Telemetry Assurance for Zero Trust Architecture

Ian Farquhar
Security Chief Technology Officer
Gigamon

Gigamon®

Ian has been in InfoSec for 33 years, in various capacities. He has previously worked at RSA, Cisco, Sun, SGI, and Macquarie University. At Gigamon Ian is the Security CTO, responsible for security architecture; Zero Trust; cryptography, and SSL/TLS; and risk management in enterprise, service provider, and government networks. Ian leads the Gigamon PSIRT team and has contributed to the design of multiple Gigamon products.

Ian Farquhar

Security Chief Technology Officer
Gigamon

@IanAtGigamon
https://www.linkedin.com/in/ianfarquhar/

Background

Gigamon and ZTA

- Gigamon has been involved in multiple early pilots around ZTA, including a pivotal one run by US DoD
- Gigamon regularly contributes to ZTA standards, and our feedback is publicly posted here ("Gigamon comments…"):

  https://blog.gigamon.com/category/zero-trust/
- Gigamon sees ZTA as something which will be (and should be) an intrinsically-interoperating multi-vendor, hybrid multi-cloud deployment
- Gigamon asserts that any operationalized ZTA infrastructure must be capable of defending managed and unmanaged devices, in any environment (physical, virtual, cloud, containerized, tactical)
- This session assumes a basic understanding of ZTA as defined by NIST SP 800-207

Agenda

- ZTA is a data-centric security architecture, based on risk-based analytics driven from telemetry
- The minimization of “implicit trust” is critical to an operationalized ZTA infrastructure
- Telemetry sensor location matters
- It is important to understand the trustworthiness of your telemetry – introducing a telemetry assurance evaluation framework
- Closing thoughts

Zero Trust Frameworks or Maturity Models

Frameworks, Maturity Models, Strategic Program Implementations etc. etc.

Model Name	Applicability	Current Status
NIST SP 800-207	All	Excepting Forrester, universal, but it's definitional
CISA Zero Trust Maturity Model v2	US Government Civilian Agencies	Advanced – all agencies engaging due to EO 14028
DOD Zero Trust Reference Framework v2	US Military and Intel (plus coalition partners?)	Advanced – all agencies engaging due to EO 14028
Forrester Zero Trust Extended	Enterprise	Widely implemented and good traction
Cloud Security Alliance Zero Trust	Enterprise	Strong focus, early days – not just cloud only
Gartner Zero Trust Strategic Roadmap	Enterprise	Very new – tracking, impressive
Zero Trust Network Access (Gartner)	Enterprise	As a step into proper ZTA? VPN replacement.
NSTAC Zero Trust Approach	Enterprise and government	Little evidence of implementation
UK NSCS Zero Trust Model	UK Government	Little evidence of implementation
Google Beyondcorp	Enterprise	Runs at Google – not much elsewhere (website last updated 2018)
Singapore GovZTA	Singapore Government	Very new – paralleling CISA's ZTA MM
No model – "we're just doing 'zero trust'"	n/a	Very common – low maturity orgs

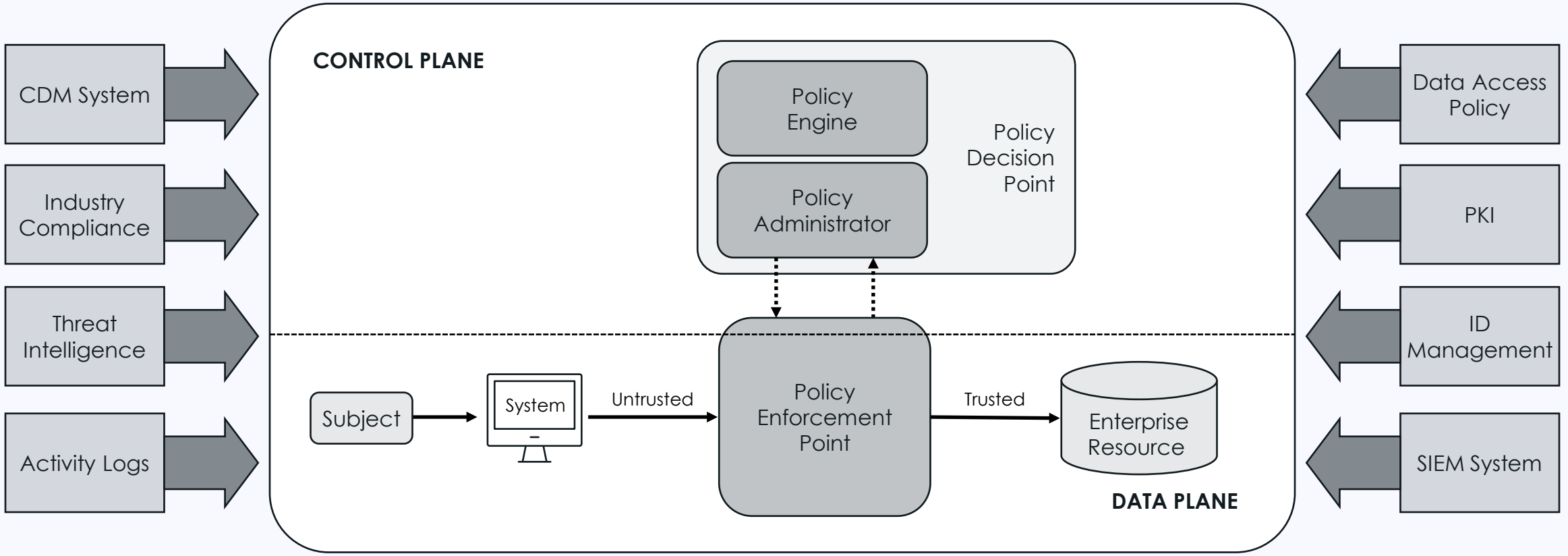
Assertion #1

ZTA is driven by analytics.

