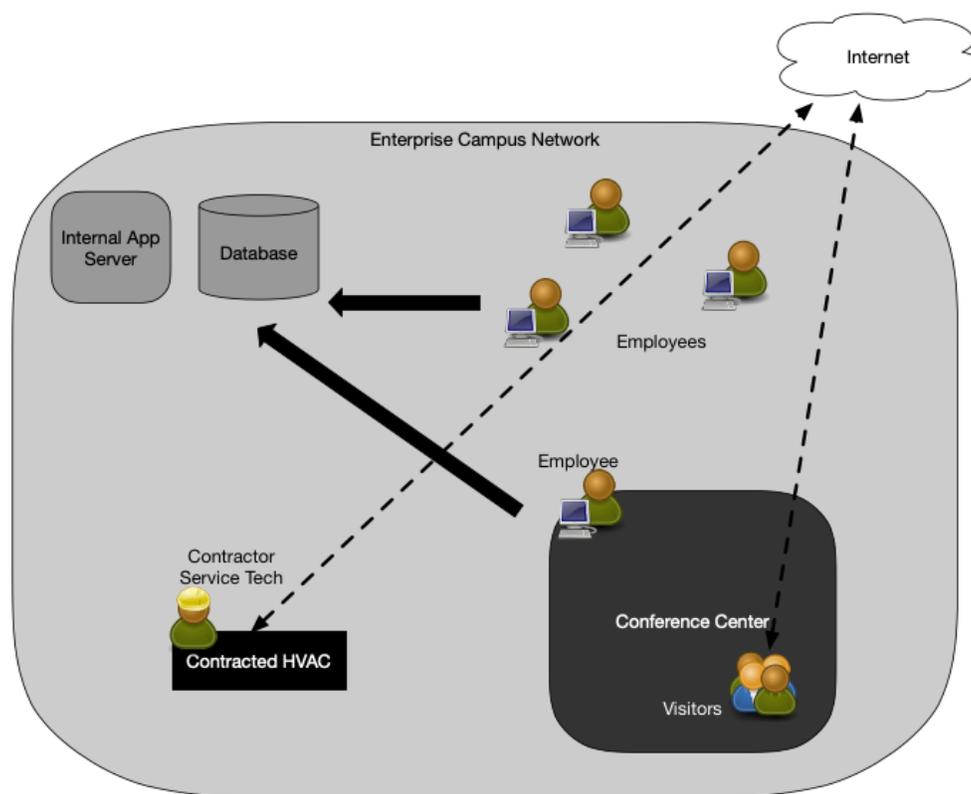
Figure 9: Multi-cloud Use Case

994 This use case is the server-server implementation of the CSA’s SDP specification [CSA-SDP].
 995 As enterprises move to more cloud-hosted applications and services, it becomes apparent that
 996 relying on the enterprise perimeter for security becomes a liability. As discussed in Section 2.2,
 997 ZT principles take the view that there should be no difference between enterprise-owned and -
 998 operated network infrastructure and infrastructure owned and operated by any other service
 999 provider. The zero trust approach to multi-cloud use is to place PEPs at the access points of each

1000 application and data source. The PE and PA may be services located in either cloud or even on a
 1001 third cloud provider. The client (via a portal or local installed agent) then accesses the PEPs
 1002 directly. That way, the enterprise can still manage access to resources even when hosted outside
 1003 the enterprise.

1004 4.3 Enterprise with Contracted Services and/or Nonemployee Access

1005 Another common scenario is an enterprise that includes on-site visitors and/or contracted service
 1006 providers that require limited access to enterprise resources to do their work (see Figure 10). For
 1007 example, an enterprise has its own internal applications, databases, and assets. These include
 1008 services contracted out to providers who may occasionally be on-site to provide maintenance
 1009 (e.g., smart heating and lighting systems that are owned and managed by external providers).
 1010 These visitors and service providers will need network connectivity to perform their tasks. A
 1011 zero trust enterprise could facilitate this by allowing these devices and any visiting service
 1012 technician access to the internet while obscuring enterprise resources.



1013

1014

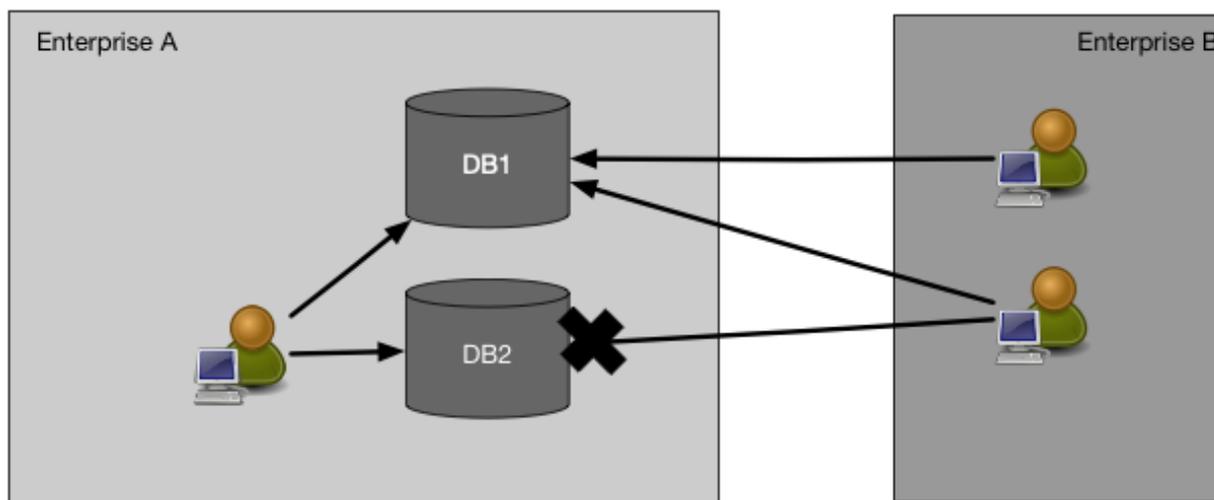
Figure 10: Enterprise with Nonemployee Access

1015 In this example, the organization also has a conference center where visitors interact with
 1016 employees. Again, with a ZTA approach of SDPs, employee devices and users are differentiated
 1017 and may be able to access appropriate enterprise resources. Visitors to the campus can have
 1018 internet access but cannot access enterprise resources. They may not even be able to discover
 1019 enterprise services via network scans (i.e., prevent active network reconnaissance/east-west
 1020 movement).

1021 In this use case, the PE(s) and PA(s) could be hosted as a cloud service or on the LAN (assuming
 1022 little or no use of cloud-hosted services). The enterprise assets could have an installed agent (see
 1023 Section 3.2.1) or access resources via a portal (see Section 3.2.3). The PA(s) ensures that all
 1024 nonenterprise assets (those that do not have installed agents or cannot connect to a portal) cannot
 1025 access local resources but may access the internet.

1026 4.4 Collaboration Across Enterprise Boundaries

1027 A fourth use case is cross-enterprise collaboration. For example, there is a project involving
 1028 employees from Enterprise A and Enterprise B (see Figure 11). The two enterprises may be
 1029 separate federal agencies (G2G) or even a federal agency and a private enterprise (G2B).
 1030 Enterprise A operates the database used for the project but must allow access to the data for
 1031 certain members of Enterprise B. Enterprise A can set up specialized accounts for the employees
 1032 of Enterprise B to access the required data and deny access to all other resources, but this can
 1033 quickly become difficult to manage. Having both organizations enrolled in a federated ID
 1034 management system would allow quicker establishment of these relationships provided that both
 1035 organizations' PEPs can authenticate subjects in a federated ID community.



1036

1037

Figure 11: Cross-Enterprise Collaboration

1038 This scenario can be similar to Use Case 1 (Section 4.1) as employees of both enterprises may
 1039 not be located on their organizations' network infrastructures, and the resource they need to
 1040 access may be within one enterprise environment or hosted in the cloud. This means that there do
 1041 not need to be complex firewall rules or enterprise-wide access control lists (ACLs) allowing
 1042 certain IP addresses belonging to Enterprise B to access resources in Enterprise A. How this
 1043 access is accomplished depends on the technology in use. Similar to Use Case 1, a PE and PA
 1044 hosted as a cloud service may provide availability to all parties without having to establish a
 1045 VPN or similar. The employees of Enterprise B may be asked to install a software agent on their
 1046 asset or access the necessary data resources through a web gateway (see Section 3.2.3).

