Predictive Analytics

Network Awareness and Predictive Cyber Analytics … For the Endpoint

Carter Bullard

Wednesday, July 29, 2020

openargus.org

qosient.com

QoSient

QoSient

Carter Bullard Founder, CEO. 2000-2020

- Academic research at FFRDC's, US National Laboratories and Industry
  - DHS – Elimination of Unmonitored Space 2017-2020
  - Co-PI – NSF Advanced Measurement Initiative (IRC-AMI) 2016-2019
  - Network Technical Lead – NSA Cyber Pilot Program 2012-2014
  - NSA Network Modernization Program 2009-2012
  - Principal Research Scientist – NSA, NRL 2006-2010
  - Director/Principal Engineer – FORE, Bay, Nortel. 1994-2000
  - Research Scientist – CMU SEI CERT 1989-1994
- Inventor of IP Flow Monitoring (commonly called "netflow")
- Directed 1st Network Forensics Research Lab SEI/CERT – 1990-1994
- ATM network security standards for the ATM Forum, and ITU 1996-2000
- Authored 20+ papers, 40+ conference presentations, 4 Patents

Problem Statement

- A primary threat to DHS enterprises is unauthorized data access and data exfiltration initiated from inside the enterprise.
- Endpoints are susceptible to malware exploitation through a number of means, creating bastion points for adversarial activity inside an enterprise.
- Remote and mobile (laptop) assets provide real opportunity for undetected malware introduction and exploitation, despite the use of CDM, HBSS, EDR style protection.
- CISA interim telework guidance, Agency Teleworker Option 1 – direct cloud provider access creates new challenges.

Job for Predictive Analytics

- Need to get ahead of the game and provide heads up to the opportunity for significant attack.
- If we knew that a remote / mobile node had interacted with a bad-actor or exhibited bad-actor or victim behavior while outside the enterprise, we would be able to realize its impact on the organization.
- Can we estimate the cyber potential of a network asset based on its historical behavior ??? We think yes !!!
- How do you realize this with todays technology …
  - Data, data and more data … but not just any data