Sub-Assertion #1A

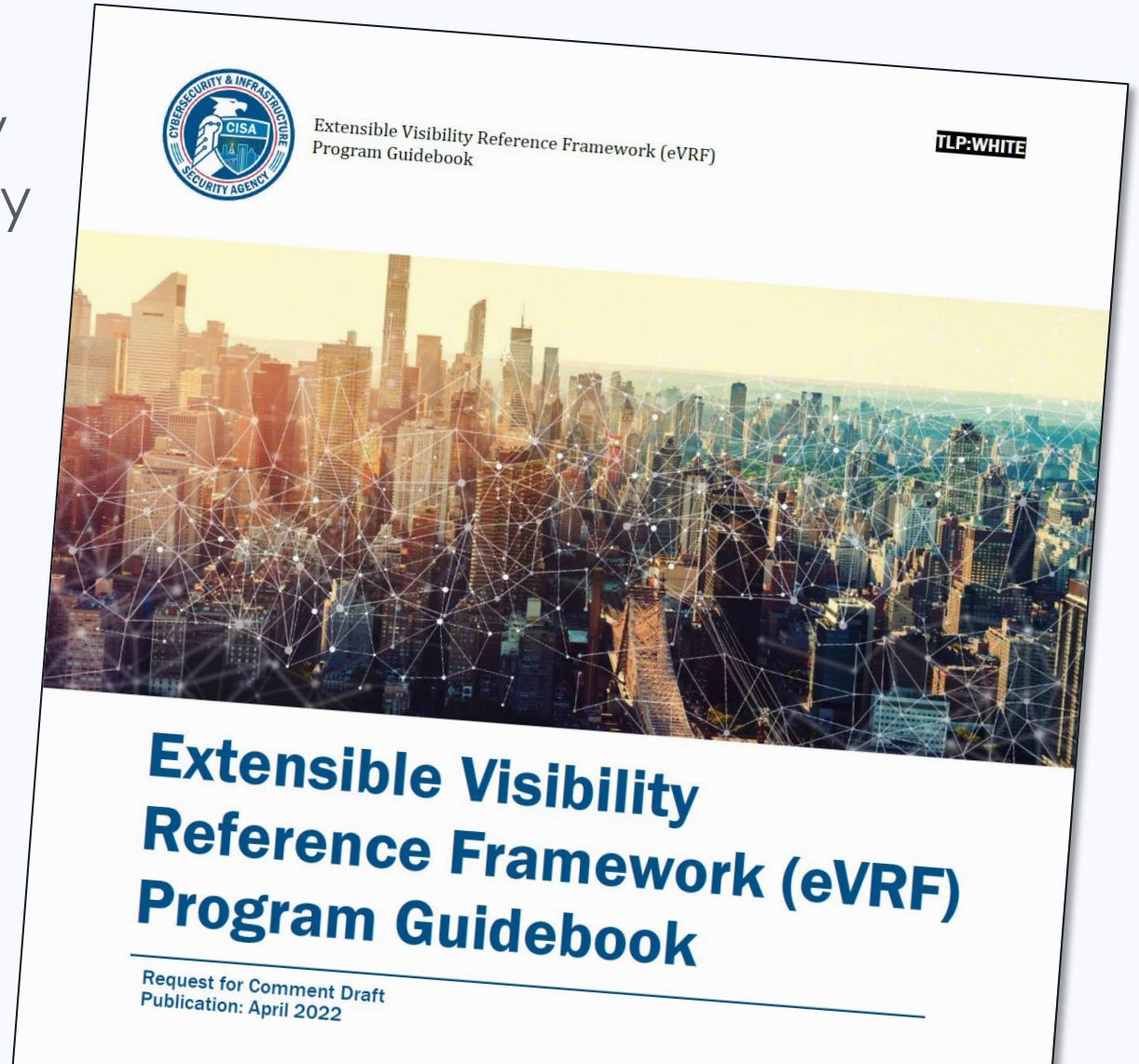
Reliable, comprehensive telemetry feeding the analytic capability is critical to a resilient and capable ZTA

The Fundamental ZTA Concept (NIST SP 800-207 Fig 2 p.9)



Extensible Visibility Reference Framework (eVRF) Program Guidebook

- Published by the US Cybersecurity and Infrastructure Security Agency (CISA)
- Publication date April 2022
- Still in draft as of November 2023
- Location: https://www.cisa.gov/sites/default/files/publications/eVRF_Guidebook_RFC_508C.pdf