1047 **4.5 Enterprise with Public- or Customer-Facing Services**

1048 A common feature in many enterprises is a public-facing service that may or may not include
1049 user registration (i.e., users must create or have been issued a set of login credentials). Such
1050 services could be for the general public, a set of customers with an existing business relationship,
1051 or a special set of nonenterprise users such as employee dependents. In all cases, it is likely that
1052 requesting assets are not enterprise-owned, and the enterprise is constrained as to what internal
1053 cybersecurity policies can be enforced.

1054 For a general, public-facing resource that does not require login credentials to access (e.g., public
1055 web page), the tenets of ZTA do not directly apply. The enterprise cannot strictly control the
1056 state of requesting assets, and public resources do not require credentials in order to be accessed.

1057 Enterprises may establish policies for registered public users such as customers (i.e., those with a
1058 business relationship) and special users (e.g., employee dependents). If the users are required to
1059 produce or are issued credentials, the enterprise may institute policies regarding password length,
1060 life cycle, and other details and may provide MFA as an option or requirement. However,
1061 enterprises are limited in the policies they can implement for this class of user. Information about
1062 incoming requests may be useful in determining the state of the public service and detecting
1063 possible attacks masquerading as legitimate users. For example, a registered user portal is known
1064 to be accessed by registered customers using one of a set of common web browsers. A sudden
1065 increase in access requests from unknown browser types or known outdated versions could
1066 indicate an automated attack of some kind, and the enterprise could take steps to limit requests
1067 from these identified clients. The enterprise should also be aware of any statutes or regulations
1068 regarding what information can be collected and recorded about the requesting users and assets.

1069 **5 Threats Associated with Zero Trust Architecture**

1070 No enterprise can eliminate cybersecurity risk. When complemented with existing cybersecurity
1071 policies and guidance, identity and access management, continuous monitoring, and general
1072 cyber hygiene, ZTA can reduce overall risk exposure and protect against common threats.
1073 However, some threats have unique features when implementing a ZTA.

1074 **5.1 Subversion of ZTA Decision Process**

1075 In ZTA, the policy engine and policy administrator are the key components of the entire
1076 enterprise. No communication between enterprise resources occurs unless it is approved and
1077 possibly configured by the PE and PA. This means that these components must be properly
1078 configured and maintained. Any enterprise administrator with configuration access to the PE's
1079 rules may be able to perform unapproved changes or make mistakes that can disrupt enterprise
1080 operations. Likewise, a compromised PA could allow access to resources that would otherwise
1081 not be approved (e.g., to a subverted, personally-owned device). Mitigating associated risks
1082 means that the PE and PA components must be properly configured and monitored, and any
1083 configuration changes must be logged and subject to audit.

1084 **5.2 Denial-of-Service or Network Disruption**

1085 In ZTA, the PA is the key component for resource access. Enterprise resources cannot connect to
1086 each other without the PA's permission and, possibly, configuration action. If an attacker
1087 disrupts or denies access to the PEP(s) or PA (i.e., DoS attack or route hijack), it can adversely
1088 impact enterprise operations. Enterprises can mitigate this threat by having the policy
1089 enforcement reside in a cloud or be replicated in several locations following guidance on cyber
1090 resiliency [SP 800-160].

1091 This mitigates the risk but does not eliminate it. Botnets such as Mirai produce massive DoS
1092 attacks against key internet service providers and disrupt service to millions of internet users.⁵ It
1093 is also possible that an attacker could intercept and block traffic to a PEP or PA from a portion or
1094 all of the user accounts within an enterprise (e.g., a branch office or even a single remote
1095 employee). In such cases, only a portion of enterprise users is affected. This is also possible in
1096 traditional VPN-based access and is not unique to ZTA.

1097 A hosting provider may also accidentally take a cloud-based PE or PA offline. Cloud services
1098 have experienced disruptions in the past, both infrastructure as a service⁶ and SaaS.⁷ An
1099 operational error could prevent an entire enterprise from functioning if the policy engine or
1100 policy administrator component becomes inaccessible from the network.

1101 There is also the risk that enterprise resources may not be reachable from the PA, so even if

⁵ <https://blog.cloudflare.com/inside-mirai-the-infamous-iot-botnet-a-retrospective-analysis/>

⁶ <https://aws.amazon.com/message/41926/>

⁷ https://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=12286870

1102 access is granted to a user, the PA cannot configure the communication path from the network.
1103 This could happen due to an attack or simply due to unexpected heavy usage. This is similar to
1104 any other network disruption in that some or all enterprise users cannot access a particular
1105 resource due to that resource not being available for some reason.

1106 **5.3 Stolen Credentials/Insider Threat**

1107 Properly implemented ZT, information security and resiliency policies, and best practices reduce
1108 the risk of an attacker gaining broad access via stolen credentials or insider attack. The ZT
1109 principle of no implicit trust based on network location means attackers need to compromise an
1110 existing account or device to gain a foothold in an enterprise. A properly implemented ZTA
1111 should prevent a compromised account or asset from accessing resources outside its normal
1112 purview or access patterns. This means that accounts with access policies around resources that
1113 an attacker is interested in would be the primary targets for attackers.

1114 Attackers may use phishing, social engineering, or a combination of attacks to obtain credentials
1115 of valuable accounts. “Valuable” may mean different things based on the attacker’s motivation.
1116 For instance, enterprise administrator accounts may be valuable, but attackers interested in
1117 financial gain may consider accounts that have access to financial or payment resources of equal
1118 value. Implementation of MFA for network access may reduce the risk of access from a
1119 compromised account. However, just like traditional enterprises, an attacker with valid
1120 credentials (or a malicious insider) may still be able to access resources for which the account
1121 has been granted access. For example, an attacker or compromised employee who has the
1122 credentials and enterprise-owned asset of a valid human resources employee may still be able to
1123 access an employee database.

1124 ZTA increases resistance to this attack and prevents any compromised accounts or assets from
1125 moving laterally throughout the network. If the compromised credentials are not authorized to
1126 access a particular resource, they will continue to be denied access to that resource. In addition, a
1127 contextual trust algorithm (see Section 3.3.1) is more likely to detect and respond quickly to this
1128 attack than when occurring in a traditional, perimeter-based network. The contextual TA can
1129 detect access patterns that are out of normal behavior and deny the compromised account or
1130 insider threat access to sensitive resources.

1131 **5.4 Visibility on the Network**

1132 As mentioned in Section 3.4.1, all traffic is inspected and logged on the network and analyzed to
1133 identify and react to potential attacks against the enterprise. However, as also mentioned, some
1134 (possibly the majority) of the traffic on the enterprise network may be opaque to traditional layer
1135 3 network analysis tools. This traffic may originate from nonenterprise-owned assets (e.g.,
1136 contracted services that use the enterprise infrastructure to access the internet) or applications
1137 that are resistant to passive monitoring. The enterprise cannot perform deep packet inspection or
1138 examine the encrypted traffic and must use other methods to assess a possible attacker on the
1139 network.

1140 That does not mean that the enterprise is unable to analyze encrypted traffic that it sees on the
1141 network. The enterprise can collect metadata about the encrypted traffic and use that to detect an

1142 active attacker or possible malware communicating on the network. Machine learning techniques
1143 [Anderson] can be used to analyze traffic that cannot be decrypted and examined. Employing
1144 this type of machine learning would allow the enterprise to categorize traffic as valid or possibly
1145 malicious and subject to remediation. In a ZTA deployment, only the traffic from nonenterprise-
1146 owned assets would need to be examined in this way as all enterprise traffic is subject to analysis
1147 by the policy administrator via the PEPs.