QoSient

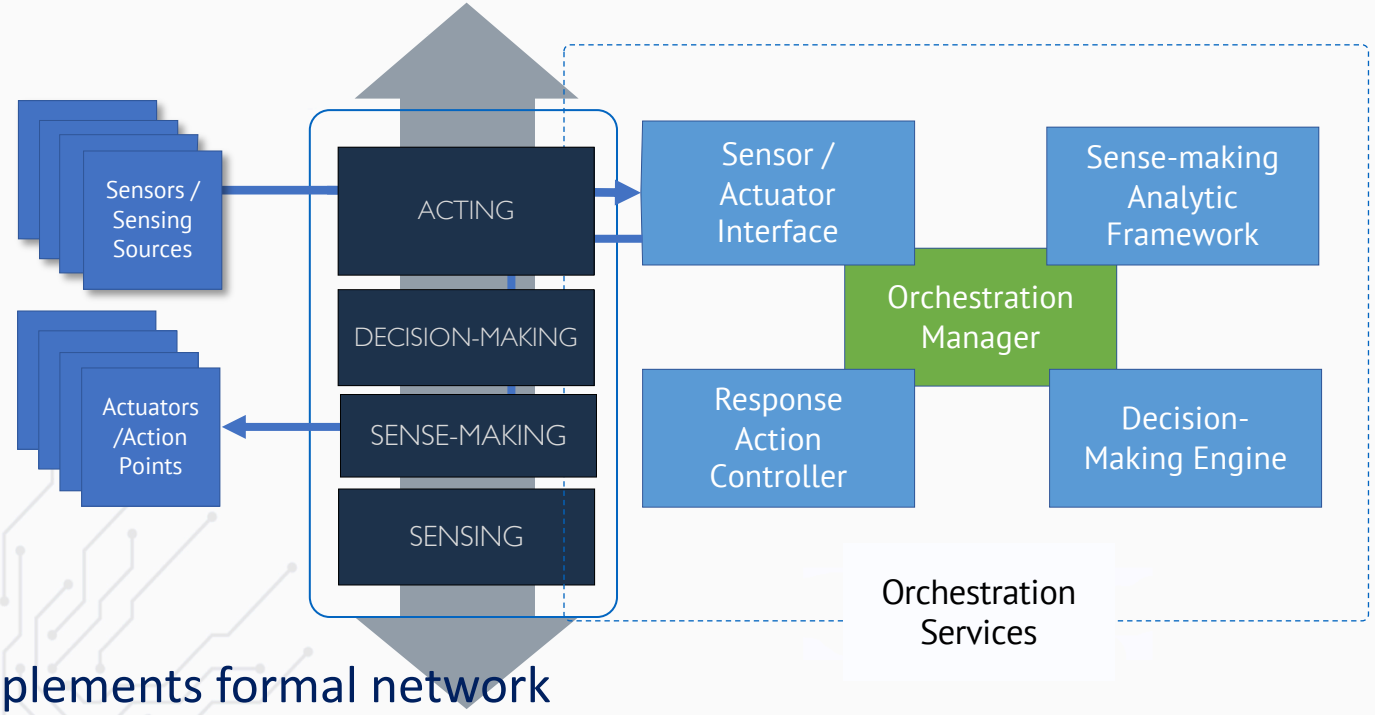
Prior Work

- DHS-OIG 'Elimination of Unmonitored Space' (EUS) Program
  - Internal Pilot to Establish Comprehensive Network Activity Information System For Enterprise Awareness
    - Modeled after the NSA's Integrated Active Cyber Defense (IACD) architecture and the US DoD CENTAUR / Acropolis programs.
    - Using argus, network flow audit system first developed at CMU SEI CERT and used operationally in US Gov't cyber security programs of record (DISA NGS/NGX/JRSS) for data generation (**network sensing**), collection, and processing (**network sense-making**).
  - Develop Sense-Making and Improve Decision-Making through inventory, categorization, baselining, classifying and comparing all network activity.
  - Enable the awareness to support identification, protection, detection, response and recovery for cyber incidents within the complete DHS-OIG enterprise.

QoSient

IACD Baseline Reference Architecture

- **Captures transition from Sense-Sense-making-Decision-making-Acting Construct to capabilities-based architecture**



Our approach implements formal network sensing and sense making as independent distributed functions, allowing for sensing at the edge.

QoSient

ARGUS

US DoD CENTAUR / Acropolis

- "Acropolis collects, stores, and analyzes network traffic on the Non-classified Internet Protocol Router Network (NIPRNet), Secret Internet Protocol Router Network (SIPRNet), and DOD enterprise services in an effort to detect and deter America's enemies."
- This is much different from conventional cyber security products that want to observe, analyze, detect and mitigate all in one thingy …
- "The environment runs on an infrastructure called CENTAUR. CENTAUR is a repository of detailed NIPRNet traffic data that enables analysts to visualize cyber activity over the past 18 months."
- The approach is all about the data, which is critically important to enabling effective cyber response.