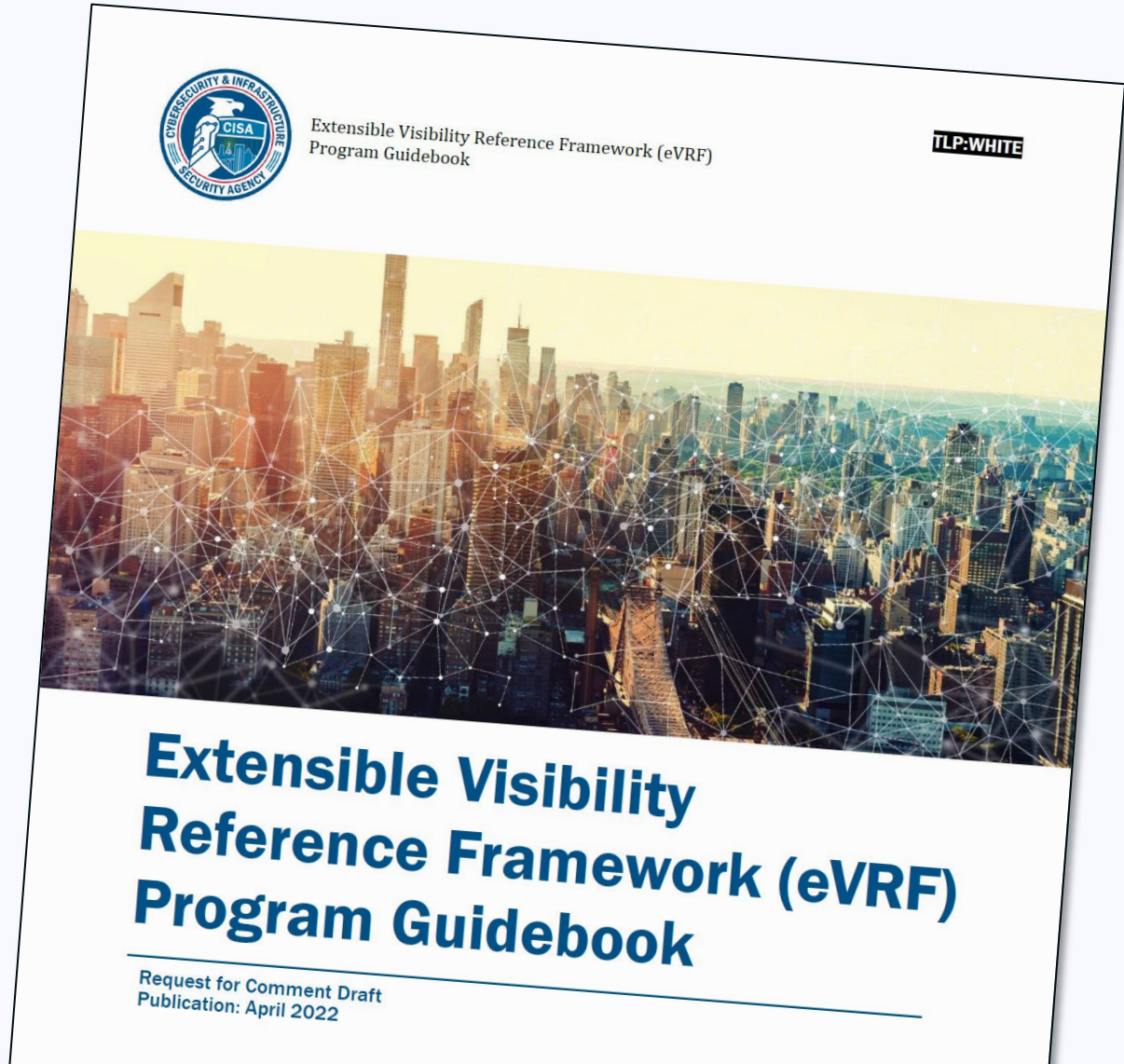
"The purpose of the extensible Visibility Reference Framework (eVRF) is to provide a framework for organizations to identify visibility data that can be used to mitigate threats, understand the extent to which specific products and services provide that visibility data, and identify potential visibility gap"