1148 **5.5 Storage of Network Information**

1149 A related threat to enterprise analysis of network traffic is the analysis component itself. If
1150 network traffic and metadata are being stored for building contextual policies, forensics, or later
1151 analysis, that data becomes a target for attackers. Just like network diagrams, configuration files,
1152 and other assorted network architecture documents, these resources should be protected. If an
1153 attacker can successfully gain access to stored traffic information, they may be able to gain
1154 insight into the network architecture and identify assets for further reconnaissance and attack.

1155 Another source of reconnaissance information for an attacker in a ZT enterprise is the
1156 management tool used to encode access policies. Like stored traffic, this component contains
1157 access policies to resources and can give an attacker information on which accounts are most
1158 valuable to compromise (e.g., the ones that have access to the desired data resources).

1159 As for all valuable enterprise data, adequate protections should be in place to prevent
1160 unauthorized access and access attempts. As these resources are vital to security, they should
1161 have the most restrictive access policies and be accessible only from designated or dedicated
1162 administrator accounts.

1163 **5.6 Reliance on Proprietary Data Formats**

1164 ZTA relies on several different data sources to make access decisions, including information
1165 about the requesting user, asset used, enterprise and external intelligence, and threat analysis.
1166 Often, the assets used to store and process this information do not have a common, open standard
1167 on how to interact and exchange information. This can lead to instances where an enterprise is
1168 locked into a subset of providers due to interoperability issues. If one provider has a security
1169 issue or disruption, an enterprise may not be able to migrate to a new provider without extreme
1170 cost (e.g., replacing several assets) or going through a long transition program (e.g., translating
1171 policy rules from one proprietary format to another). Like DoS attacks, this risk is not unique to
1172 ZTA, but because ZTA is heavily dependent on the dynamic access of information (both
1173 enterprise and service providers), disruption can affect the core business functions of an
1174 enterprise. To mitigate associated risks, enterprises should evaluate service providers on a
1175 holistic basis by considering factors such as vendor security controls, enterprise switching costs,
1176 and supply chain risk management.

1177 **5.7 Use of Non-person Entities (NPE) in ZTA Administration**

1178 Artificial intelligence and other software-based agents are being deployed to manage security
1179 issues on enterprise networks. These components need to interact with the management
1180 components of ZTA (e.g., policy engine, policy administrator), sometimes in lieu of a human
1181 administrator. How these components authenticate themselves in an enterprise implementing a

1182 ZTA is an open issue. It is assumed that most automated technology systems will use some
1183 means to authenticate when using an API to resource components.

1184 The biggest risk when using automated technology for configuration and policy enforcement is
1185 the possibility of false positives (innocuous actions mistaken for attacks) and false negatives
1186 (attacks mistaken for normal activity). This can be reduced with regular retuning analysis to
1187 correct mistaken decisions and improve the decision process.

1188 The associated risk is that an attacker will be able to induce or coerce an NPE to perform some
1189 task that the attacker is not privileged to perform. The software agent may have a lower bar for
1190 authentication (e.g., API key versus MFA) to perform administrative or security-related tasks
1191 compared with a human user. If an attacker can interact with the agent, they could theoretically
1192 trick the agent into allowing the attacker greater access or into performing some task on behalf of
1193 the attacker. There is also a risk that an attacker could gain access to a software agent's
1194 credentials and impersonate the agent when performing tasks.

1195 **6 Zero Trust Architecture and Possible Interactions with Existing Federal** 1196 **Guidance**

1197 Several existing federal policies and guidance intersect with the planning, deployment, and
1198 operation of a ZTA. These policies do not prohibit an enterprise from moving to a more zero
1199 trust-oriented architecture but can influence development of a zero trust strategy for an agency.
1200 When complemented with existing cybersecurity policies and guidance, ICAM, continuous
1201 monitoring, and general cyber hygiene, ZTA may reinforce an organization's security posture
1202 and protect against common threats.

1203 **6.1 ZTA and NIST Risk Management Framework**

1204 A ZTA deployment involves developing access policies around acceptable risk to the designated
1205 mission or business process (see Section 7.3.3). It is possible to deny all network access to a
1206 resource and allow access only via a connected terminal, but this is disproportionately restrictive
1207 in the majority of cases and inhibits work from being accomplished. For a federal agency to
1208 perform its mission, there is an acceptable level of risk. The risks associated with performing the
1209 given mission must be identified, evaluated, and mitigated. To assist in this, the NIST Risk
1210 Management Framework (RMF) was developed.

1211 ZTA planning and implementation may change the authorization boundaries defined by the
1212 enterprise. This is due to the addition of new components (e.g., policy engine, policy
1213 administrator, and PEPs) and a reduction of reliance on network perimeter defenses. The overall
1214 process described in the RMF will not change in a ZTA.

1215 **6.2 ZT and NIST Privacy Framework**

1216 Protecting the privacy of users and private information (e.g., personally identifiable information)
1217 is a prime concern for organizations. Privacy and data protections are included in compliance
1218 programs such as FISMA and the Health Insurance Portability and Accountability Act (HIPAA).
1219 In response, NIST produced a Privacy Framework for use by organizations [NISTPRIV]. This
1220 document provides a framework to describe privacy risks and mitigation strategies, as well as a
1221 process for an enterprise to identify, measure, and mitigate risks to user privacy and private
1222 information stored and processed by an organization. This includes personal information used by
1223 the enterprise to support ZTA operations and any biometric attributes used in access request
1224 evaluations.

1225 Part of the core requirements for ZTA is that an enterprise should inspect and log traffic (or
1226 metadata when dealing with encrypted traffic) in its environment. Some of this traffic may
1227 contain private information or have associated privacy risks. Organizations will need to identify
1228 any possible risks associated with intercepting, scanning, and logging network traffic [NISTIR
1229 8062]. This may include actions such as informing users, obtaining consent (via a login page,
1230 banner, or similar), and educating enterprise users. The NIST Privacy Framework could help in
1231 developing a formal process to identify and mitigate any privacy-related risks to an enterprise
1232 developing a zero trust architecture.

1233 **6.3 ZTA and Federal Identity, Credential, and Access Management Architecture**

1234 User provisioning is a key component of ZTA. The policy engine cannot determine if attempted
1235 connections are authorized to connect to a resource if the PE has insufficient information to
1236 identify associated users and resources. Strong user provision and authentication policies need to
1237 be in place before moving to a more zero trust–aligned deployment. Enterprises need a clear set
1238 of user attributes and policies that can be used by a PE to evaluate access requests.

1239 The Office of Management and Budget (OMB) issued M-19-17 on improving identity
1240 management for the Federal Government. The goal of the policy is to develop “...a common
1241 vision for identity as an enabler of mission delivery, trust, and safety of the Nation” [M-19-17].
1242 The memo calls on all federal agencies to form an ICAM office to govern efforts related to
1243 identity issuance and management. Many of these management policies should use the
1244 recommendations in NIST SP 800-63-3, *Digital Identity Guidelines* [SP800-63]. As ZTA is
1245 heavily dependent on precise identity management, any ZTA effort will need to integrate the
1246 agency’s ICAM policy.

1247 **6.4 ZTA and Trusted Internet Connections 3.0**

1248 Trusted Internet Connections (TIC) is a federal cybersecurity initiative jointly managed by the
1249 Office of Management and Budget (OMB), the Department of Homeland Security Cybersecurity
1250 & Infrastructure Security Agency (DHS CISA), and the General Services Administration to
1251 establish a network security baseline across the Federal Government. Historically, TIC was a
1252 perimeter-based cybersecurity strategy that required agencies to consolidate and monitor their
1253 external network connections. Inherent in TIC 1.0 and TIC 2.0 is the assumption that the inside
1254 of the perimeter is trusted, whereas ZTA assumes that network location does not infer trust (i.e.,
1255 there is no trust on an agency’s internal network). TIC 2.0 provides a list of network-based
1256 security capabilities (e.g., content filtering, monitoring, authentication) to be deployed at the TIC
1257 access point at the agency’s perimeter; many of these capabilities are aligned with ZTA.