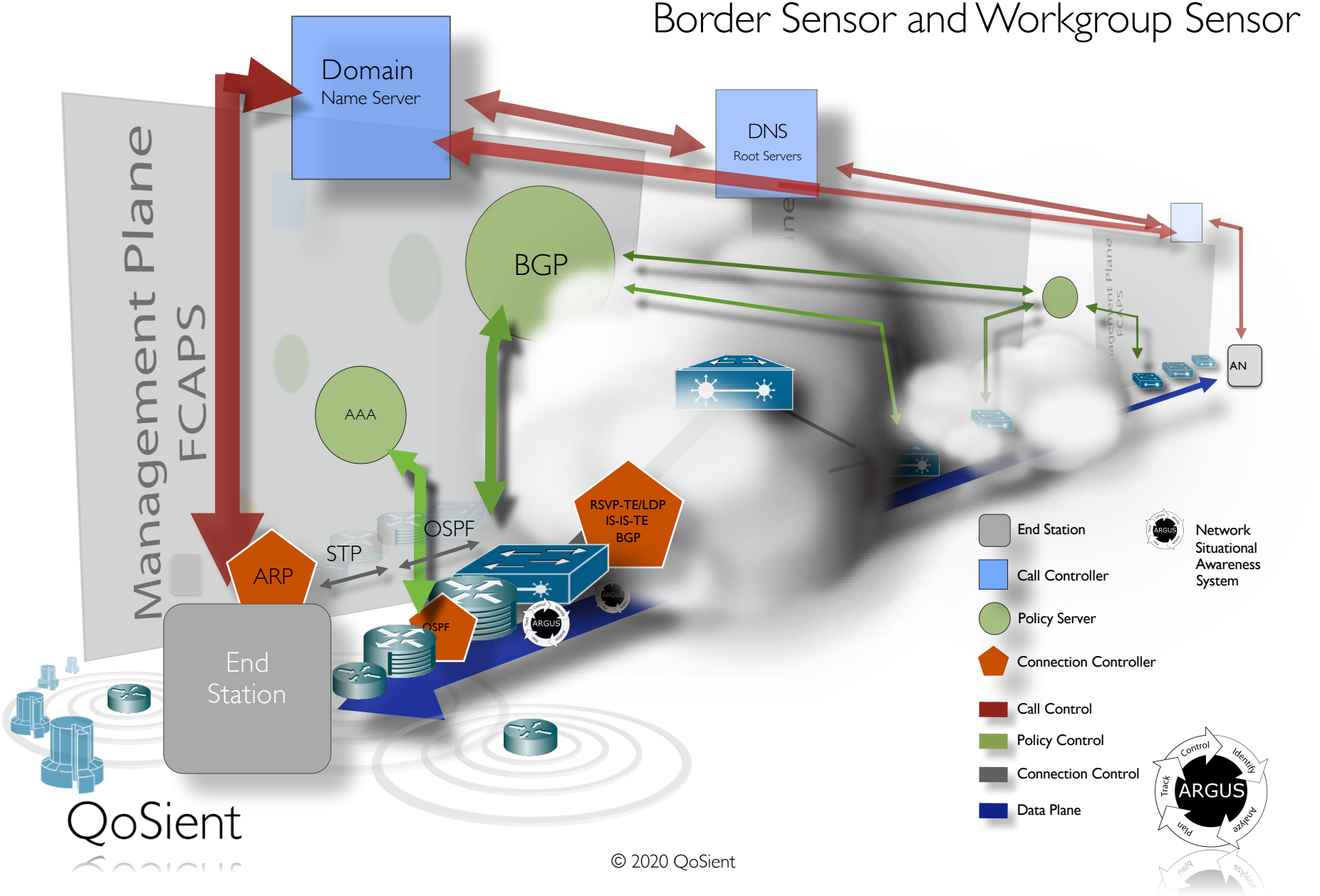
QoSient

Expanding to the Edge

QoSient

Traditional Visibility Deployment Strategies

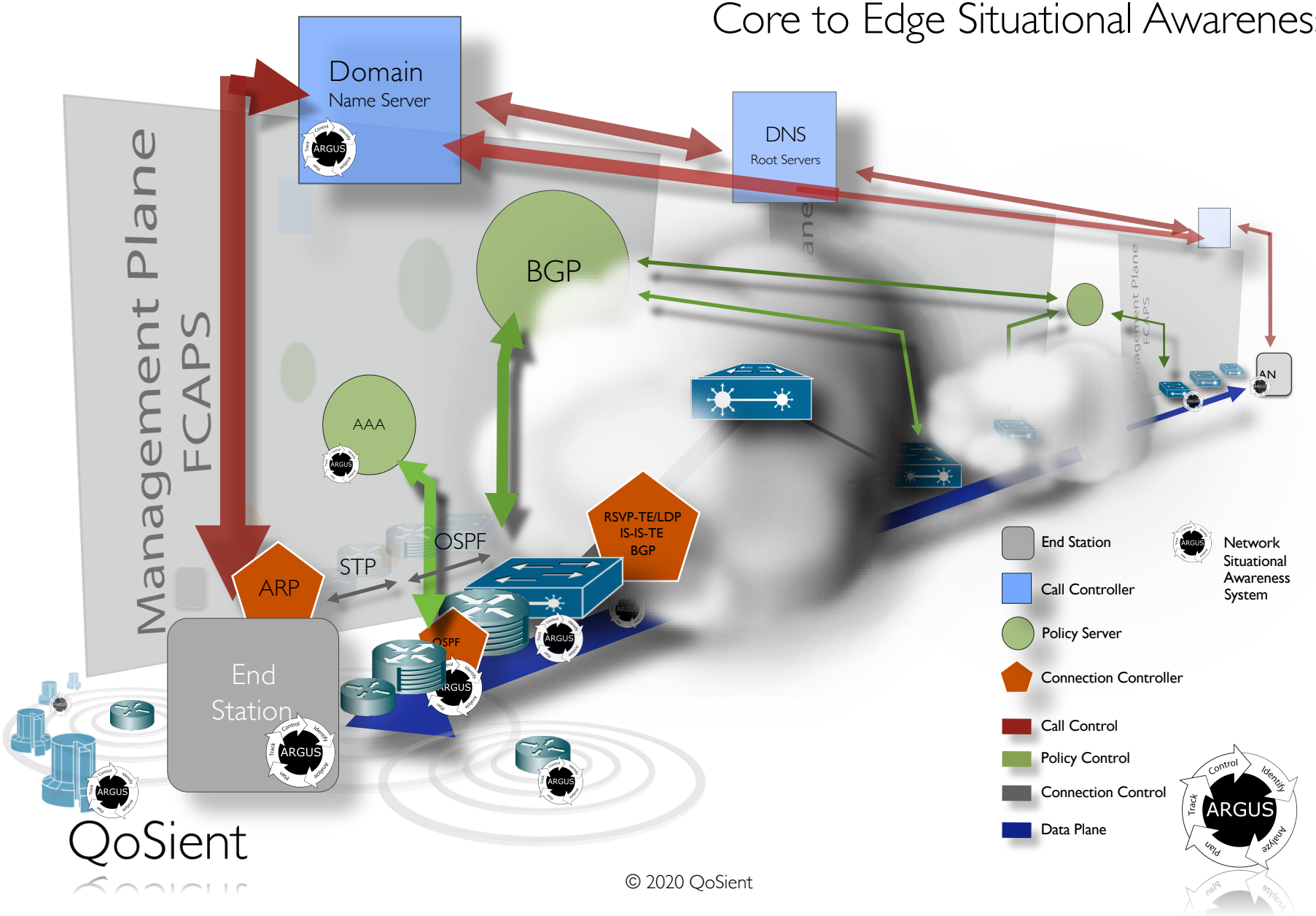
Border Sensor and Workgroup Sensor



QoSient

Argus Deployment Strategies

Core to Edge Situational Awareness



QoSient

Expanding to the Edge

- Going to the Edge falls into the Endpoint Detection and Response (EDR) sector of the cyber security marketplace.
- To add new capabilities to EDR technology, any resource impacts on the end device MUST be minimal.