Telemetry

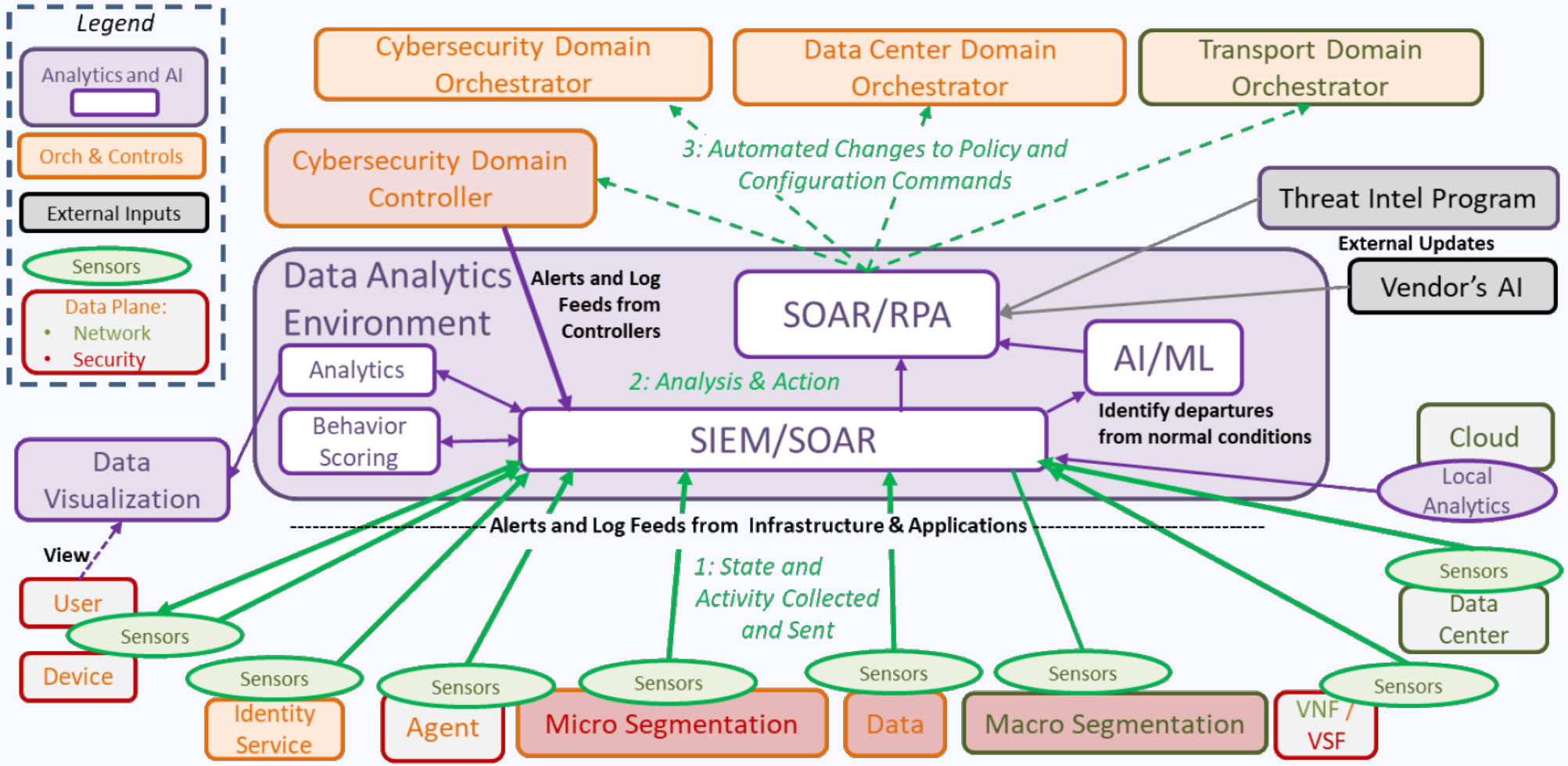
Artifacts derived from security capabilities that provide visibility into security posture, often through automated collections

Extensible Visibility Reference Framework (eVRF) Program Guidebook

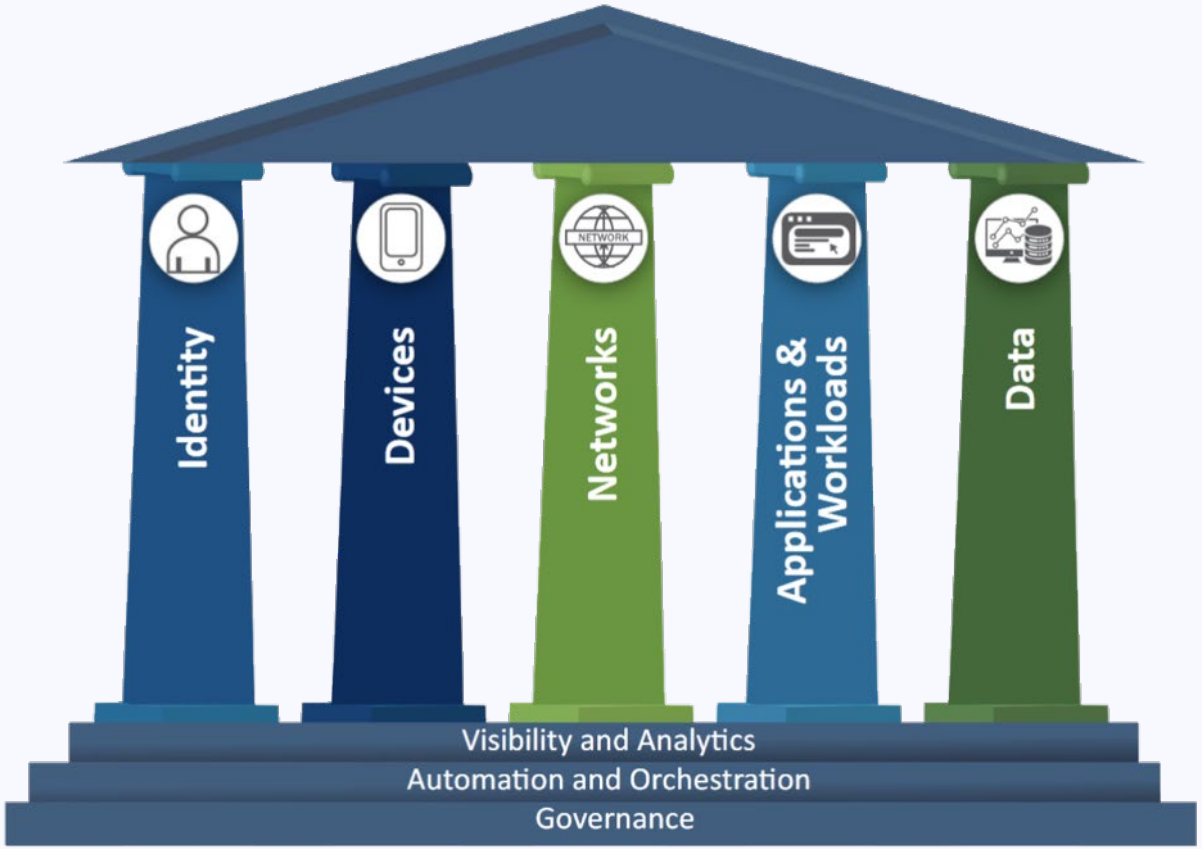
- **VISIBILITY SURFACE**: A digital environment for which cyber-observable data exists, or should exist
- **DOMAINS**: Dividing the enterprise into components to manage visibility requirements
- **OBSERVATION POINT**: Defines the architectural location of "telemetry sources" in the given domain
- **SENSOR**: Telemetry collection entity at an observation point
- **COVERAGE MAP**: Maps the visibility gained into the MITRE ATT&CK framework, to determine whether a specific technique is visible to the enterprise

