Network Flow Data Fusion

GeoSpatial and NetSpatial Data Enhancement

FloCon 2010
New Orleans, La

Carter Bullard
QoSient, LLC

carter@qosient.com

GIGEF
Global Information Grid
Evaluation Facilities

LD
Large Data JCTD

Naval Research Laboratory
Washington, D.C.

ARGUS
Control
Identify
Analyze
Plan
Track

Carter Bullard   carter@qosient.com

- QoSient - Research and Development Company
  - Naval Research Laboratory (NRL), GIG-EF, JCTD-LD, DISA, DoD
    Network Performance and Security Research
- Inventor/Developer Argus   http://qosient.com/argus
- FBI/CALEA Data Wire-Tapping Working Group
- QoS/Security Network Management - Nortel/Bay
- Security Product Manager – FORE Systems
- CMU/SEI CERT
  - Network Intrusion Research and Analysis
  - NAP Site Security Policy Development
  - Network Security Incident Coordinator
- NFSnet Core Administrator (SURAnet)
- Standards Efforts
  - Editor of ATM Forum Security Signaling Standards
  - IETF Working Group(s) Contributor
  - Internet2 Security WG
  - NANOG

FloCon 2010 Flow Innovation

This year's conference will focus on flow data analysis within the context of other data sources. Presenters are encouraged to consider how flow is a piece of the puzzle.

Which Puzzle?

- Cyber-Situational Awareness and Network Defense (CND)
  - Near real-time awareness of threats, status, and performance, with awareness of external attacks and insider abuse/misuse.
- Assured Enterprise Management and Control
  - Critical infrastructure must operate as intended, with management, control and information assured.

Cyber-Situational Awareness

Level 1 SA - Perception

- The perception of elements in the environment within a volume of time and space
- Involves timely sensing, data generation, distribution, collection, combination, filtering, enhancement, processing, storage, retention and access.

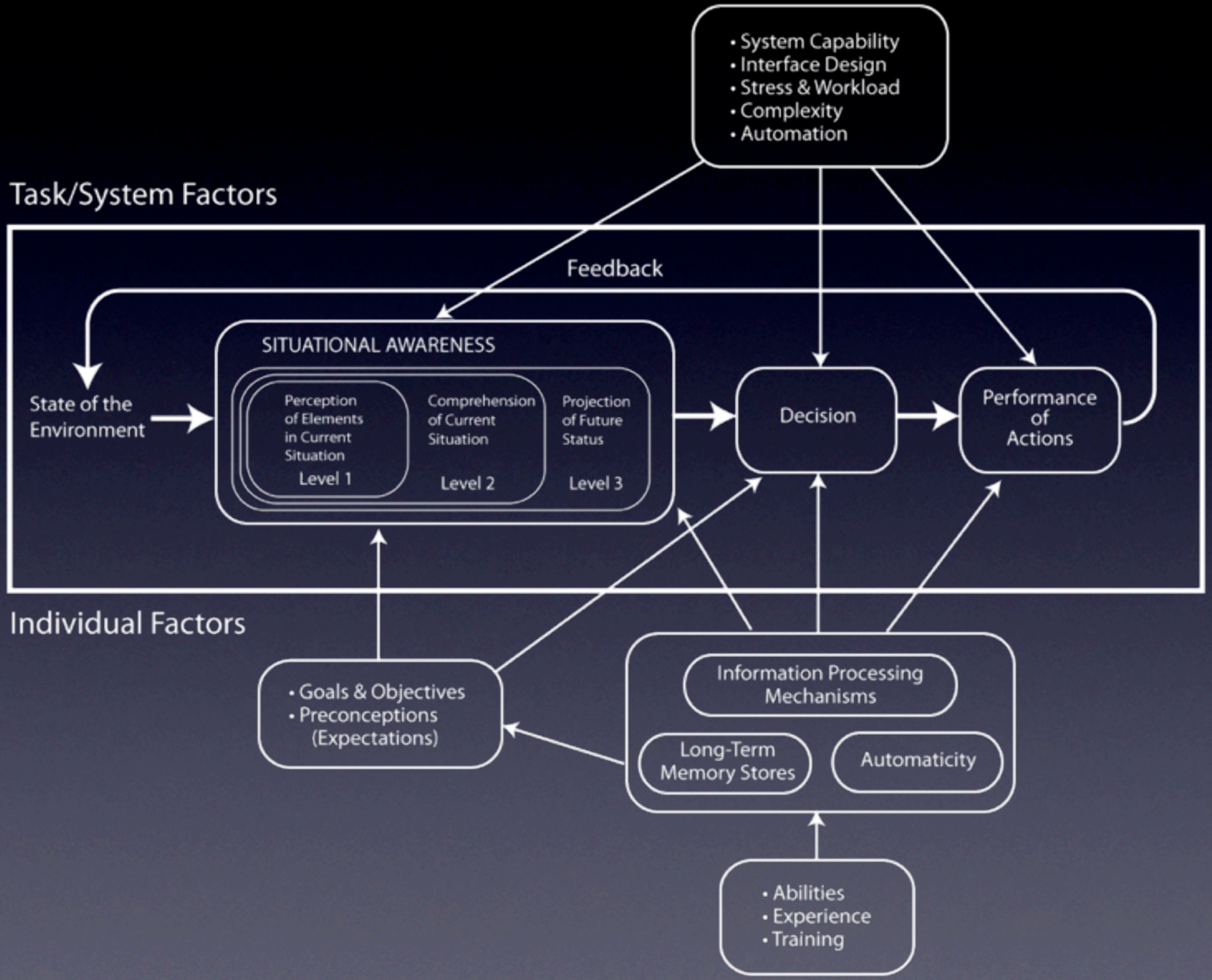
Level 2 SA - Comprehension

- Understanding significance of perceived elements in relation to relevant goals and objectives.
- Involves integration, correlation, knowledge generation.

Level 3 SA - Projection of Future Status

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), 32-64.

Model of Situational Awareness in Dynamic Decision Making



- System Capability
- Interface Design
- Stress & Workload
- Complexity
- Automation

Task/System Factors

Feedback

SITUATIONAL AWARENESS

State of the Environment

Perception of Elements in Current Situation Level 1

Comprehension of Current Situation Level 2

Projection of Future Status Level 3

Decision

Performance of Actions

Individual Factors

- Goals & Objectives
- Preconceptions (Expectations)