  - Few cycles and no additional battery life is available.
  - No one wants another agent
- Our approach is to separate advanced network flow data generation within endpoints from data processing.
  - Upload data to an audit and analytics information system.
  - Establish analytics framework around large data archive.
  - Perform lazy IDS and behavioral anomaly detection.
  - Enable ML model development, feature extraction, testing and deployment.

QoSient

Expanding to the Edge

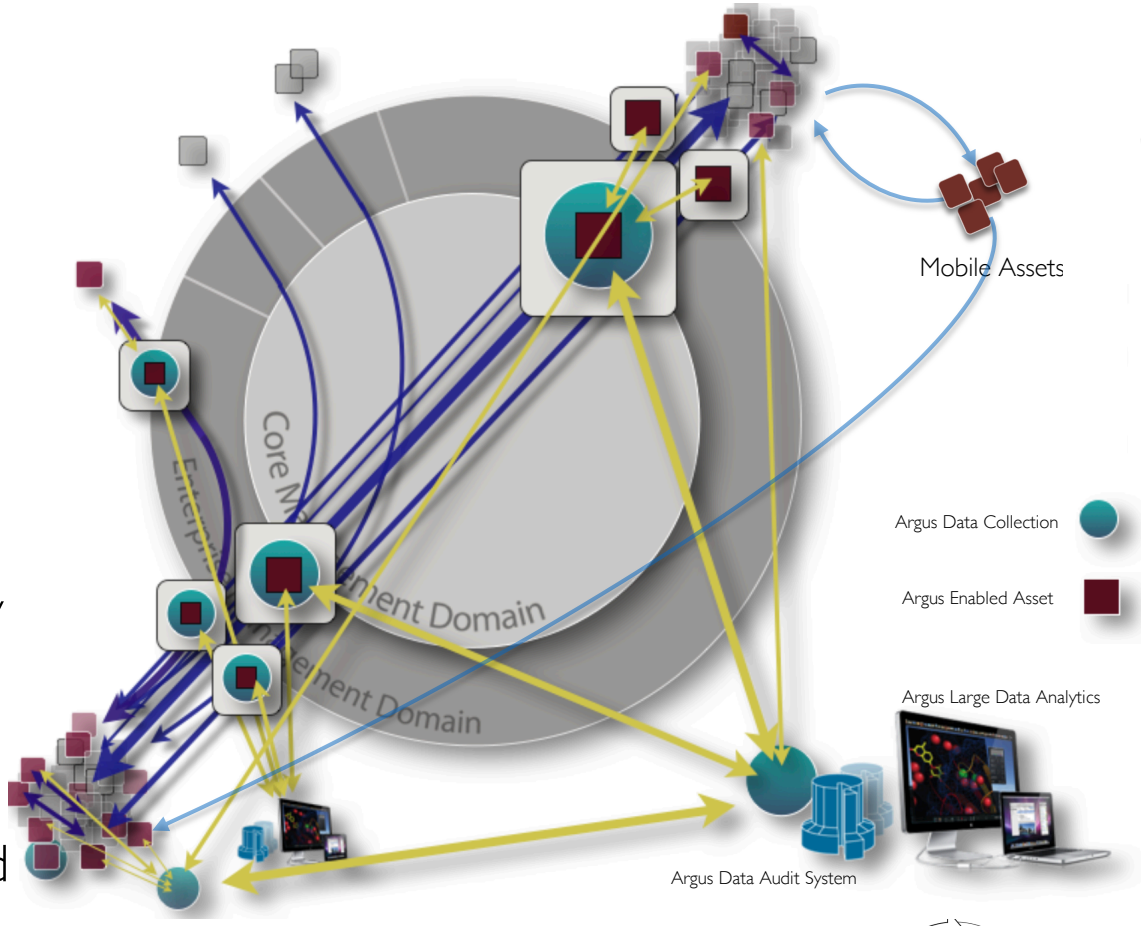
Primary Benefits of End System Deployment