1258 TIC 3.0 will be updated to accommodate cloud services and mobile devices [M-19-26]. In TIC
1259 3.0, agencies can define trust zones as low trust, moderate trust, and high trust based on the level
1260 of control, transparency, and verification that an agency has over a particular computing
1261 environment as well as the sensitivity of data associated with that environment. In addition, TIC
1262 3.0 has updated the network-based security capabilities to be applied to multiple PEPs, which are
1263 located at the boundary of a trust zone and not at a single PEP at the agency perimeter. Many of
1264 these TIC 3.0 security capabilities directly support ZTA (e.g., encrypted traffic, default/deny,
1265 virtualization security, network and asset inventory). TIC 3.0 defines specific use cases that
1266 describe the implementation of trust zones and security capabilities across specific applications,
1267 services, and environments.

1268 TIC 3.0 is focused on network-based security protections, whereas ZTA is a more inclusive
1269 architecture that addresses application, user, and data protections. As TIC 3.0 evolves its use
1270 cases, it is likely that a ZTA TIC use case will be developed to define the network protections to
1271 be deployed at ZTA enforcement points.

1272 **6.5 ZTA and EINSTEIN (NCPS – National Cybersecurity Protection System)**

1273 NCPS (also known as EINSTEIN) is an integrated system-of-systems that delivers intrusion
1274 detection, advanced analytics, information sharing, and intrusion prevention capabilities to
1275 defend the Federal Government from cyber threats. The goals of NCPS, which align with the
1276 overarching goals of zero trust, are to manage cyber risk, improve cyber protection, and
1277 empower partners to secure cyber space. EINSTEIN sensors enable CISA’s National
1278 Cybersecurity and Communications Integration Center to defend federal networks and respond
1279 to significant incidents at federal agencies.

1280 The placement of NCPS sensors is based on a perimeter network defense in the Federal
1281 Government, while zero trust architectures move protections closer to the data and resources. If
1282 ZTA is adopted across the Federal Government, the NCPS implementation would need to
1283 evolve, or new capabilities would need to be deployed to fulfill NCPS objectives. Incident
1284 responders could potentially leverage information from authentication, traffic inspection, and
1285 logging of agency traffic available to federal agencies that have implemented a zero trust
1286 architecture. Information generated in a ZTA may better inform event impact quantification.
1287 Machine learning tools could use ZTA data to improve detection, and additional logs from ZTA
1288 may be saved for after-the-fact analyses by incident responders.

1289 **6.6 ZTA and DHS Continuous Diagnostics and Mitigations (CDM) Program**

1290 The DHS CDM program is an effort to improve federal agency information technology (IT).
1291 Vital to that posture is an agency’s insight into the assets, configuration, and users within itself.
1292 To protect a system, agencies need to set up processes to discover and understand the basic
1293 components and actors in their infrastructure:

- 1294 • **What is connected?** What devices, applications, and services are used by the
1295 organization? This includes observing and improving the security posture of these
1296 artifacts as vulnerabilities and threats are discovered.
- 1297 • **Who is using the network?** What users are part of the organization or are external and
1298 allowed to access enterprise resources? These include NPEs that may be performing
1299 autonomous actions.
- 1300 • **What is happening on the network?** An enterprise needs insight into traffic patterns
1301 and messages between systems.
- 1302 • **How is data protected?** The enterprise needs a set policy on how information is
1303 protected at rest, in transit, and in use.

1304 Having a strong CDM program implementation is key to the success of ZTA. For example, to
1305 move to ZTA, an enterprise must have a system to discover and record physical and virtual
1306 assets to create a usable inventory. The DHS CDM program has initiated several efforts to build
1307 the capabilities needed within federal agencies to move to a ZTA. For example, the DHS
1308 Hardware Asset Management (HWAM) [HWAM] program is an effort to help agencies identify
1309 devices on their network infrastructure to deploy a secure configuration. This is similar to the
1310 first steps in developing a road map to ZTA. Agencies must have visibility into the assets active
1311 on the network (or those accessing resources remotely) to categorize, configure, and monitor the
1312 network’s activity.

1313 **6.7 ZTA, Cloud Smart, and the Federal Data Strategy**

1314 The Cloud Smart⁸ strategy, updated Data Center Optimization Initiative [M-19-19] policy, and
1315 Federal Data Strategy⁹ all influence some requirements for agencies planning a ZTA. These
1316 policies require agencies to inventory and assess how they collect, store, and access data, both on
1317 premises and in the cloud.

1318 This inventory is critical to determining what business processes and resources would benefit
1319 from implementing a ZTA. Data resources and applications that are primarily cloud-based or
1320 primarily used by remote workers are good candidates for a ZTA approach (see Section 7.3.3)
1321 because the users and resources are located outside of the enterprise network perimeter and are
1322 likely to see the most benefit in use, scalability, and security.

1323 One additional consideration with the Federal Data Strategy is how to make agency data assets
1324 accessible to other agencies or the public. This corresponds with the cross-enterprise
1325 collaboration ZTA use case (see Section 4.4). Agencies using a ZTA for these assets may need to
1326 take collaboration or publication requirements into account when developing the strategy.

1327

⁸ Federal Cloud Computing Strategy: <https://cloud.cio.gov/strategy/>

⁹ Federal Data Strategy: <https://strategy.data.gov/>

1328 **7 Migrating to a Zero Trust Architecture**

1329 Implementing a ZTA is a journey rather than a wholesale replacement of infrastructure or
1330 processes. An organization should seek to incrementally implement zero trust principles, process
1331 changes, and technology solutions that protect its highest value data assets. Most enterprises will
1332 continue to operate in a hybrid zero-trust/perimeter-based mode for an indefinite period while
1333 continuing to invest in ongoing IT modernization initiatives.

1334 How an enterprise migrates to a strategy depends on its current cybersecurity posture and
1335 operations. An enterprise should reach a baseline of competence before it becomes possible to
1336 deploy a significant ZT-focused environment [ACT-IAC]. This baseline includes having assets,
1337 users, and business processes identified and cataloged for the enterprise. The enterprise needs
1338 this information before it can develop a list of candidate business processes and the users/assets
1339 involved in this process.

1340 **7.1 Pure Zero Trust Architecture**

1341 In a greenfield approach, it would be possible to build a zero trust architecture from the ground
1342 up. Assuming the enterprise knows the applications and workflows that it wants to use for its
1343 operations, it can produce an architecture based on zero trust tenets for those workflows. Once
1344 the workflows are identified, the enterprise can narrow down the components needed and begin
1345 to map how the individual components interact. From that point, it is an engineering and
1346 organizational exercise in building the infrastructure and configuring the components. This may
1347 include additional organizational changes depending on how the enterprise is currently set up
1348 and operating.

1349 In practice, this is rarely a viable option for federal agencies or any organization with an existing
1350 network. However, there may be times when an organization is asked to fulfill a new
1351 responsibility that would require building its own infrastructure. In these cases, it might be
1352 possible to introduce ZT concepts to some degree. For example, an agency may be given a new
1353 responsibility that entails building a new application and database. The agency could design the
1354 newly needed infrastructure around ZT principles, such as having users' trust evaluated before
1355 access is granted and having micro-perimeters around new resources. The degree of success
1356 depends on how dependent this new infrastructure is on existing resources (e.g., ID management
1357 systems).