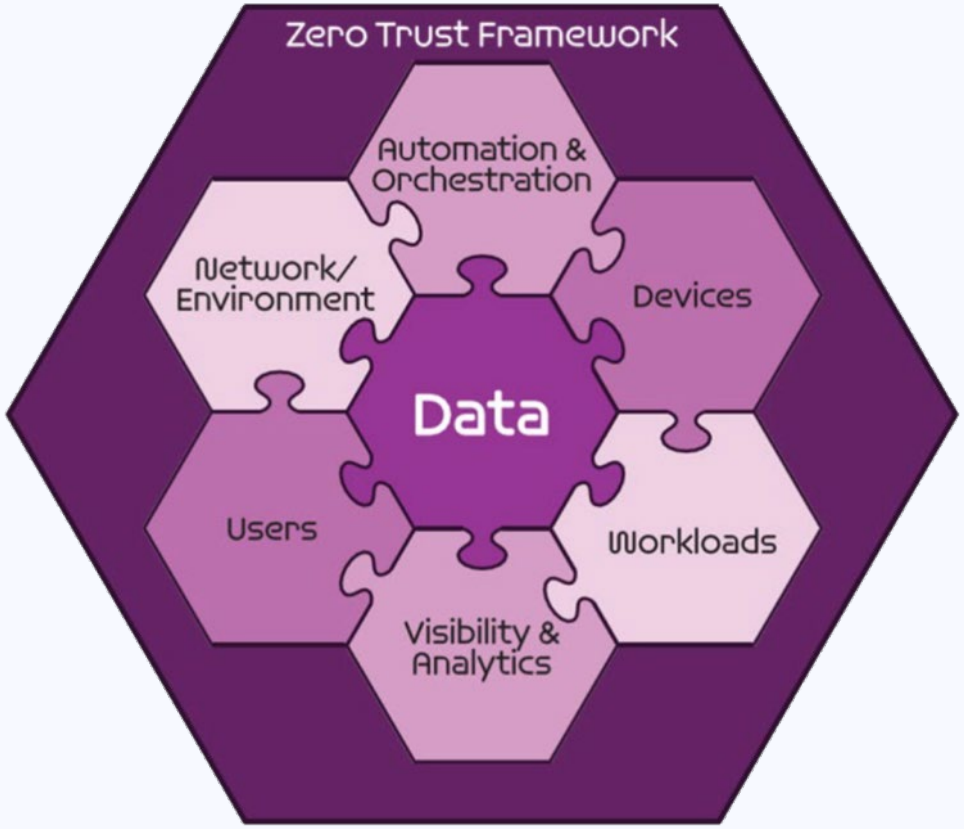


DoD's Policy Engine (Data Analytics & AI [SV-1])



Source: "Department of Defense Zero Trust Reference Architecture v2.0", by DISA and NSA, July 2002, Figure 17, p.44.





Source: "Department of Defense Zero Trust Reference Architecture v2.0", by DISA and NSA, July 2002, Figure 5, p.22.
Source: CISA Zero Trust Maturity Model v2, figure 1 page 7.