- Network Forensics Data Generation
  - Whether data is collected or not, having a log of every network interaction as a network log is a huge improvement over existing strategies
  - Provides contextual visibility regardless of where the end system is.
  - Provides basic data needed for analytics like network activity baselining which is a requirement for network behavioral anomaly detection.
- Best opportunity for seeing Shadow IT, tunneled access, covert entry points into the enterprise.
- Opportunity to provide near-real time access to network activity data in the enterprise
- Cost effective strategy to turn existing assets into cyber sensors.

QoSient

Data - Analytics Framework

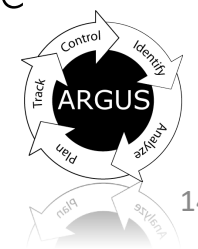
- Advanced network forensics data generated within mobile assets (laptops).
- On-demand and scheduled distributed collection with centralization for data processing and management.
- Large data analytics platform for asset context tracking, behavioral baselining, anomaly detection, deferred IDS, and intelligence assessment, and **Predictive Analytics.**
- Web-based dashboarding and data navigation.



QoSient

DHS Operational Pilot

- The operational pilot has been successful demonstrating that a network sensing and sense-making data approach can enable new capabilities for the component.
- EUS program has introduced new cyber security analytics for incident identification, detection and response.
  - SOC Notification investigation and response (ESOC, SEN)
  - Malware Identification, Analysis and Response (Kofter)
  - Behavioral Anomaly Detection (Resource Use, Lateral Movement)
  - Critical Infrastructure Architecture Assurance (Rogue Server, Access)
  - Continuous Access Control and Acceptable Use Policy Verification (DNS)
- We have identified a class of threat from mobile assets (laptops) where external compromises have threatened the enterprise.
- Real need to 'know' what mobile assets experience while outside the protective DHS environment (forensics analysis, attribution).



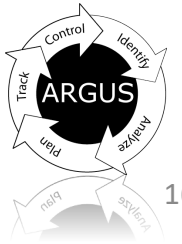
QoSient

Predictive Analytics

- Mobile assets that are attacked, or express bad actor or victim behavior, have a predictable impact when they return.
- Discreet and continuous predictive analytics and machine learning (ML) to identify **intrusion** and **anomalous network behavior** for endpoints while outside the enterprise.
  - The operational large data approach provides seed data for behavioral baselining, anomaly detection and predictive analytic development.
  - Training features include network context and behaviors such as used ports and protocols used, communicating peers, services and resources used (loads, demand, volume), session dynamics, geo and net-spatial metrics, producer / consumer ratios, DHCP and DNS usage.
  - Assessments against deviations from historical traffic baselines, tailored 3rd party intelligence data and whether any observed traffic challenges or violates known enterprise access control and acceptable use policies.

Predictive Analytics

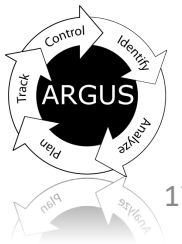
- EUS Cyber Data Model
  - Argus used to develop the **UNSW-NB15** (2015) intrusion detection dataset which is heavily used in AI and ML cyber security research (935 research articles in Google Scholar)
  - Oak Ridge National Laboratory (ORNL) uses real-time streaming argus data in their operational SITU unsupervised anomaly detection system.
  - NSF Advanced Measurement Initiative (IRNC AMI InSight2) uses argus data to develop multi-stage Markov chain modeling for predictive operational awareness.
- Use predictive analytics to assess the cyber security threat potential of mobile assets when they access or return to the enterprise network.
  - Multiple endpoint classification schemes that contribute to an ML assisted asset cyber threat score. (known end point experienced unfriendly behavior with unknown entity in new network context without subsequent change in network behavior).
  - Markov chain modeling to estimate / predict a cyber threat posture index for each mobile asset (nothing bad, no problem, some bad things, real problem)
  - Use threat posture index and predicted "next states" to set cyber threat expectations for asset while accessing enterprise assets.