Information Processing Mechanisms

Long-Term Memory Stores

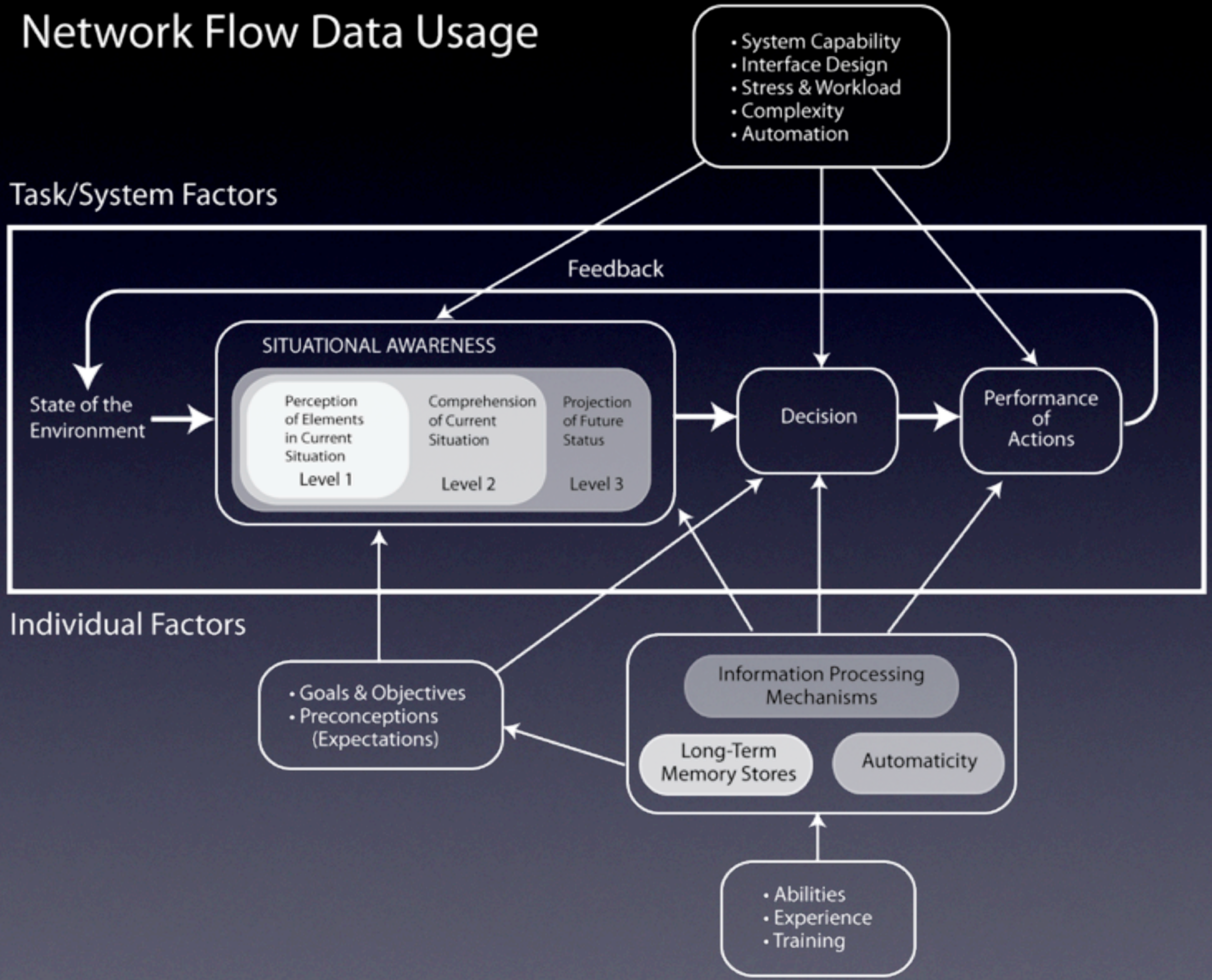
Automaticity

- Abilities
- Experience
- Training

Endsley, M. R., ***Human Factors***, Vol. 37, No. 1, 1995. Copyright 1995 by the Human Factors and Erogonomics Society. All rights reserved.

Model of Situational Awareness in Dynamic Decision Making

Network Flow Data Usage



- System Capability
- Interface Design
- Stress & Workload
- Complexity
- Automation

Task/System Factors

Feedback

SITUATIONAL AWARENESS

State of the Environment

Perception of Elements in Current Situation
Level 1

Comprehension of Current Situation
Level 2

Projection of Future Status
Level 3

Decision

Performance of Actions

Individual Factors

- Goals & Objectives
- Preconceptions (Expectations)

Information Processing Mechanisms

Long-Term Memory Stores

Automaticity

- Abilities
- Experience
- Training

Endsley, M. R., ***Human Factors***, Vol. 37, No. 1, 1995. Copyright 1995 by the Human Factors and Erogonomics Society. All rights reserved.

Who/What/When/Where

Sometimes, 'Where' is the only criteria for comprehending that there is a problem.

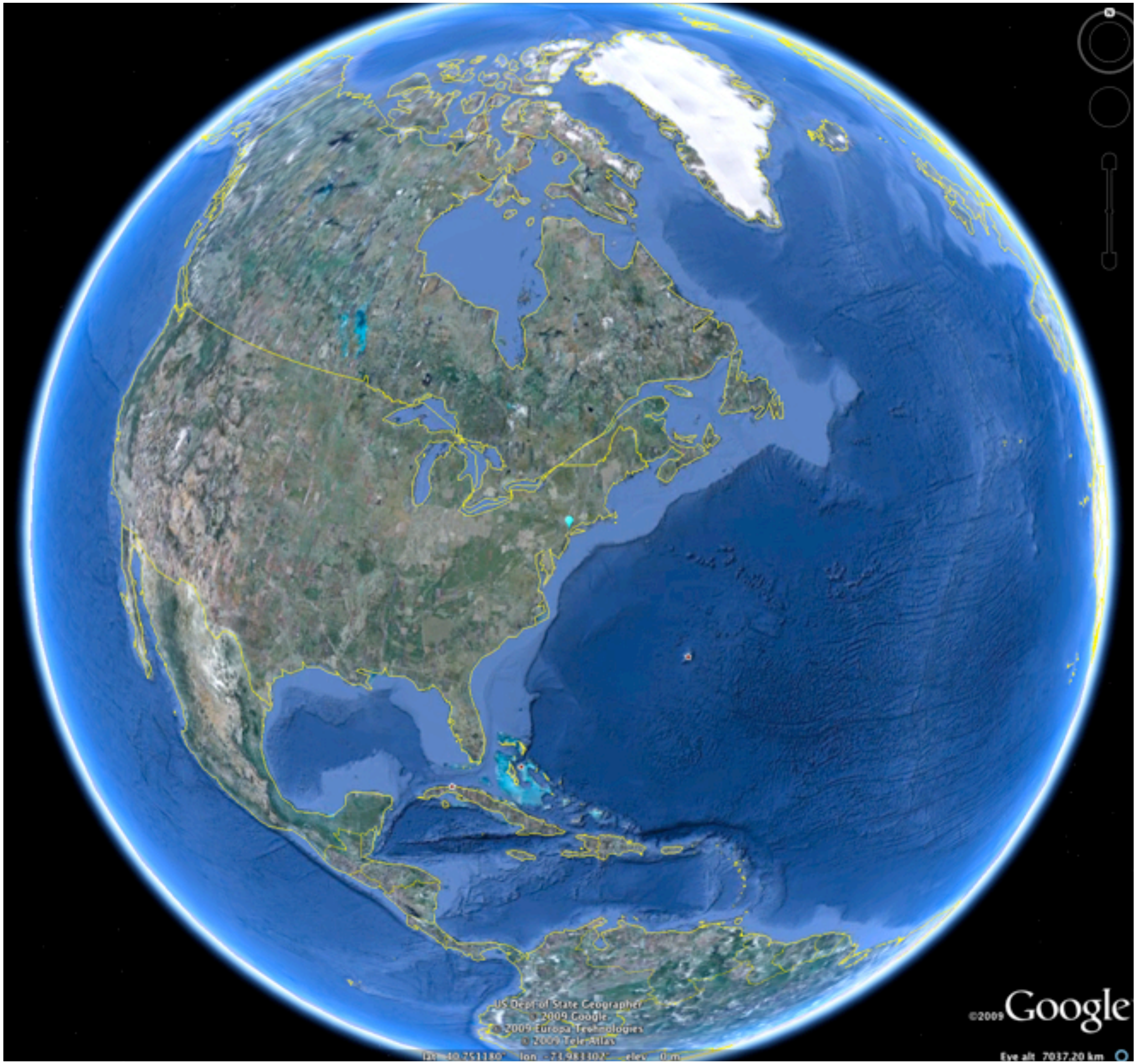
- Data isn't suppose to be coming from there.
- Data isn't suppose to be going that way.
- Data should to be coming from there but ....
- Where is this data coming from !!!!!!

Network flow data can be used in perception and comprehension of some of these very complex concepts, but the data needs to have some specific qualities in order to successively support 'where' functions.