Assertion #2

The minimization of “implicit trust” is critical to an operationalized ZTA infrastructure

Sub-Assertion #2A

… especially in government, military and DIB environments

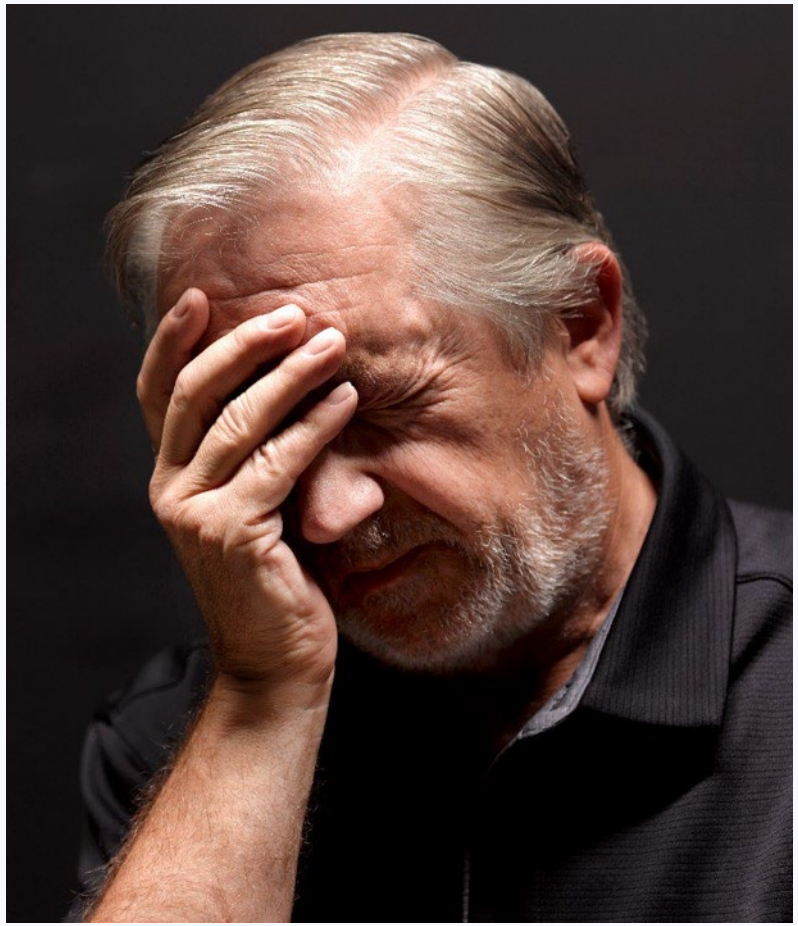
Implicit Trust and ZTA Controls

- Implicit trust is when you trust something without verification:

  "In this new paradigm, an enterprise must assume no implicit trust and continually analyze and evaluate the risks to its assets and business functions and then enact protections to mitigate these risks."

  NIST SP 800-207 Section1, p.1

- Note that it doesn't say:
  - Users
  - Managed endpoints
  - Endpoints, servers, cloud workloads
- If you are limiting implicit trust to users and managed devices, you are **missing the point**



Implicit Trust and ZTA Controls

- This is an incomplete list of "assets" we should be minimizing implicit trust in:

Managed endpoints	Unmanaged endpoints	BYOD	WiFI Infrastructure	Security agents
Servers (physical)	Servers (virtual private)	Cloud Workloads (IaaS)	Smart Phones	Tablets
Switches	Routers	Firewalls	Public cloud infrastructure	Private Cloud Hypervisors
Load Balancers	Printers	Photocopiers	Application Software	Operating Systems
Document Centers	HVAC	BMCs	Video Cameras	Peripherals
IoT	OT	ICS/SCADA	Drivers`	Etc. etc. etc.

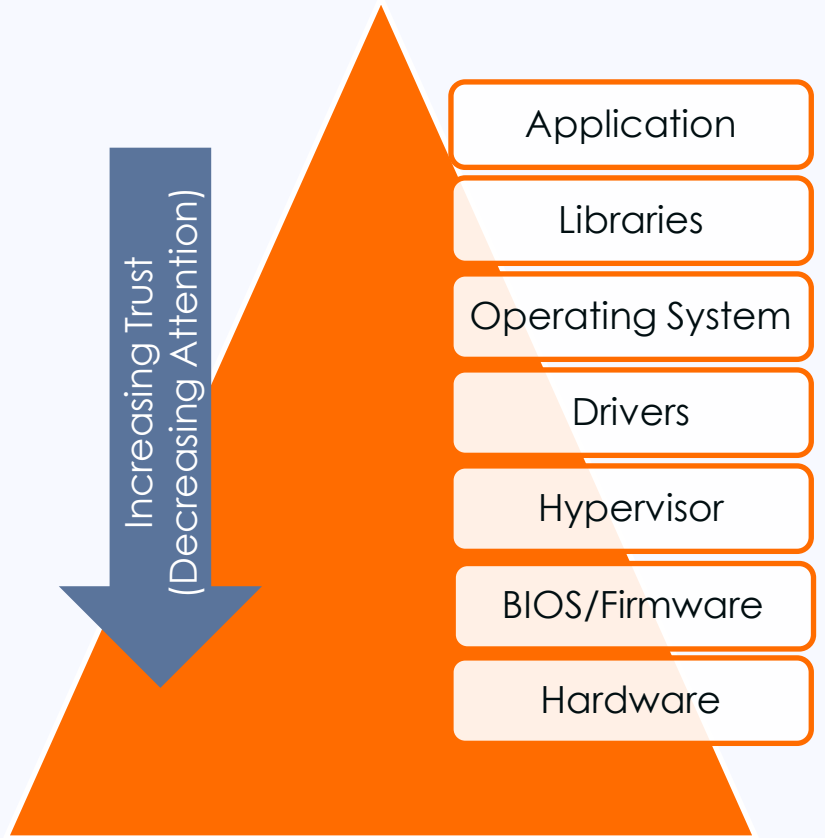
- **Interesting question to ponder**: how do we avoid doing implicit trust in the policy engine?
  - Replicated policy engines (homogenous/heterogenous/interlocked?)
  - Traditional "high assurance" techniques for isolating it as much as possible (cross-domain solutions etc,)
- **"Think like an attacker"**: given that predictive AI and machine learning will be so critical to a mature policy engine, how do we prevent data poisoning attacks?