Predictive Analytics

- What do we want to observe as predictive indicators
  - Any indication of successful intrusion
    - STIX/TAXI signature analysis
    - Control Plane Manipulation (DNS, DHCP, ARP exploits)
    - Association with other Bad Actors (Reputation)
  - Indications of bad actor behavior
    - Exfiltration Behavior
    - Covert Channel Use
    - Discovery / Lateral Movement Behavior
  - Indications of victim behavior
    - Permissive access
    - Stepping Stone / Split Tunnel Behavior
    - Beaconing

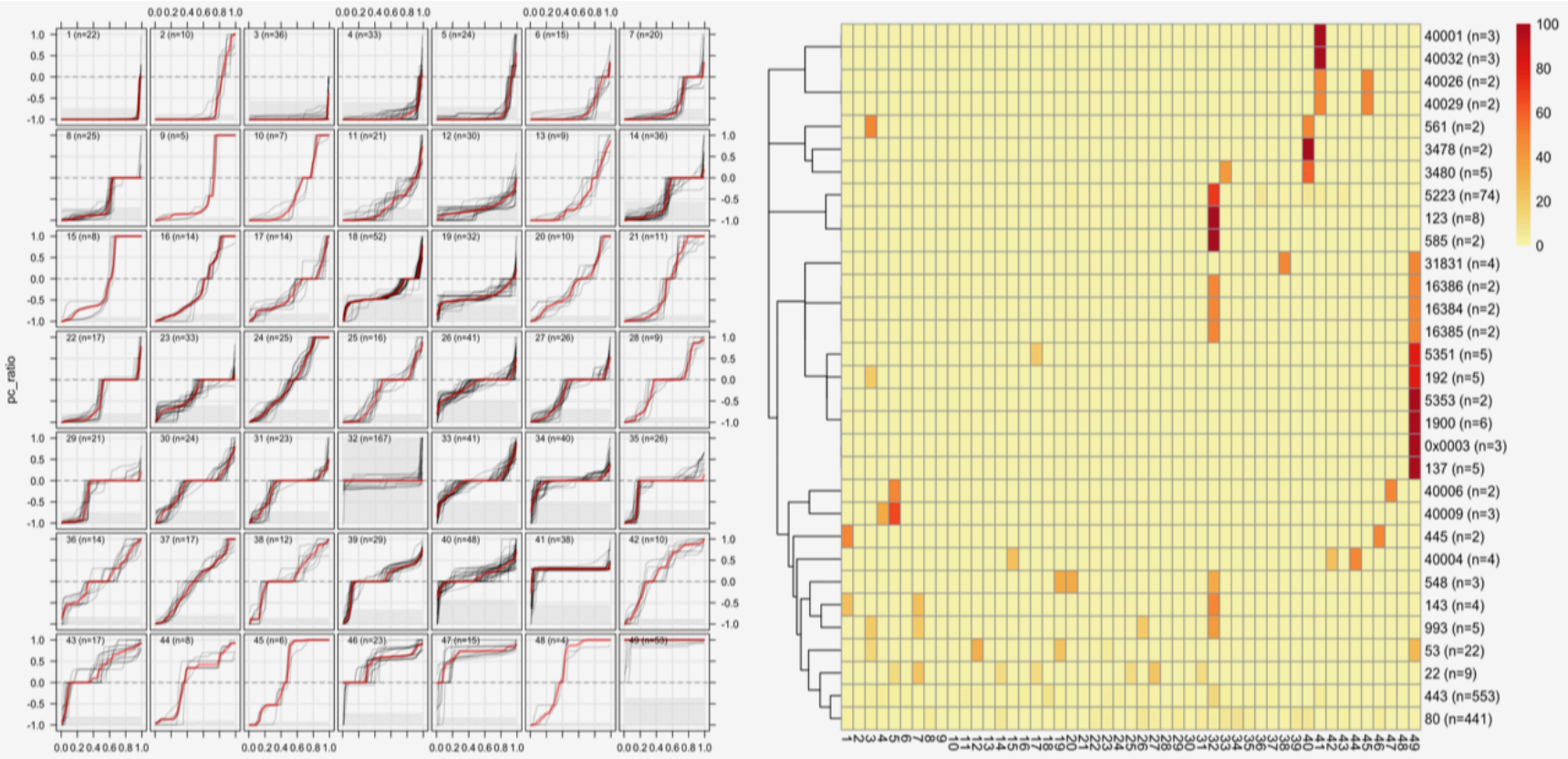
QoSient



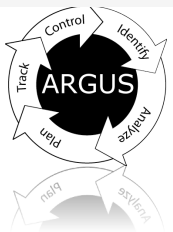
Exfiltration Behavioral Baselining

PCR Traffic Classification Analysis

Enterprise Services



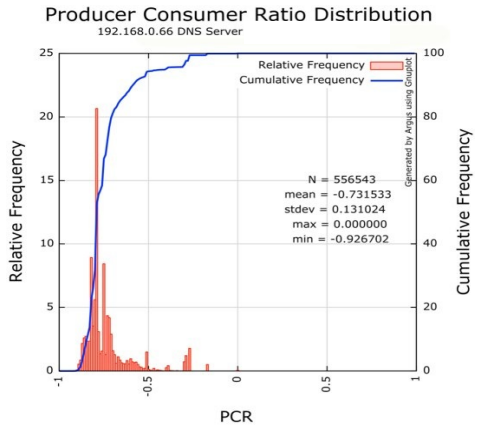
QoSient

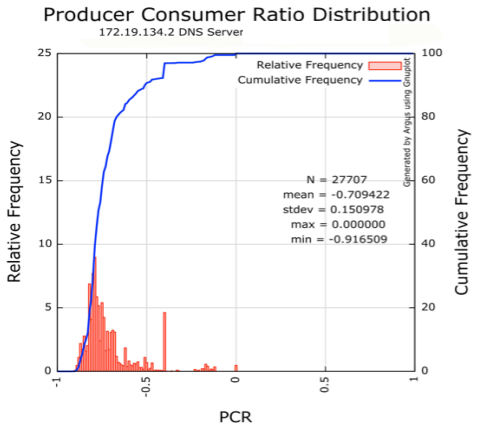


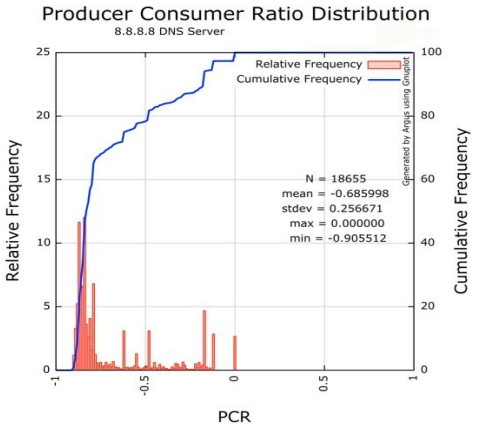
Service Behavioral Baselining

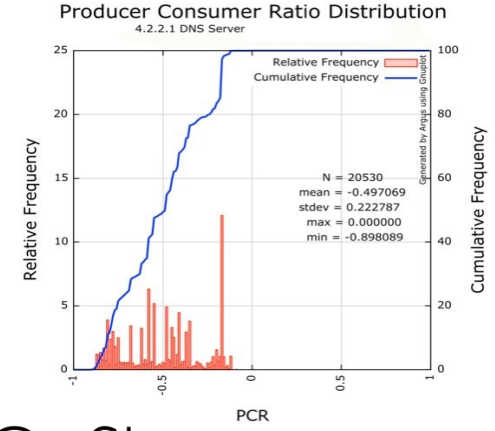
PCR Application Analysis

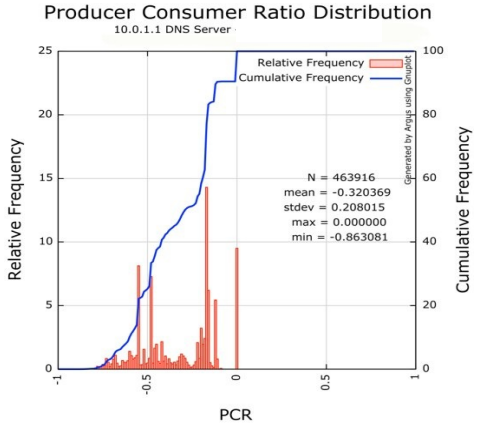
Domain Name Servers









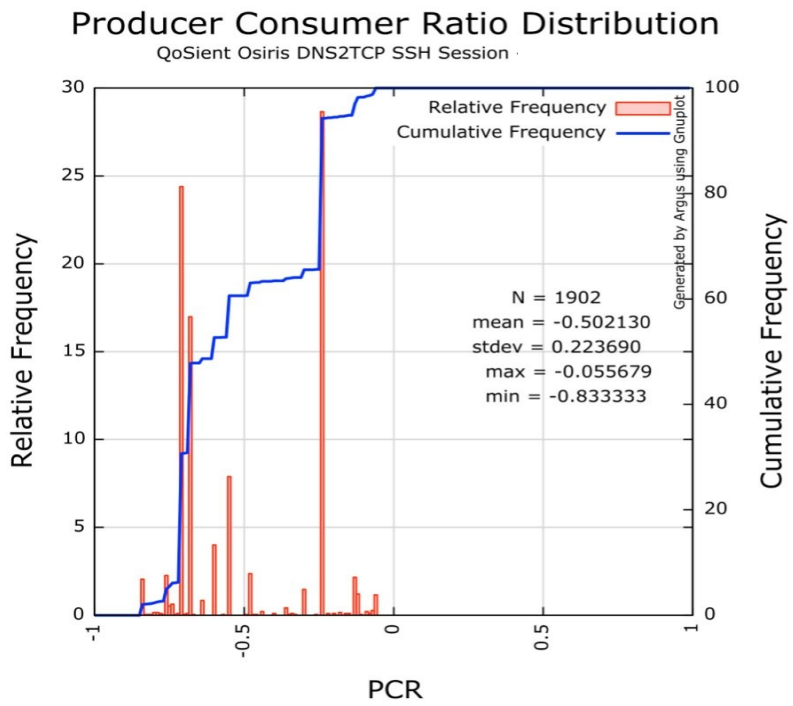