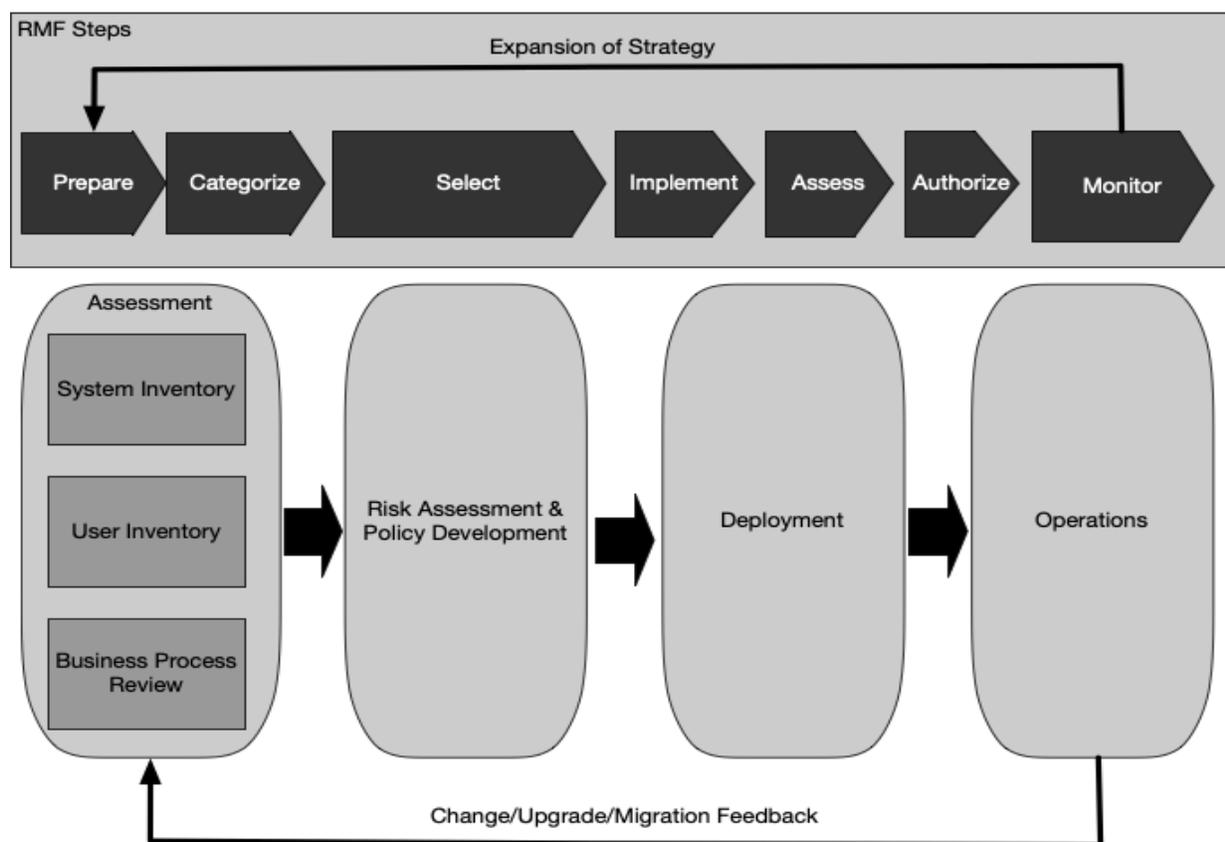
1358 **7.2 Hybrid ZTA and Perimeter-Based Architecture**

1359 It is unlikely that any significant enterprise can migrate to zero trust in a single technology
1360 refresh cycle. There may be an indefinite period when ZTA workflows coexist in a traditional
1361 enterprise. Migration to a ZTA approach to the enterprise may take place one business process at
1362 a time. The enterprise needs to make sure that the common elements (e.g., ID management,
1363 device management, event logging) are flexible enough to operate in a ZTA and perimeter-based
1364 hybrid security architecture. Enterprise architects may also want to restrict ZTA candidate
1365 solutions to those that can interface with existing components.

1366 **7.3 Steps to Introducing ZTA to a Perimeter-Based Architected Network**

1367 Migrating to ZTA requires an organization to have detailed knowledge of its assets (physical and
 1368 virtual), users (including user privileges), and business processes. This knowledge is accessed by
 1369 the PE when evaluating resource requests. Incomplete knowledge will most often lead to a
 1370 business process failure where the PE denies requests due to insufficient information. This is
 1371 especially an issue if there are unknown “shadow IT” deployments operating within an
 1372 organization.

1373 Before undertaking an effort to bring ZTA to an enterprise, there should be a survey of assets,
 1374 users, data flows, and workflows. This is the foundation state that must be reached before a ZTA
 1375 deployment is possible. These surveys can be conducted in parallel, but both are tied to
 1376 examination of the business processes of the organization. These steps can be mapped to the
 1377 steps in the RMF [SP800-37] as any adoption of a ZTA is a process to reduce risk to an agency’s
 1378 business functions. The pathway to implementing a ZTA can be visualized in Figure 12.



1379

1380

Figure 12: ZTA Deployment Cycle

1381 After the initial inventory is created, there is a regular cycle of maintenance and updating. This
 1382 updating may either change business processes or not have any impact, but an evaluation of
 1383 business processes should be conducted. For example, a change in digital certificate providers
 1384 may not appear to have a significant impact but may involve certificate root store management,
 1385 Certificate Transparency log monitoring, and other factors that are not apparent at first.

1386 **7.3.1 Identify Actors on the Enterprise**

1387 For a zero trust enterprise to operate, the PE must have knowledge of enterprise subjects.
1388 Subjects could encompass both human and possible NPEs, such as service accounts that interact
1389 with resources.

1390 Users with special privileges, such as developers or system administrators, require additional
1391 scrutiny when being assigned attributes or roles. In a traditional security architecture, these
1392 accounts may have blanket permission to access all enterprise resources. ZTA should allow
1393 developers and administrators to have sufficient flexibility to satisfy their business requirements
1394 while using logs and audit actions to identify access behavior patterns. ZTA deployments may
1395 require administrators to satisfy a more stringent confidence level or criteria as outlined in NIST
1396 SP 800-63A, Section 5 [SP800-63A].

1397 **7.3.2 Identify Assets Owned by the Enterprise**

1398 As mentioned in Section 2.1, one of the key requirements of ZTA is the ability to identify and
1399 manage devices. ZTA also requires the ability to identify and monitor nonenterprise-owned
1400 devices that may be on enterprise-owned network infrastructure or that access enterprise
1401 resources. The ability to manage enterprise assets is key to the successful deployment of ZTA.
1402 This includes hardware components (e.g., laptops, phones, IoT devices) and digital artifacts (e.g.,
1403 user accounts, applications, digital certificates). It may not be possible to conduct a complete
1404 census on all enterprise-owned assets, so an enterprise should consider building the capability to
1405 quickly identify, categorize, and assess newly discovered assets that are on enterprise-owned
1406 infrastructure.

1407 This goes beyond simply cataloging and maintaining a database of enterprise assets. This also
1408 includes configuration management and monitoring. The ability to observe the current state of an
1409 asset is part of the process of evaluating access requests (see Section 2.1). This means that the
1410 enterprise must be able to configure, survey, and update enterprise assets, such as virtual assets
1411 and containers. This also includes both its physical (as best estimated) and network location. This
1412 information should inform the PE when making resource access decisions.

1413 Nonenterprise-owned assets and enterprise-owned “shadow IT” should also be cataloged as well
1414 as possible. This may include whatever is visible by the enterprise (e.g., MAC address, network
1415 location) and augmented by administrator data entry. This information is not only used for access
1416 decisions (as collaborator and BYOD assets may need to contact PEPs) but also for monitoring
1417 and forensics logging by the enterprise. Shadow IT presents a special problem in that these
1418 resources are enterprise-owned but not managed like other resources. Certain ZTA approaches
1419 (mainly network-based) may even cause shadow IT components to become unusable as they may
1420 not be known and included in network access policies.

1421 Many federal agencies have already begun identifying enterprise assets. Agencies that have
1422 established CDM program capabilities, such as HWAM [HWAM] and Software Asset
1423 Management (SWAM) [SWAM], have a rich set of data to draw from when enacting a ZTA.
1424 Agencies may also have a list of ZTA candidate processes that involve High Value Assets
1425 (HVA) [M-19-03] that have been identified as key to the agency mission. This work would need

1426 to exist enterprise- or agency-wide before any business process could be (re)designed with a
1427 ZTA. These programs must be designed to be expandable and adaptable to changes in the
1428 enterprise, not only when migrating to ZTA but also when accounting for new assets, services,
1429 and business processes that become part of the enterprise.

1430 **7.3.3 Identify Key Processes and Evaluate Risks Associated with Executing Process**

1431 The third inventory that an agency should undertake is to identify and rank the business
1432 processes, data flows, and their relation in the missions of the agency. Business processes should
1433 inform the circumstances under which resource access requests are granted and denied. An
1434 enterprise may wish to start with a low-risk business process for the first transition to ZTA as
1435 disruptions will likely not negatively impact the entire organization. Once enough experience is
1436 gained, more critical business processes can become candidates.