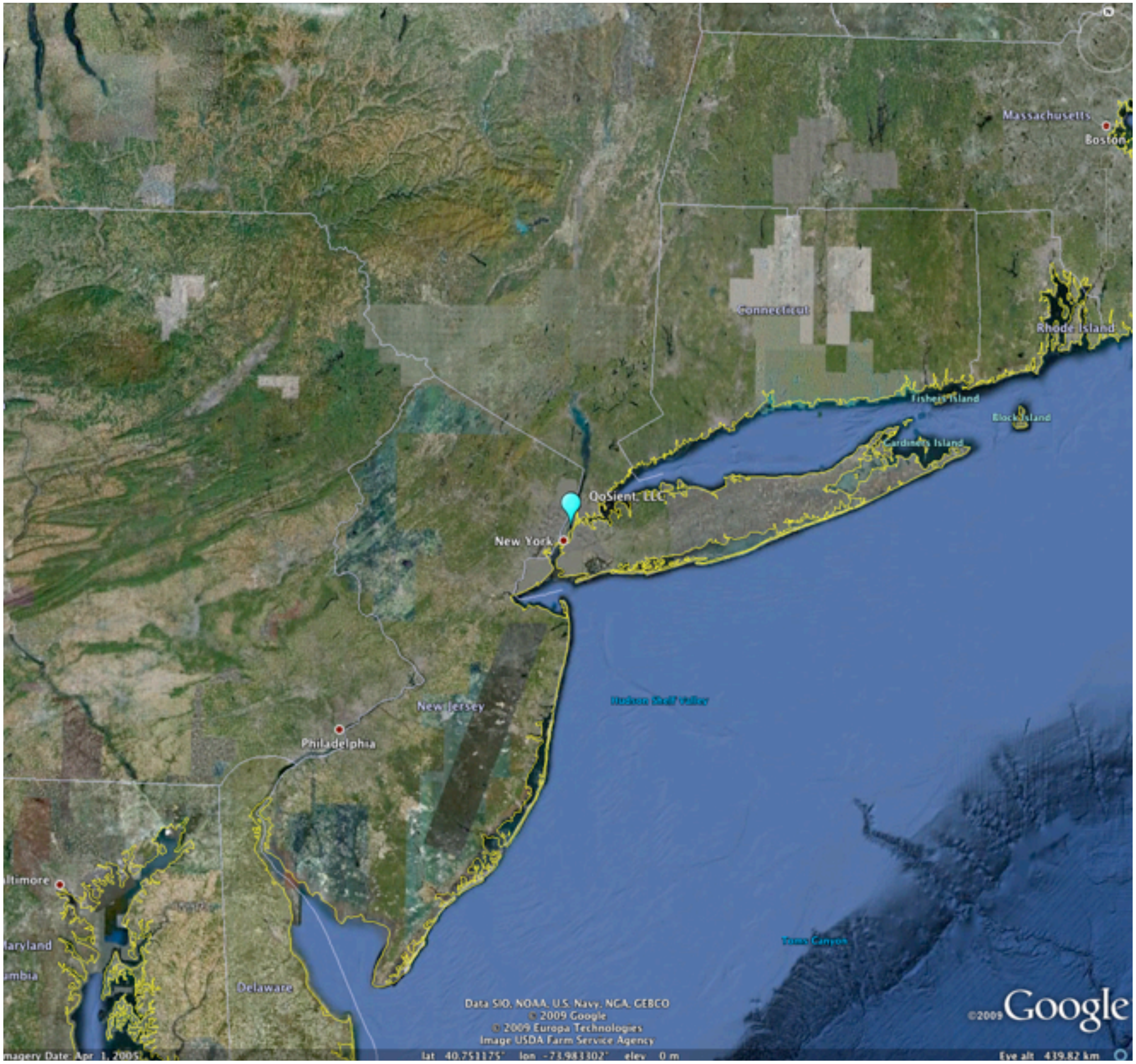
Who/What/When/Where

- GeoSpatial Information
  - Association with geographic information, GIS and Geomatics.
    - GeoLocation
      - Identification of 'real-world' geographic location information
      - Generally IP Address, MAC, RFID, Triangulation Based (rarely GPS based).
      - Time Zone, Country Codes, Region, City, Postal/Zip Codes, Lat/Lon
      - Commercial/Open Source Data Sources
        - Regional Internet Registries
        - ISP Provided Information
  - Used primarily for Marketing and Directed Advertisement.
  - Applications to E-commerce are emerging (taxation).
  - VoIP/SIP Based Emergency Services.
  - Lots of Standards (OGC, IEEE, W3C, ITU, IETF)
  - Little guidelines for privacy protection issues
  - Important Issue in Mobile Ad-hoc Networking
    - Gradient, aspect and visibility
    - Distance optimizations for power minimization and path length

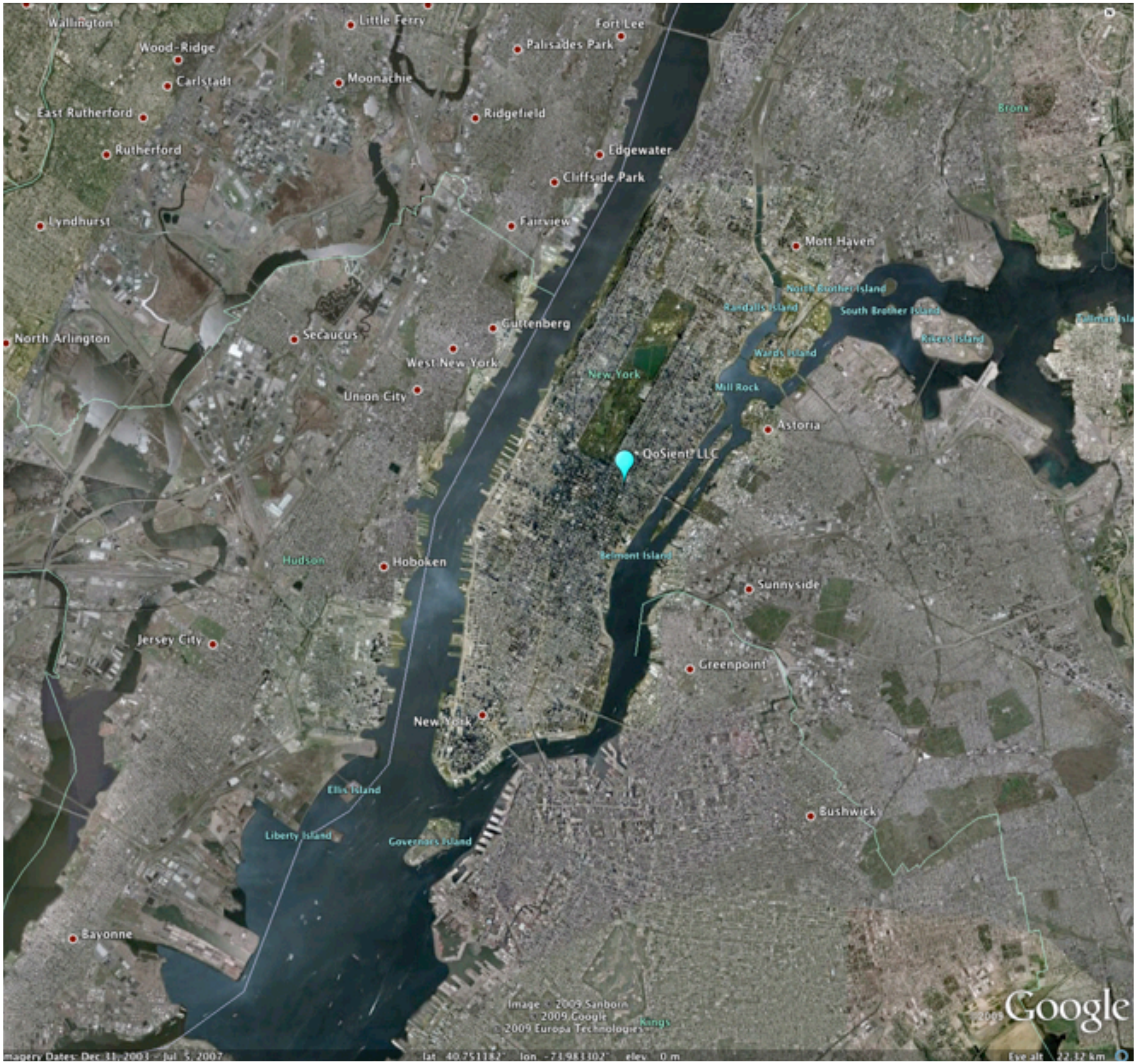
Where is QoSient.com?