Implicit Trust and ZTA Controls

- And don't forget our users, staff, suppliers, customers, visitors etc.
  - Don't limit the "identity" pillar just to the contents of your IDM
- It's not just "assets", we need to consider provenance
  - Vendor
  - Software and hardware components (and sub-components)
  - Delivery (interdiction attacks), installation, deployment
  - Ability for a government to force the vendor to act on their behalf
- Did anyone say "supply chain risk"? 😉
- The only rational approach is to consider everything potentially compromised, and build the ZTA infrastructure to detect compromised assets

The Risk of "Trusting Down"

- Blindly trusting down as been axiomatic for years
  - This is literally a form of implicit trust
  - It's also been literally exploited for years by malware, and while there are techniques to prevent it, they're rarely deployed outside narrow fields like DRM.
- Note that NIST SP 800-207 draws the boundary at the device edge, which is probably pragmatic but not optimal
  - Ian's sub-assertion: everything in ZTA is a maturity model 😉
- Even finding programmable attack surface, let alone securing it, is quite difficult
- Further details: see my MILCIS 2015 presentation on "Securing IoT"



Hardware Attack Surface: Intel Celeron N5101

Advanced Technologies	
Intel® Gaussian & Neural Accelerator	2.0
Intel® Image Processing Unit	6.0
Intel® Smart Sound Technology	Yes
Intel® Wake on Voice	Yes
Intel® High Definition Audio	Yes
Intel® Optane™ Memory Supported ‡	No
Intel® Speed Shift Technology	Yes
Intel® Turbo Boost Max Technology 3.0 ‡	No
Intel® Turbo Boost Technology ‡	No
Intel® Hyper-Threading Technology ‡	No
Intel® 64 ‡	Yes
Instruction Set	64-bit
Instruction Set Extensions	Intel® SSE4.2
Idle States	Yes
Enhanced Intel SpeedStep® Technology	Yes
Thermal Monitoring Technologies	Yes
Intel® Identity Protection Technology ‡	Yes
Intel® Smart Response Technology	No

Intel® Gaussian & Neural Accelerator

Intel® Gaussian & Neural Accelerator (GNA) is an ultra-low power accelerator block designed to run audio and speed-centric AI workloads. Intel® GNA is designed to run audio based neural networks at ultra-low power, while simultaneously relieving the CPU of this workload.

Intel® Image Processing Unit

Intel® Image Processing Unit is an integrated image signal processor with advanced hardware implementation that improves image and video quality of cameras.

Intel® Smart Sound Technology

Intel® Smart Sound Technology is an integrated audio DSP (Digital Signal Processor) built to handle audio, voice, and speech interaction. It allows the latest Intel® Core™ processor-based PCs to respond to your voice command quickly and offer high fidelity audio without impact system performance and battery life.

Intel® Wake on Voice

Intel® Wake on Voice allows your device to wait and listen for your command without consuming excessive power and battery life, as well as wake from modern standby.

Intel® Identity Protection Technology *

Intel® Identity Protection Technology is a built-in security token technology that helps provide a simple, tamper-resistant method for protecting access to your online customer and business data from threats and fraud. Intel® IPT provides a hardware-based proof of a unique user's PC to websites, financial institutions, and network services; providing verification that it is not malware attempting to login. Intel® IPT can be a key component in two-factor authentication solutions to protect your information at websites and business log-ins.

Zero Trust "Strategy" – The Case for Robust Telemetry

- Conclusions: trust NOTHING
- Understand modern devices might have programmable attack surface you don't even know if there
- Everything can be compromised by a sufficiently resources and motivated threat actor
  - Including our service providers
  - Including our ZTA controls and infrastructure
- Expect data poisoning attacks targeting our AI/ML
- We need to monitor EVERYTHING
- Hence, resilient and reliable telemetry is essential
  - Actually, we need to have telemetries – from as many eVRF "visibility surfaces" as possible
  - Then we have a chance

Assertion #3

Sensor location matters

Sub-assertion #3A

A sensor inside the blast radius of a compromise is potentially compromised!

Sub-assertion #3B

Water is wet.