1437 Business processes that utilize cloud-based resources or are used by remote workers are often
1438 good candidates for ZTA and would likely see improvements to availability and security. Rather
1439 than project the enterprise perimeter into the cloud or bring clients into the enterprise network
1440 via a VPN, enterprise clients can request cloud services directly. The enterprise's PEPs ensure
1441 that enterprise policies are followed before resource access is granted to a client.

1442 **7.3.4 Formulating Policies for the ZTA Candidate**

1443 The process of identifying a candidate application or business workflow depends on several
1444 factors: the importance of the process to the organization, the group of users affected, and the
1445 current state of resources used for the workflow. The value of the asset or workflow based on
1446 risk to the asset or workflow can be evaluated using the NIST Risk Management Framework
1447 [SP800-37].

1448 After the asset or workflow is identified, identify all upstream resources (e.g., ID management
1449 systems, databases, micro-services), downstream resources (e.g., logging, security monitoring),
1450 and entities (e.g., users, service accounts) that are used or affected by the workflow. This may
1451 influence the candidate choice as a first migration to ZTA. An application used by an identified
1452 subset of enterprise users (e.g., a purchasing system) may be preferred over one that is vital to
1453 the entire user base of the enterprise (e.g., email).

1454 The enterprise administrators then need to determine the set of criteria (if using a criteria-based
1455 TA) or confidence level weights (if using a score-based TA) for the resources used in the
1456 candidate business process (see Section 3.3.1). Administrators may need to adjust these criteria
1457 or values during the tuning phase. These adjustments are necessary to ensure that policies are
1458 effective but do not hinder access to resources.

1459 **7.3.5 Identifying Candidate Solutions**

1460 Once a list of candidate business processes has been developed, enterprise architects can
1461 compose a list of candidate solutions. Some deployment models (see Section 3.1) are better
1462 suited to particular workflows and current enterprise ecosystems. Likewise, some vendor
1463 solutions are better suited to some use cases than others. These are some factors to consider:

- 1464 • **Does the solution require that components be installed on the client asset?** This may
1465 limit business processes where nonenterprise-owned assets are used or desired, such as
1466 BYOD or cross-agency collaborations.
- 1467 • **Does the solution work where the business process resources exist entirely on**
1468 **enterprise premises?** Some solutions assume that requested resources will reside in the
1469 cloud (so-called north-south traffic) and not within an enterprise perimeter (east-west
1470 traffic). The location of candidate business process resources will influence candidate
1471 solutions as well as the ZTA for the process.
- 1472 • **Does the solution provide a means to log interactions for analysis?** A key component
1473 of ZT is the collection and use of data related to the process flow that feeds back into the
1474 PE when making access decisions.

1475 One solution is to model an existing business process as a pilot program rather than just a
1476 replacement. This pilot program could be made general to apply to several business processes or
1477 be made specific to one use case. The pilot can be used as a “proving ground” for ZTA before
1478 transitioning users to the ZTA deployment and away from the traditional process infrastructure.

1479 **7.3.6 Initial Deployment and Monitoring**

1480 Once the candidate workflow and ZTA components are chosen, the initial deployment can start.
1481 Enterprise administrators must implement the developed policies by using the selected
1482 components but may wish to operate in an observation and monitoring mode at first. Few
1483 enterprise policy sets are complete in their first iterations: important user accounts (e.g.,
1484 administrator accounts) may be denied access to resources they need or may not need all the
1485 access privileges they have been assigned.

1486 The new ZT business workflow could be operated in reporting-only mode for some time to make
1487 sure the policies are effective and workable. Reporting-only means that access should be granted
1488 for most requests, and logs and traces of connections should be compared with the initial
1489 developed policy. Basic policies such as denying requests that fail MFA or appear from known,
1490 blacklisted IP addresses should be enforced and logged, but after initial deployment, access
1491 polices should be more lenient to collect data from actual interactions of the ZT workflow. If it is
1492 not possible to operate in a more lenient nature, enterprise network operators should monitor logs
1493 closely and be prepared to modify access policies based on operational experience.

1494 **7.3.7 Expanding the ZTA**

1495 When enough confidence is gained and the workflow policy set is refined, the enterprise enters
1496 the steady operational phase. The network and assets are still monitored, and traffic is logged
1497 (see Section 2.2.1), but responses and policy modifications are done at a lower tempo as they
1498 should not be severe. The users and stakeholders of the resources and processes involved should
1499 also provide feedback to improve operations. At this stage, the enterprise administrators can
1500 begin planning the next phase of ZT deployment. Like the previous rollout, a candidate
1501 workflow and solution set need to be identified and initial policies developed.

1502 However, if a change occurs to the workflow, the operating ZT architecture needs to be

1503 reevaluated. Significant changes to the system—such as new devices, major updates to software
1504 (especially ZT logical components), and shifts in organizational structure—may result in changes
1505 to the workflow or policies. In effect, the entire process should be reconsidered with the
1506 assumption that some of the work has already been done. For example, new devices have been
1507 purchased, but no new user accounts have been created, so only the device inventory needs to be
1508 updated.

1509 **References**

- [ACT-IAC] American Council for Technology and Industry Advisory Council (2019) *Zero Trust Cybersecurity Current Trends*. Available at <https://www.actiac.org/zero-trust-cybersecurity-current-trends>
- [Anderson] Anderson B, McGrew D (2017) Machine Learning for Encrypted Malware Traffic Classification: Accounting for Noisy Labels and Non-Stationarity. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, Halifax, Nova Scotia, Canada), pp 1723-1732. <https://doi.org/10.1145/3097983.3098163>
- [BCORE] Department of Defense CIO (2007). Department of Defense Global Information Grid Architecture Vision Version 1.0 June 2007. <http://www.acqnotes.com/Attachments/DoD%20GIG%20Architectural%20Vision,%20June%202007.pdf>
- [CSA-SDP] Cloud Security Alliance (2015) SDP Specification 1.0. April 2015. <https://cloudsecurityalliance.org/artifacts/sdp-specification-v1-0/>
- [FIPS199] National Institute of Standards and Technology (2004) Standards for Security Categorization of Federal Information and Information Systems. (U.S. Department of Commerce, Washington, DC), Federal Information Processing Standards Publication (FIPS) 199. <https://doi.org/10.6028/NIST.FIPS.199>
- [Gilman] Gilman E, Barth D (2017) *Zero Trust Networks: Building Secure Systems in Untrusted Networks* (O'Reilly Media, Inc., Sebastopol, CA), 1st Ed.
- [HWAM] Department of Homeland Security (2015) *Hardware Asset Management (HWAM) Capability Description*. Available at https://www.us-cert.gov/sites/default/files/cdm_files/HWAM_CapabilityDescription.pdf
- [IBNVN] R. Cohen, K. Barabash, B. Rochwerger, L. Schour, D. Crisan, R. Birke, C. Minkenberg, M. Gusat, R. Recio and V. Jain. An Intent-based Approach for Network Virtualization. 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp 42-50. <https://ieeexplore.ieee.org/xpl/conhome/6560458/proceeding>
- [JERICHO] The Jericho Forum (2007) *Jericho Forum Commandments*, version 1.2. Available at https://collaboration.opengroup.org/jericho/commandments_v1.2.pdf
- [M-19-03] Office of Management and Budget (2018) Strengthening the Cybersecurity of Federal Agencies by Enhancing the High Value Asset Program. (The White House, Washington, DC), OMB Memorandum M-