US Dept of State Geographer
© 2009 Google
© 2009 Europa Technologies
© 2009 Tele Atlas
lat 40.751180° lon -73.983302° elev 0 m
©2009 Google
Eye alt 7037.20 km

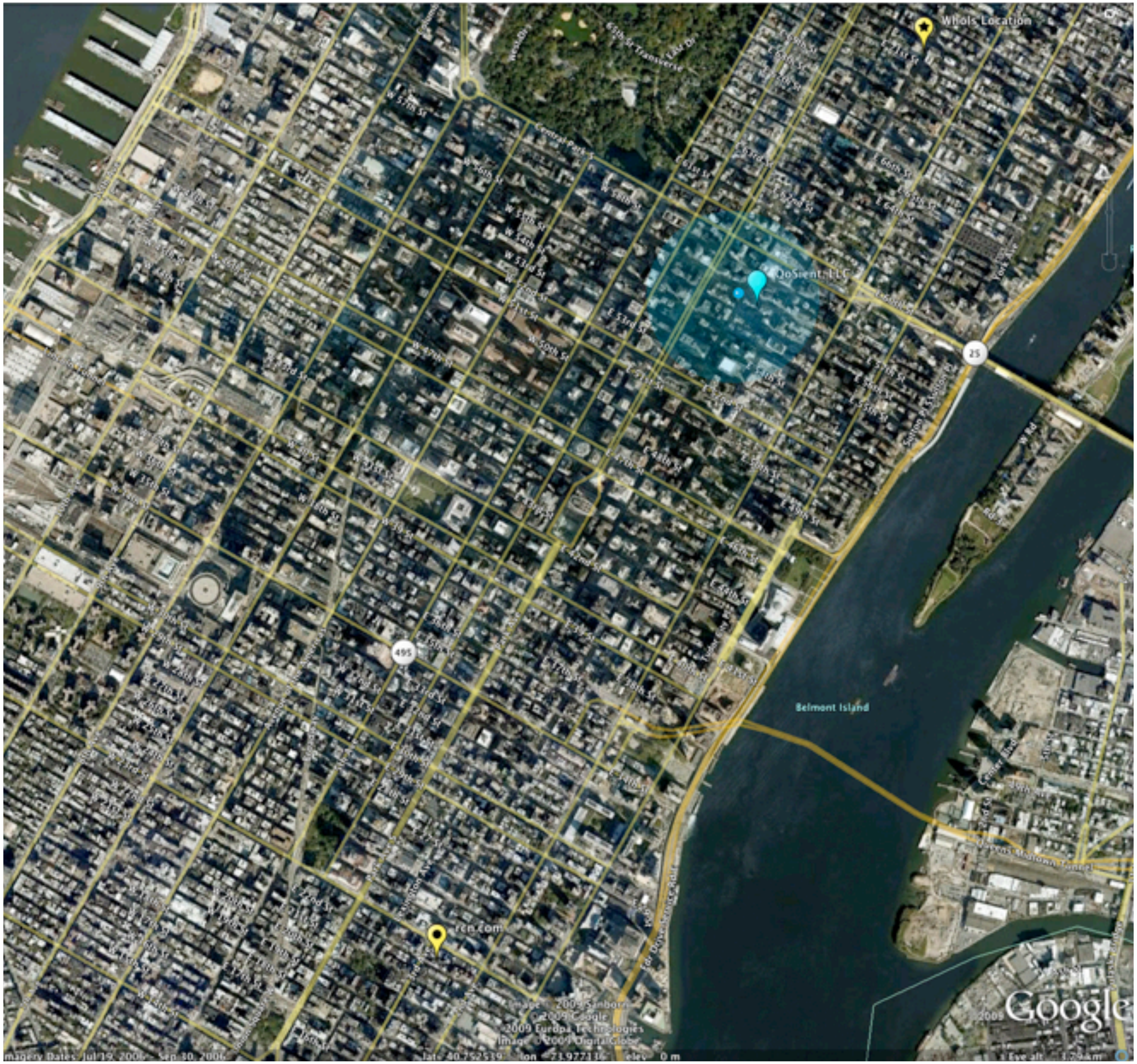
Where is QoSient.com?



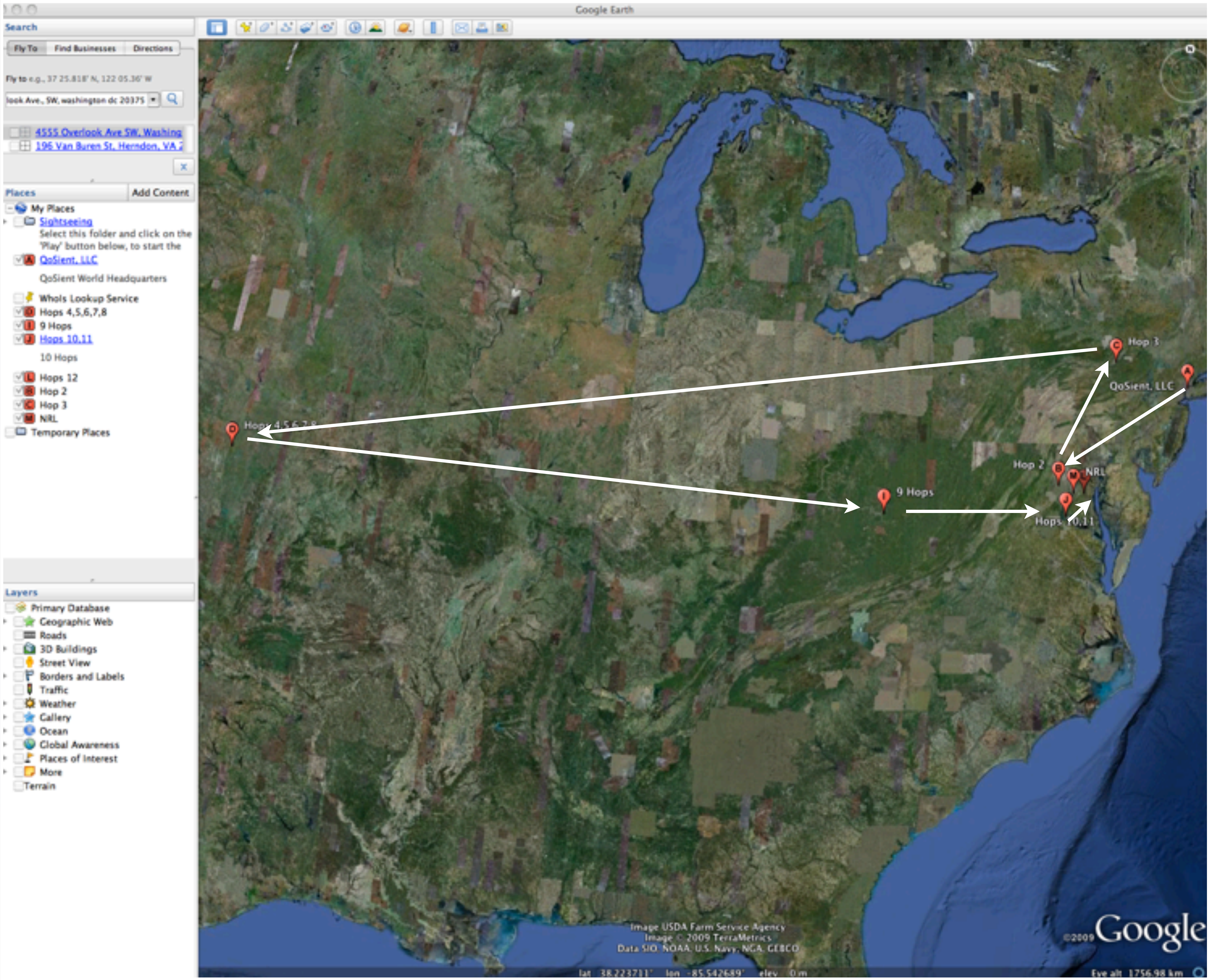
Where is QoSient.com?