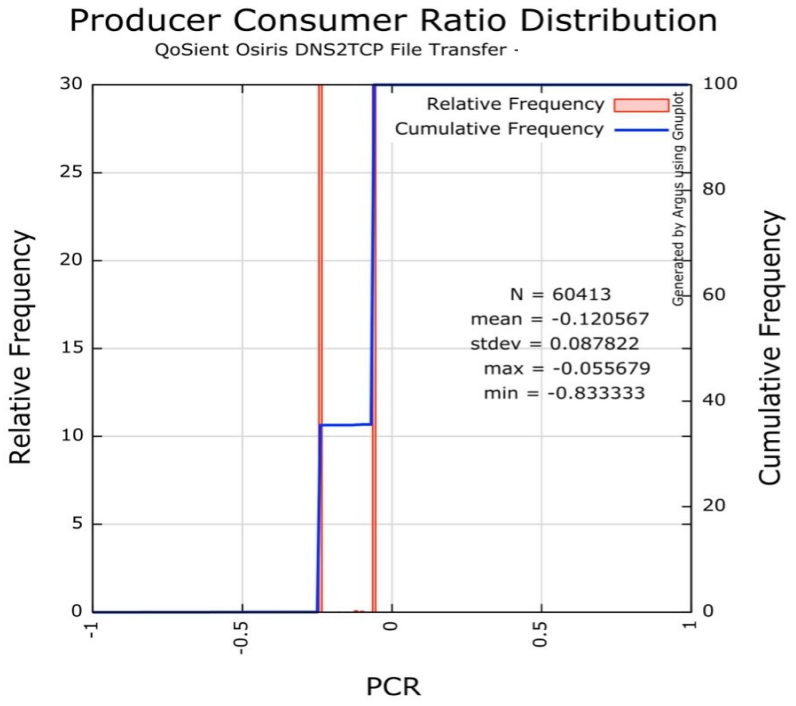
QoSient

Behavioral Anomaly Detection

PCR Covert Channel Analysis

Domain Name Service - dns2tcp



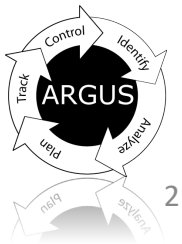


Farnham, G. and Atlasis, A. *Detecting DNS Tunneling,* SANS Institute InfoSec Reading Room, Feb, 2013.
https://www.sans.org/reading-room/whitepapers/dns/detecting-dns-tunneling-34152

QoSient

What do we now

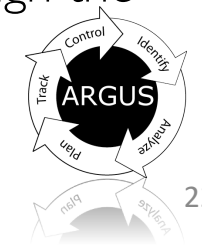
- Predictive analytic scoring needs to be available to the traditional enterprise cyber security management framework.
  - Create Alarms / Alerts for scoring changes
  - APIs to fetch cyber posture scoring for Layer 2, 3 addresses and names
  - Provide query capability to find nodes that are in trouble
- Dashboard visibility to see scoring and anomalies
- Data browsing capabilities to understand scoring changes
- Human in the loop technology to insert human modifications to automated scoring and actions.
- SOAR API integration to provide complete automatiom.



QoSient

Conclusions and Invitation

- The DHS-OIG EUS pilot approach is a great operational foundation for developing cyber security predictive analytics.
  - Establishing a network sensing and sense-making capability for endpoints that feeds a network audit data analytic platform is an innovative approach to delivering operational network security awareness analytics.
  - The approach creates a corpus of operational network awareness data that developers can use to build, model, and test site-specific cyber security analytics, rather than just generating alarms.
- The remote / mobile device is vulnerable when outside the protective enterprise, and will be challenged, possibly compromised, and may express the only evidence of compromise when outside the enterprise.
- Predictive analytics that can assess the threat potential of these devices based on the assets network experience addresses a huge visibility gap, that could be extended to all endpoints in the enterprise.
- We invite your involvement to help bring this capability to life through the open source effort.



QoSient