- 19-03, December 10, 2018. Available at <https://www.whitehouse.gov/wp-content/uploads/2018/12/M-19-03.pdf>
- [M-19-17] Office of Management and Budget (2019) Enabling Mission Delivery through Improved Identity, Credential, and Access Management. (The White House, Washington, DC), OMB Memorandum M-19-17, May 21, 2019. Available at <https://www.whitehouse.gov/wp-content/uploads/2019/05/M-19-17.pdf>
- [M-19-19] Office of Management and Budget (2019) Update on Data Center Optimization Initiative (DCOI). (The White House, Washington, DC), OMB Memorandum M-19-19, June 25, 2019. Available at https://datacenters.cio.gov/assets/files/m_19_19.pdf
- [M-19-26] Office of Management and Budget (2019) Update to the Trusted Internet Connections (TIC) Initiative. (The White House, Washington, DC), OMB Memorandum M-19-26, September 12, 2019. Available at <https://www.whitehouse.gov/wp-content/uploads/2019/09/M-19-26.pdf>
- [NISTIR 7987] Ferraiolo DF, Gavrila S, Jansen W (2015) Policy Machine: Features, Architecture, and Specification. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) 7987, Rev. 1. <https://doi.org/10.6028/NIST.IR.7987r1>
- [NISTIR 8062] Brooks SW, Garcia ME, Lefkovitz NB, Lightman S, Nadeau EM (2017) An Introduction to Privacy Engineering and Risk Management in Federal Systems. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) 8062. <https://doi.org/10.6028/NIST.IR.8062>
- [NISTPRIV] National Institute of Standards and Technology (2020) Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management. Version 1.0 January 16, 2020. <https://www.nist.gov/privacy-framework/privacy-framework>
- [SDNBOOK] T. Nadeau and K. Gray (2013) *SDN: Software Defined Networks: An Authoritative Review of Network Programmability Technologies*. (O'Reilly) 1st Ed.
- [SP800-37] Joint Task Force (2018) Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-37, Rev. 2. <https://doi.org/10.6028/NIST.SP.800-37r2>
- [SP800-63] Grassi PA, Garcia ME, Fenton JL (2017) Digital Identity Guidelines. (National Institute of Standards and Technology, Gaithersburg, MD),

- NIST Special Publication (SP) 800-63-3, Includes updates as of December 1, 2017. <https://doi.org/10.6028/NIST.SP.800-63-3>
- [SP800-63A] Grassi PA, Fenton JL, Lefkovitz NB, Danker JM, Choong Y-Y, Greene KK, Theofanos MF (2017) Digital Identity Guidelines: Enrollment and Identity Proofing. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63A, Includes updates as of December 1, 2017. <https://doi.org/10.6028/NIST.SP.800-63A>
- [SP800-63-3] Grassi PA, Garcia ME, Fenton JL (2017) Digital Identity Guidelines. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-63-3, Includes updates as of December 1, 2017. <https://doi.org/10.6028/NIST.SP.800-63-3>
- [SP800-160] Ross R, Pillitteri V, Graubart R, Bodeau D, and McQuaid R (2019) Developing Cyber Resilient Systems: A Systems Security Engineering Approach. (National Institute of Standards and Technology, Gaithersburg, MD), Final Public Draft NIST Special Publication (SP) 800-160, Vol. 2. Available at <https://csrc.nist.gov/publications/detail/sp/800-160/vol-2/draft>
- [SP800-162] Hu VC, Ferraiolo DF, Kuhn R, Schnitzer A, Sandlin K, Miller R, Scarfone KA (2014) Guide to Attribute Based Access Control (ABAC) Definition and Considerations. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-162, Includes updates as of February 25, 2019. <https://doi.org/10.6028/NIST.SP.800-162>
- [SWAM] Department of Homeland Security (2015) *Software Asset Management (SWAM) Capability Description*. Available at https://www.us-cert.gov/sites/default/files/cdm_files/SWAM_CapabilityDescription.pdf

1510

1511

1512 **Appendix A—Acronyms**

CDM	Continuous Diagnostics and Mitigation
DHS	Department of Homeland Security
NIST	National Institute of Standards and Technology
PA	Policy Administrator
PE	Policy Engine
PEP	Policy Enforcement Point
RMF	NIST Risk Management Framework
SIEM	Security Information and Event Monitoring
ZTA	Zero Trust Architecture

1513

1514 **Appendix B—Identified Gaps in the Current State-of-the-Art in ZTA**

1515 The current maturity of zero trust components and solutions was surveyed during the research
 1516 conducted in the development of this document. This survey concluded that the current state of
 1517 the ZTA ecosystem is not mature enough for widespread adoption. While it is possible to use
 1518 ZTA strategies to plan and deploy an enterprise environment, there is no single solution that
 1519 provides all the necessary components. Also, few ZTA components available today can be used
 1520 for all of the various workflows present in an enterprise.

1521 The following is a summary of identified gaps in the ZTA ecosystem and areas that need further
 1522 investigation. Some of these areas have some foundation of work, but how ZTA tenets change
 1523 these areas is not well-known as there is not enough experience with diverse ZTA-focused
 1524 enterprise environments.

1525 **B.1 Technology Survey**

1526 Multiple vendors were invited to present their products and views on zero trust. The goal of this
 1527 survey was to identify missing pieces that prevent agencies from moving to a zero trust based
 1528 enterprise infrastructure now or maintaining an existing ZTA implementation. These gaps can be
 1529 categorized into immediate deployment (immediate or short term), systemic gaps that affect
 1530 maintenance or operations (short or midterm), and missing knowledge (areas for future research).
 1531 They are summarized in Table B-1.

1532 **Table B-1: Summary of Identified Deployment Gaps**

Category	Example Questions	Identified Gaps
Immediate deployment	<ul style="list-style-type: none"> • How should procurement requirements be written? • How does a ZTA plan work with TIC, FISMA, and other requirements? 	<ul style="list-style-type: none"> • Lack of a common framework and vocabulary for ZTA • Perception that ZTA conflicts with existing policy
Systemic	<ul style="list-style-type: none"> • How can vendor lock-in be prevented? • How do different ZTA environments interact? 	<ul style="list-style-type: none"> • Too much reliance on vendor APIs
Areas needing more research	<ul style="list-style-type: none"> • How will threats evolve in the face of ZTA? • How will business processes change in the face of ZTA? 	<ul style="list-style-type: none"> • What a successful compromise looks like in an enterprise with a ZTA • Documented end user experience in an enterprise with a ZTA

1533 **B.2 Gaps that Prevent an Immediate Move to ZTA**

1534 These are the issues that are slowing adoption of a ZTA at present. These were classified as
1535 immediate issues, and no thought of future maintenance or migration was considered for this
1536 category. A forward-thinking enterprise may also consider the maintenance category to be of
1537 immediate concern in preventing the initial deployment of ZTA components, but these issues are
1538 considered a separate category for this analysis.

1539 **B.2.1 Lack of Common Terms for ZTA Design, Planning, and Procurement**

1540 Zero trust as a strategy for the design and deployment of enterprise infrastructure is still a
1541 forming concept. Industry has not yet coalesced around a single set of terms or concepts to
1542 describe ZTA components and operations. This makes it difficult for organizations (e.g., federal
1543 agencies) to develop coherent requirements and policies for designing zero trust enterprise
1544 infrastructure and procuring components.

1545 The driver for Sections 2.1 and 3.1 is an initial attempt to form a neutral base of terms and
1546 concepts to describe ZTA. The abstract ZTA components and deployment models were
1547 developed to serve as basic terms and ways to think about ZTA. The goal is to provide a
1548 common way to view, model, and discuss ZTA solutions when developing enterprise
1549 requirements and performing market surveys. The above sections may prove to be incomplete as
1550 more experience is gained with ZTA in federal agencies, but they currently serve as a base for a
1551 common conceptual framework.

1552 **B.2.2 Perception that ZTA Conflicts with Existing Federal Cybersecurity Policies**

1553 There is a misconception that ZTA is a single framework with a set of solutions that are
1554 incompatible with the existing view of cybersecurity. Zero trust should instead be viewed as an
1555 evolution of current cybersecurity strategies as many of the concepts and ideas have been
1556 circulating for a long time. Federal agencies have been encouraged to take a more zero trust
1557 approach to cybersecurity through existing guidance (see Section 6). If an agency has a mature
1558 ID management system and robust CDM capabilities in place, it is on the road to a ZTA (see
1559 Section 7.3). This gap is based on a misconception of ZTA and how it has evolved from previous
1560 cybersecurity paradigms.

1561 **B.3 Systemic Gaps that Impact ZTA**

1562 These are the gaps that affect initial implementation and deployment of ZTA and continued
1563 operation/maturity. These gaps could slow the adoption of ZTA in agencies or result in
1564 fragmentation of the ZTA component industry. Systemic gaps are areas where open standards
1565 (produced either by a standards development organization [SDO] or industry consortium) can
1566 help.