Where is QoSient.com?


WhoIs Location
QoSient, LLC
Belmont Island
rcn.com
25
495
Image © 2009 Sanborn
© 2009 Google
©2009 Europa Technologies
Image © 2009 DigitalGlobe
Google
Imagery Dates: Jul 19, 2006 - Sep 30, 2006
lat 40.752539 lon -73.977116 elev 0 m
Eye alt 1.79 km

Network Path Information?



Who/What/When/Where

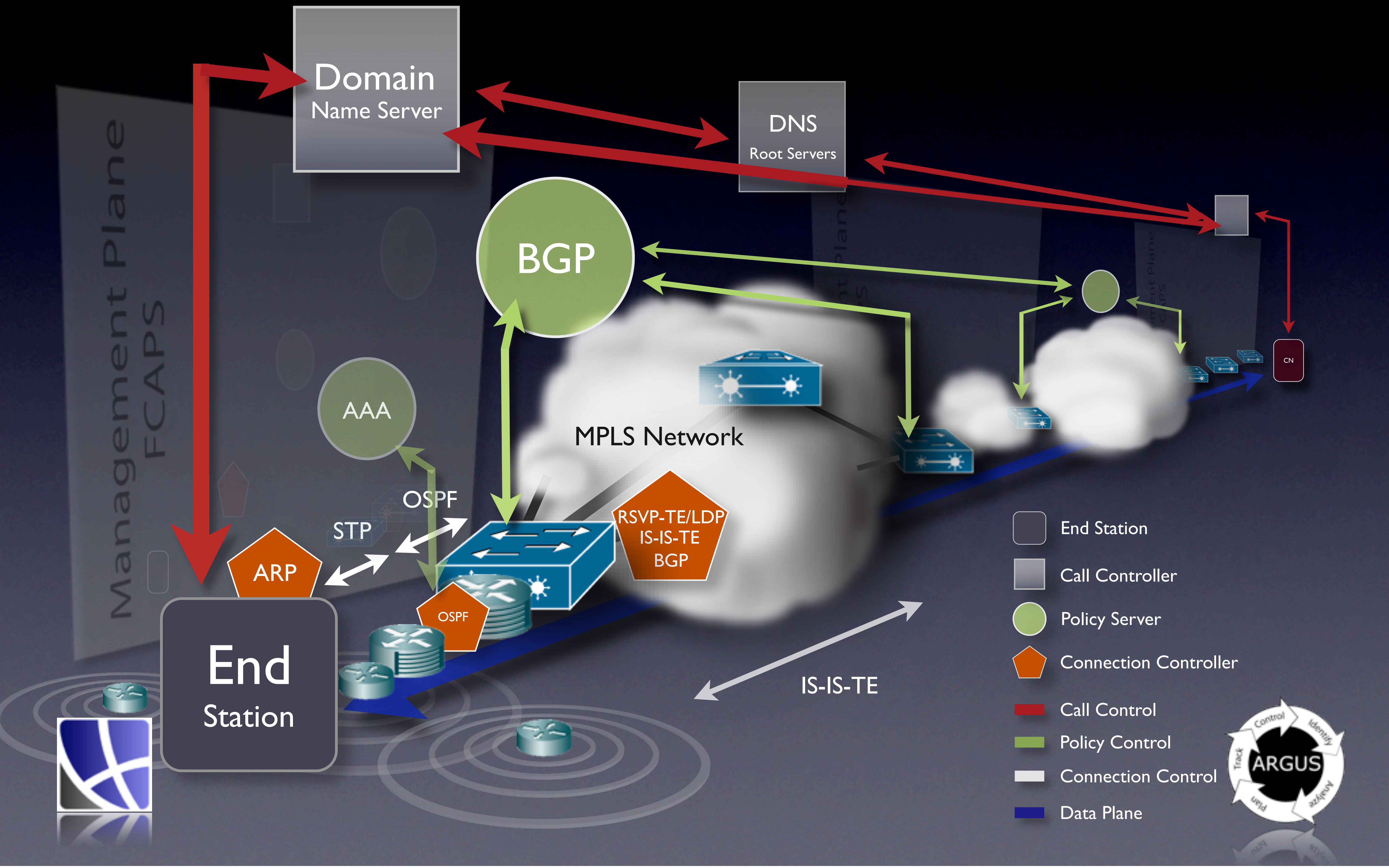
- Issues using geospatial information and flow data
  - There is no GeoSpatial Information in data packets
  - Most network flow data must be enhanced external to sensor
    - Flow data enhancement during/after data collection or distribution
  - No relational algebraic constraints on geospatial identifiers
    - IP addresses are not globally unique.
    - IP Address / Geolocation mappings are not formally managed/maintained.
  - Issues involve accuracy, relevancy, dynamism and time
  - IPFIX has not discussed geospatial/netspatial data support.

Who/What/When/Where

- Argus geospatial support
    - Flow Data Semantic Enhancement
        - radium() - collection based enhancement
        - ralabel() - post collection enhancement
    - metadata insertion strategy
        - saddr:lat=42.246532, lon=18.345261
        - geospatial information embedded in each record
    - direct GPS data insertion when available
- Support for printing, graphing, filtering, aggregation, and anonymization.
    - aniso lat/lon aggregation generates bounding box
    - lat/lon anonymization
        - constant offset projected onto either poles or ocean/land boundaries

China Syndrome

- But not all is as it appears to be.
  - QoSient.com constantly scanned by IP addresses from CN
    - Using ARIN databases for country codes.
    - Not a bother at all really.
  - One flow presented with estimated hop-count of 4 hops
    - Argus uses TTL to estimate hops (nearest 2^x - observed TTL).
  - Modified ping to source yielded RTT of less than 5 mSecs
    - Speed of light distance is 465.71 miles one-way.
    - Network distance estimates usually put 5 mSecs close, 0-20 miles.
    - So, what's up with this.
- Source address spoofing?
- Router root-kit attack?
- Routing infrastructure attack?