Assertion #4

It is important to understand the trustworthiness of your telemetry – telemetry assurance matters

Where this Concept Started – Evaluating the Assurance of Logging

At the log source

- The cyber-relevant event must be noticeable by the "system"
- It must have been anticipated by the code developer
- The code developer must have coded it into the software
- Logging must be configured correctly – with the logging level set to generate the log

In Transit

- The event must be sent from the log source to the log collection system without loss, duplication or modification

At the log collector

- The log message must be ingested successfully
- It must be parsed into an informational taxonomy
- That taxonomy must allow it to be understood (in isolation or with other messages) to map to the cyber-relevant event
- An alert to a person or system must occur, which allows the event to be managed
- That person or system has to actually do it to remediate the risk

Telemetry Assurance Evaluation Framework

Objective

- The aim of this framework is not to stack rank, but to provide an understanding of the assurance level provided by a specific telemetry type
- It needs to be applied to a specific architecture and deployment
- Where gaps or deficiencies are identified, other forms of telemetry can be deployed to compensate
- Reminder: predictive AI is very good at correlating expected behaviors across multiple data types vs. time
- Does not consider cost (cost is not assurance)

Criteria

- **Reliability**: How comprehensive/ accurate is the telemetry?
- **Evadability Resistance**: How readily can an attacker evade detection in the telemetry produced?
- **Resilience**: Is it possible to detect the evasion?
- **Stealthiness**: Are the controls covert or detectable by a threat actor?

Evaluating Traditional Logging (e.g. Syslog) from a Linux Cloud Workload

Example of use

Criteria	Evaluation	Commentary
Reliability	Low/Medium	See the previous slide.
Evadability Resistance	Low	Disabling/degrading/spoofing logging is a standard attacker TTP (T1070.001, T1070.002, T1070.003, T1562.002, T1562.003, T1562.008) and can be done from inside the workload.
Resilience	Low/Medium	While techniques exist to detect attacks on logging, their efficacy is low, producing both false positives and negatives.
Stealthiness	Low	Logging is configured inside the workload. An attacker who has compromised the workload will be aware that it is logging.

Evaluating Agent-Based Network Traffic Access in a Public Cloud

Example of use

Criteria	Evaluation	Commentary
Reliability	High	Delivers all traffic to and from this workload
Evadability Resistance	Low	Agent can be disabled by an attacker who has compromised the workload.
Resilience	Low/Medium	While cessation of traffic streams can be detected, a very sophisticated attacker could theoretically mask or filter out their traffic, or even traffic record or spoof.
Stealthiness	Low	The running agent is visible to an attacker who has compromised the agent

Evaluating Agent-Based Network Traffic Access in a Public Cloud

Example of use

Criteria	Evaluation	Commentary
Reliability	High (If available)	Delivers all traffic to and from this workload. (But… not available in all Cloud Service Providers)
Evadability Resistance	High	Configured in the cloud management environment, which cannot be disrupted from the workload
Resilience	High	See above
Stealthiness	High	Not visible from the workload

Closing Thoughts

Closing Thoughts

- This is the start of a framework, which we believe has merit
  - In the context of CISA's eVRF – we developed this and proposed it to them in their response
  - In the context of Zero Trust Architectural Design
- Note that cost is a real-world consideration which is not included in the current framework
  - Should be be? It's not actually an assurance consideration.
- Low/Medium/High remain "thumb in the air" for the moment, but seem to work well enough for this framework to be useful
- Two most useful outcomes from this framework:
  - Gap identification
  - Minimization of "familiarity bias" in architectural design

What Does This Mean for Zero Trust?

- There are typically 3.5 forms of telemetry used for ZTA
  - Metrics, Events. Logs and Traces (MELT) – but mostly logs. These come from OS, applications, services and appliances
  - Security Workload Agents (e.g. EDR) – security focused applications which provide endpoint assurance telemetry
  - Network Behavioral Telemetry (e.g. NDR) – external observations of the behavior of a workload or service
  - Telemetry from APIs and things like eBPF – external to the workload, but reporting on the workload
- So which to choose? As many as you can! Defense in depth!
- Loss of telemetry, or deviations from normal, indicate:
  - One telemetry deviates: possibility that the telemetry sensor has been compromised
  - More than one deviation: likely presence of atypical behavior corresponding to a risk

What Does This Mean for Zero Trust?

(Continued…)

- Ian's unpopular opinion: anyone not planning a policy engine but claiming they're doing "Zero Trust" in 2023 is being dishonest
- These are going to require huge ingest and processing rates, and the ability to scale AI/ML models (everything from DNNs, SVMs even things like Bayesian inference) is essential
  - Our tradition SIEMs aren't going to cut this
  - Consider that telemetry sources and volumes will evolve over time
  - Consider the need to run AI/ML over historical data as well as telemetry streaming in
- Consider the risk of data poisoning – typically this will be from a single telemetry source, and you could consider blocking AI/ML learning off that while you clean up the attack

Thank You

Ian Farquhar
ian.farquhar@gigamon.com
+61 437 992 219

@IanAtGigamon
https://www.linkedin.com/in/ianfarquhar/

Gigamon®