1567 **B.3.3 Standardization of Interfaces Between Components**

1568 During the technology survey, it became apparent that no one vendor offers a single solution that
1569 will provide zero trust. Furthermore, it might not be desirable to use a single-vendor solution to

1570 achieve zero trust and thereby risk vendor lock-in. This leads to interoperability within
1571 components not only at the time of purchase but also over time.

1572 The spectrum of components within the wider enterprise is vast, with many products focusing on
1573 a single niche within zero trust and relying on other products to provide either data or some
1574 service to another component (e.g., integration of MFA for resource access). Vendors too often
1575 rely on proprietary APIs provided by partner companies rather than standardized, vendor-
1576 independent APIs to achieve this integration. The problem with this approach is that these APIs
1577 are proprietary and single-vendor controlled. The controlling vendor can change the API
1578 behavior, and integrators are required to update their products in response. This requires close
1579 partnerships between communities of vendors to ensure early notification of modifications
1580 within APIs, which may affect compatibility between products. This adds an additional burden
1581 on vendors and consumers: vendors need to expend resources to change their products, and
1582 consumers need to apply updates to multiple products when one vendor makes a change to its
1583 proprietary API. Additionally, vendors are required to implement and maintain wrappers for each
1584 partner component to allow maximum compatibility and interoperability. For example, many
1585 MFA product vendors are required to create a different wrapper for each cloud provider or
1586 identity management system to be usable in different kinds of client combinations.

1587 On the customer side, this generates additional problems when developing requirements for
1588 purchasing products. There are no standards that purchasers can rely on to identify compatibility
1589 between products. Hence, it is very difficult to create a multiyear road map for moving into ZTA
1590 because it is impossible to identify a minimum set of compatibility requirements for components.

1591 **B.3.4 Emerging Standards that Address Overreliance on Proprietary APIs**

1592 As there is no single solution to developing a ZTA, there is no single set of tools or services for a
1593 zero trust enterprise. Thus, it is impossible to have a single protocol or framework that enables an
1594 enterprise to move to a ZTA. Currently, there is a wide variety of models and solutions seeking
1595 to become the leading authority of ZTA.

1596 This indicates that there is an opportunity for a set of open, standardized protocols or frameworks
1597 to be developed to aid organizations in migrating to a ZTA. SDOs like the Internet Engineering
1598 Task Force (IETF) have specified protocols that may be useful in exchanging threat information
1599 (called XMPP-Grid [1]). The Cloud Security Alliance (CSA) has produced a framework for
1600 Software Defined Perimeter (SDP) [2] that may also be useful in ZTA. Efforts should be directed
1601 toward surveying the current state of ZTA-related frameworks or the protocols necessary for a
1602 useful ZTA and toward identifying places where work is needed to produce or improve these
1603 specifications.

1604 **B.4 Knowledge Gaps in ZTA and Future Areas of Research**

1605 The gaps listed here do not hinder an organization from adopting a ZTA for its enterprise. These
1606 are gray areas in knowledge about operational ZTA environments, and most arise from a lack of
1607 time and experience with mature zero trust deployments. These are areas of future work for
1608 researchers.

1609 **B.4.5 Attacker Response to ZTA**

1610 A properly implemented ZTA for an enterprise will improve the enterprise's cybersecurity
1611 posture over traditional network perimeter-based security. The tenets of ZTA aim to reduce the
1612 exposure of resources to attackers and minimize or prevent lateral movement within an
1613 enterprise should a host asset be compromised.

1614 However, determined attackers will not sit idle but will instead change behavior in the face of
1615 ZTA. The open issue is how the attacks will change. One possibility is that attacks aimed at
1616 stealing credentials will be expanded to target MFA (e.g., phishing, social engineering). Another
1617 possibility is that in a hybrid ZTA/perimeter-based enterprise, attackers will focus on the
1618 business processes that have not had ZTA tenets applied (i.e., follow traditional network
1619 perimeter-based security)—in effect, targeting the low-hanging fruit in an attempt to gain some
1620 foothold in the ZTA business process.

1621 As ZTA matures, more deployments are seen, and experience is gained, the effectiveness of ZTA
1622 in shrinking the attack surface of resources may become apparent. The metrics of success of
1623 ZTA over older cybersecurity strategies will also need to be developed.

1624 **B.4.6 User Experience in a ZTA Environment**

1625 There has not been a rigorous examination of how end users act in an enterprise that is using a
1626 ZTA. This is mainly due to the lack of large ZTA use cases available for analysis. There have,
1627 however, been studies on how users react to MFA and other security operations that are part of a
1628 ZTA enterprise, and this work could form the basis of predicting end user experience and
1629 behavior when using ZTA workflows in an enterprise.

1630 One set of studies that can predict how ZTA affects end user experience is the work done on the
1631 use of MFA in enterprises and security fatigue. Security fatigue [3] is the phenomenon wherein
1632 end users are confronted with so many security policies and challenges that it begins to impact
1633 their productivity in a negative way. Other studies show that MFA may alter user behavior, but
1634 the overall change is mixed [4] [5]. Some users readily accept MFA if the process is streamlined
1635 and involves devices they are used to using or having with them (e.g., applications on a
1636 smartphone). However, some users resent having to use personally-owned devices for business
1637 processes or feel that they are being constantly monitored for possible violations of IT policies.

1638 **B.4.7 Resilience of ZTA to Enterprise and Network Disruption**

1639 The survey of the ZTA vendor ecosystem displayed the wide range of infrastructure that an
1640 enterprise deploying a ZTA would need to consider. As previously noted, there is no single
1641 provider of a full zero trust solution at this time. As a result, enterprises will purchase several
1642 different services and products, which can lead to a web of dependencies for components. If one
1643 vital component is disrupted or unreachable, there could be a cascade of failures that impact one
1644 or multiple business processes.

1645 Most products and services surveyed relied on a cloud presence to provide robustness, but even
1646 cloud services have been known to become unreachable through either an attack or simple error.
1647 When this happens, key components used to make access decisions may be unreachable or may

1648 not be able to communicate with other components. For example, PE and PA components
1649 located in a cloud may be reachable during a distributed denial-of-service (DDoS) attack but may
1650 not be able to reach all PEPs located with resources. Research is needed on discovering the
1651 possible choke points of ZTA deployment models and the impact on network operations when a
1652 ZTA component is unreachable or has limited reachability.

1653 The continuity of operations (COOP) plans for an enterprise will likely need revision when
1654 adopting a ZTA. A ZTA makes many COOP factors easier as remote workers may have the
1655 same access to resources that they had on-premises. However, policies like MFA may also have
1656 a negative impact if users are not properly trained or lack experience. Users may forget or not
1657 have access to tokens and enterprise devices during an emergency, and that will impact the speed
1658 and effectiveness of enterprise business processes.

1659 **B.5 ZTA Test Environment**

1660 TBD – describe NCCoE test lab and tests to be performed

1661 **B.6 References**

- [1] Cam-Winget N (ed.), Appala S, Pope S, Saint-Andre P (2019) Using Extensible Messaging and Presence Protocol (XMPP) for Security Information Exchange. (Internet Engineering Task Force (IETF)), IETF Request for Comments (RFC) 8600. <https://doi.org/10.17487/RFC8600>
- [2] Software Defined Perimeter Working Group “SDP Specification 1.0” Cloud Security Alliance. April 2014.
- [3] Stanton B, Theofanos MF, Spickard Prettyman S, Furman S (2016) Security Fatigue. *IT Professional* 18(5):26-32. <https://doi.org/10.1109/MITP.2016.84>
- [4] Strouble D, Shechtman GM, Alsop AS (2009) Productivity and Usability Effects of Using a Two-Factor Security System. *S AIS 2009 Proceedings* (AIS, Charleston, SC), p 37. Available at <http://aisel.aisnet.org/sais2009/37>
- [5] Weidman J, Grossklags J (2017) I Like It but I Hate It: Employee Perceptions Towards an Institutional Transition to BYOD Second-Factor Authentication. *Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC 2017)* (ACM, Orlando, FL), pp 212-224. <https://doi.org/10.1145/3134600.3134629>

1662