Management Plane
FCAPS

Domain
Name Server

DNS
Root Servers

BGP

AAA

MPLS Network

OSPF

STP

ARP

OSPF

RSVP-TE/LDP
IS-IS-TE
BGP

End
Station

IS-IS-TE

CN

□ End Station

▨ Call Controller

● Policy Server

⬟ Connection Controller

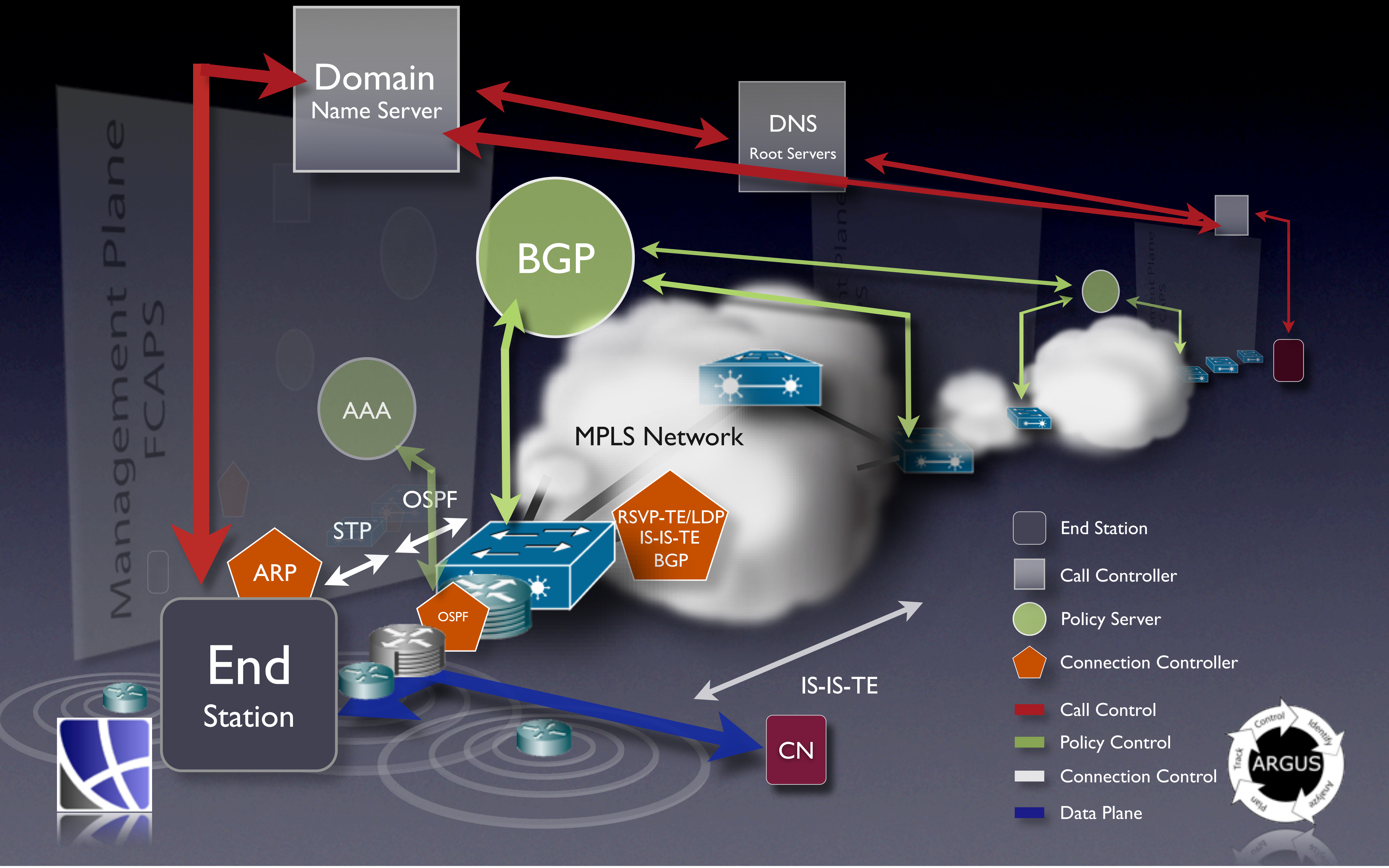
🟥 Call Control

🟩 Policy Control

⬜ Connection Control

🟦 Data Plane

ARGUS



- ▢ End Station
- ▨ Call Controller
- ● Policy Server
- ⬟ Connection Controller
- 🟥 Call Control
- 🟩 Policy Control
- ⬜ Connection Control
- 🟦 Data Plane

Who/What/When/Where

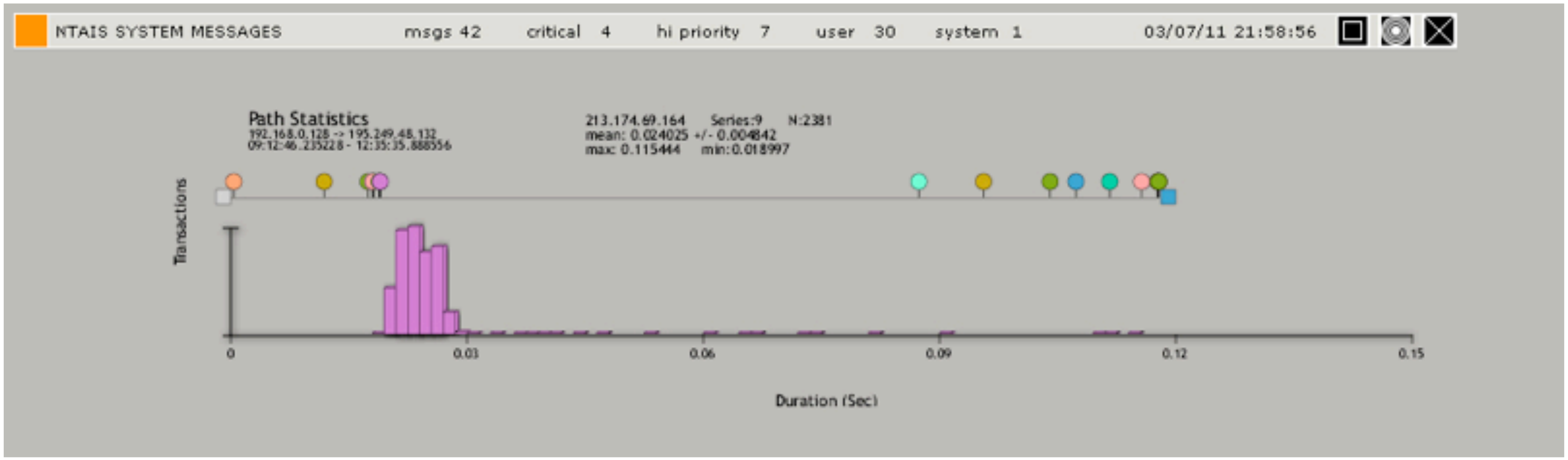
- How Do You Detect This?
  - Geospatial /Netspatial Incongruity
  - Network Distance Estimation and Correlation
    - Service Discovery, Service Usage Optimization, Group Join Optimizations, Shortest Path Routing
    - Methods
      - Global Network Positioning (GNP and NPS), CDN (Akamai), Internet Iso-bar, Internet Distance Maps (IDMaps), Vivaldi, Dynamic Distance Maps (DDM), RON, Landmark Clustering, Dynamic Landmark Triangles, Netvigator
    - All Network Distance Estimation Methods use simple active RTT metrics such as ping() and traceroute(), differentiations involve sampling strategies and statistical analysis.

Need For Active Elements?

- Timeliness of Determination/Validation
  - See a packet from some interesting IP address
    - Need to timely propagate the perception of address
    - Make GeoSpatial assessment
    - Decide to make some form of network estimation
    - Schedule ping/traceroute/probe....
- Flow sensors passively capturing network distance estimation metrics
  - Bi-directional flow monitors
    - Capture RTT regardless of protocol type
    - P1-P2 flow tracking captures traceroute information
  - Billions of Location Metrics Per Day
    - 8-10K Host Associations Per Work Group
    - 25% Infrequent, 35% Transient, 40% Persistent

Geo/NetSpatial Correlation

- Network Distance Estimation Accuracy
  - Large samples generate good results

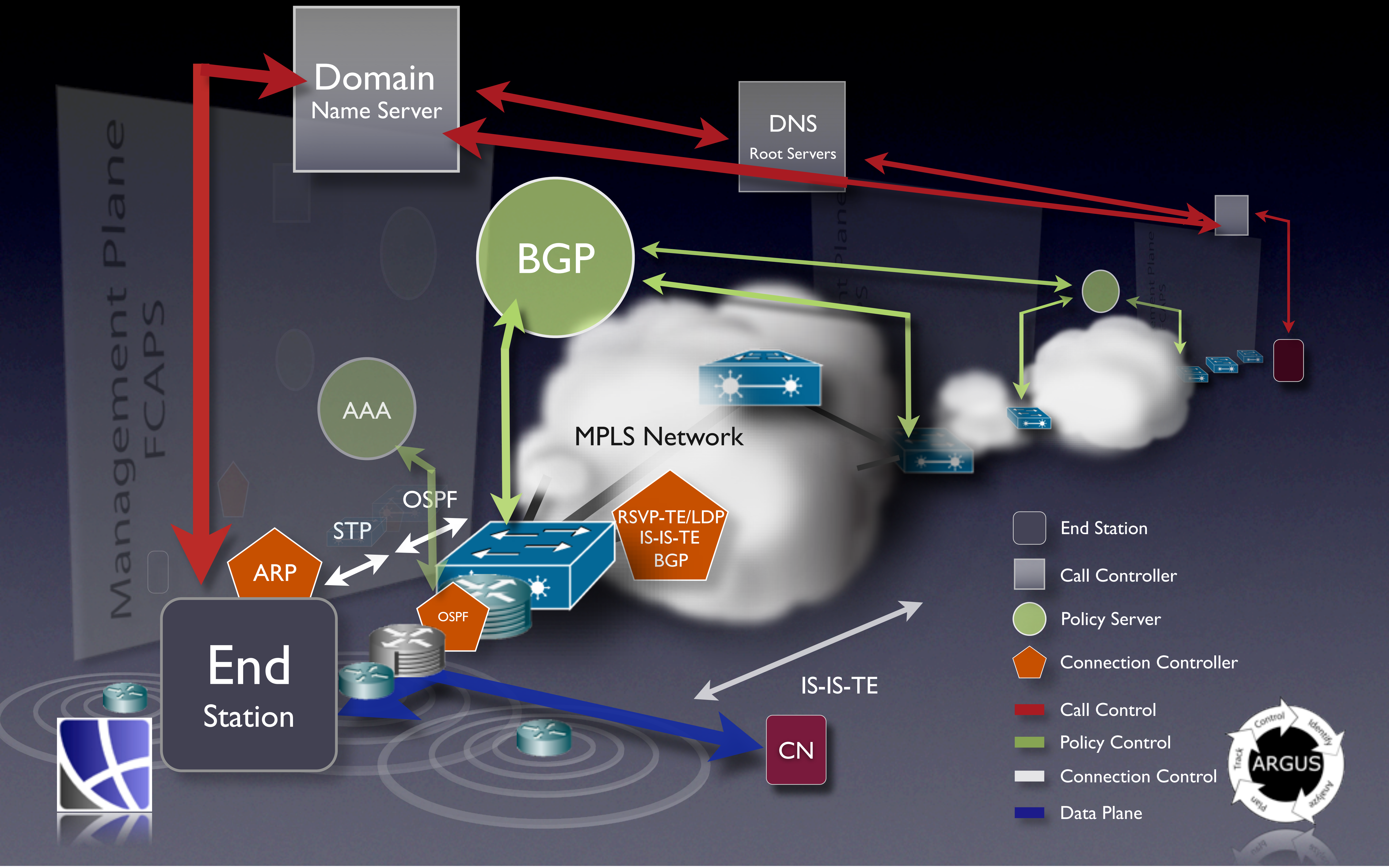


NTAIS SYSTEM MESSAGES msgs 42 critical 4 hi priority 7 user 30 system 1 03/07/11 21:58:56

Path Statistics
192.168.0.128 -> 195.249.48.132
09:12:46.235228 - 12:35:35.888556

213.174.69.164 Series:9 N:2381
mean: 0.024025 +/- 0.004842
max: 0.115444 min:0.018997

  - But, network distance does not relate to physical distance.
  - At least you can get a sense that something is up



Management Plane
FCAPS

Domain
Name Server

DNS
Root Servers

BGP

AAA

MPLS Network

OSPF

STP

ARP

OSPF

RSVP-TE/LDP
IS-IS-TE
BGP

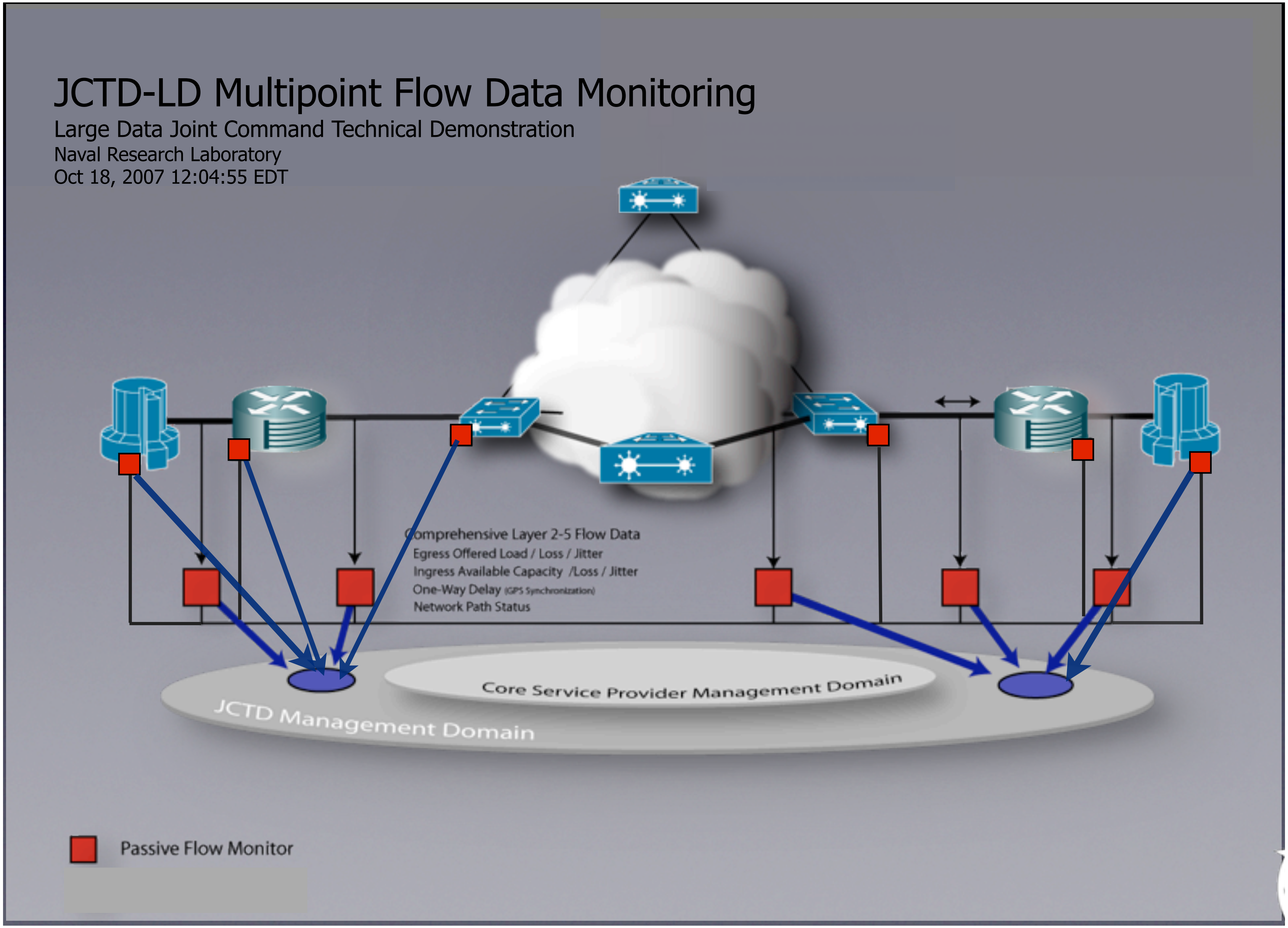
End
Station

IS-IS-TE

CN

- ▢ End Station
- ■ Call Controller
- ● Policy Server
- ⬟ Connection Controller
- 🟥 Call Control
- 🟩 Policy Control
- ▭ Connection Control
- 🟦 Data Plane

Multi-Point Monitoring



JCTD-LD Multipoint Flow Data Monitoring

Large Data Joint Command Technical Demonstration
Naval Research Laboratory
Oct 18, 2007 12:04:55 EDT

Comprehensive Layer 2-5 Flow Data
Egress Offered Load / Loss / Jitter
Ingress Available Capacity /Loss / Jitter
One-Way Delay (GPS Synchronization)
Network Path Status

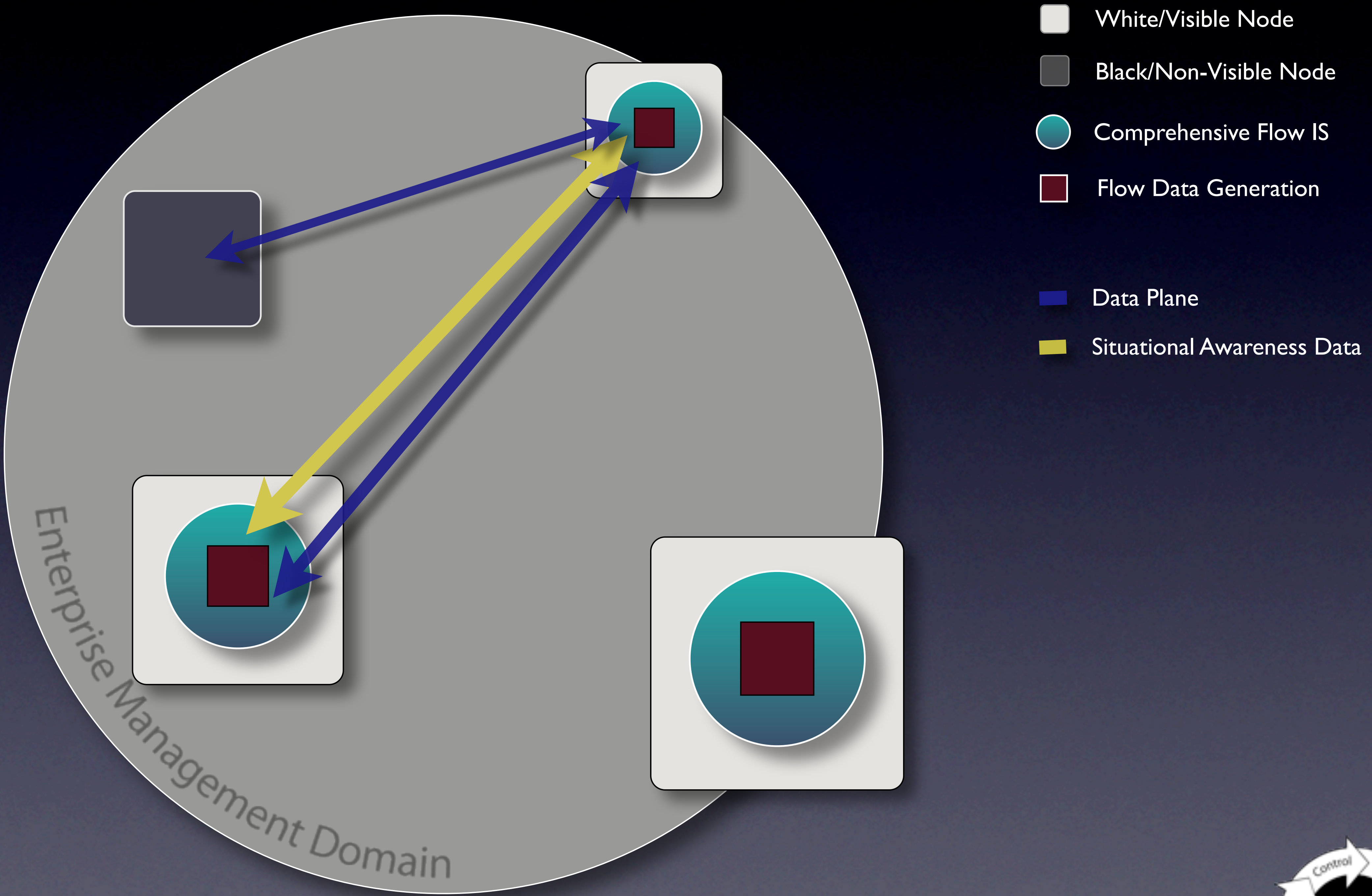
🟥 Passive Flow Monitor

MultiProbe Correlation

- Look for flows at multiple points
  - Differential analysis
    - One-way delay
    - Loss statistics
  - Path assurance
- Sensor placement provides utility
  - Exploit geospatial nature of observation domain
  - Validate explicit compartmentalizaiton
    - Exterior / Interior verification

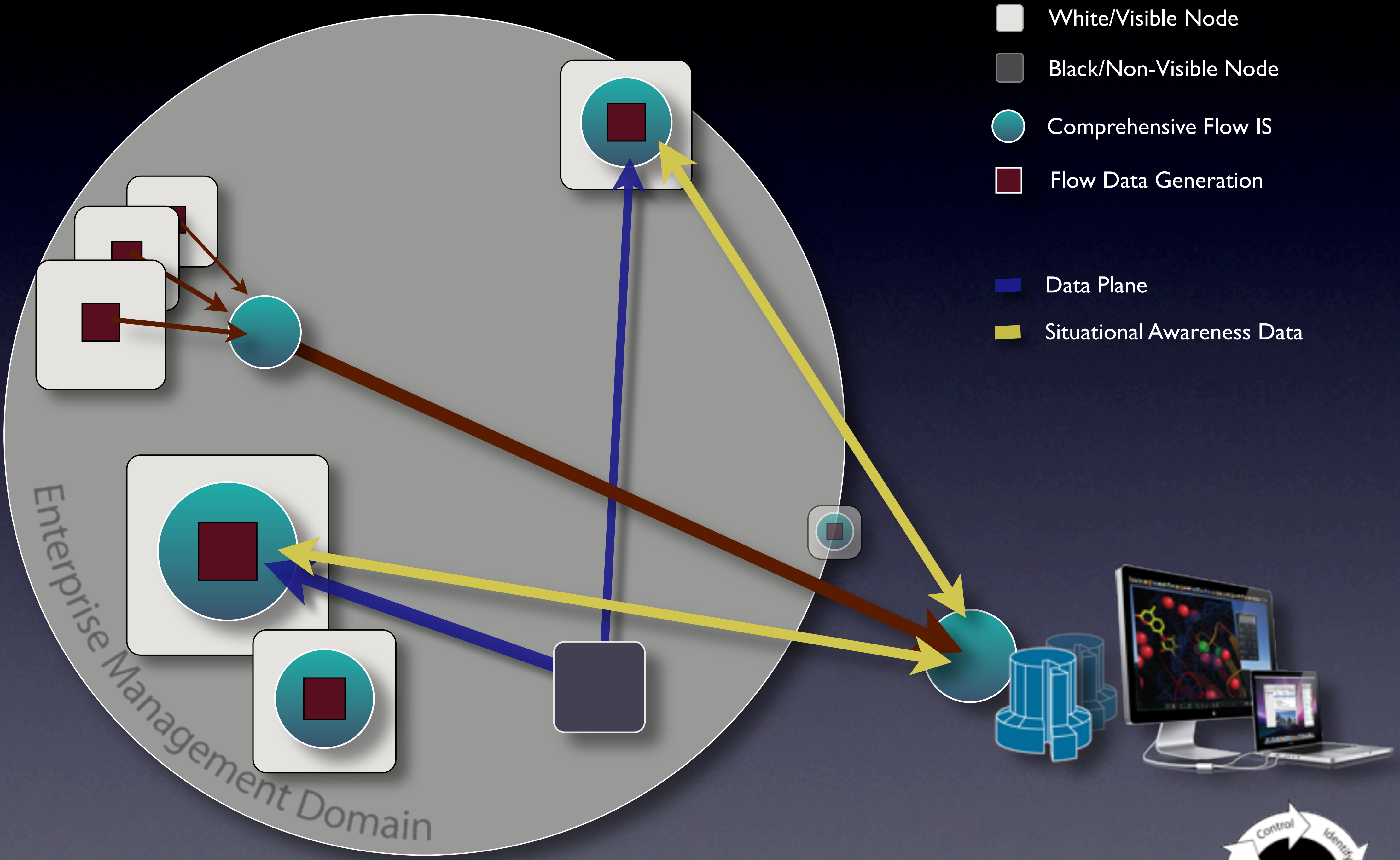
GeoSpatial Situational Awareness System

Mixed Black-box White-box Approach
Local Area Network Implementation



GeoSpatial Situational Awareness System

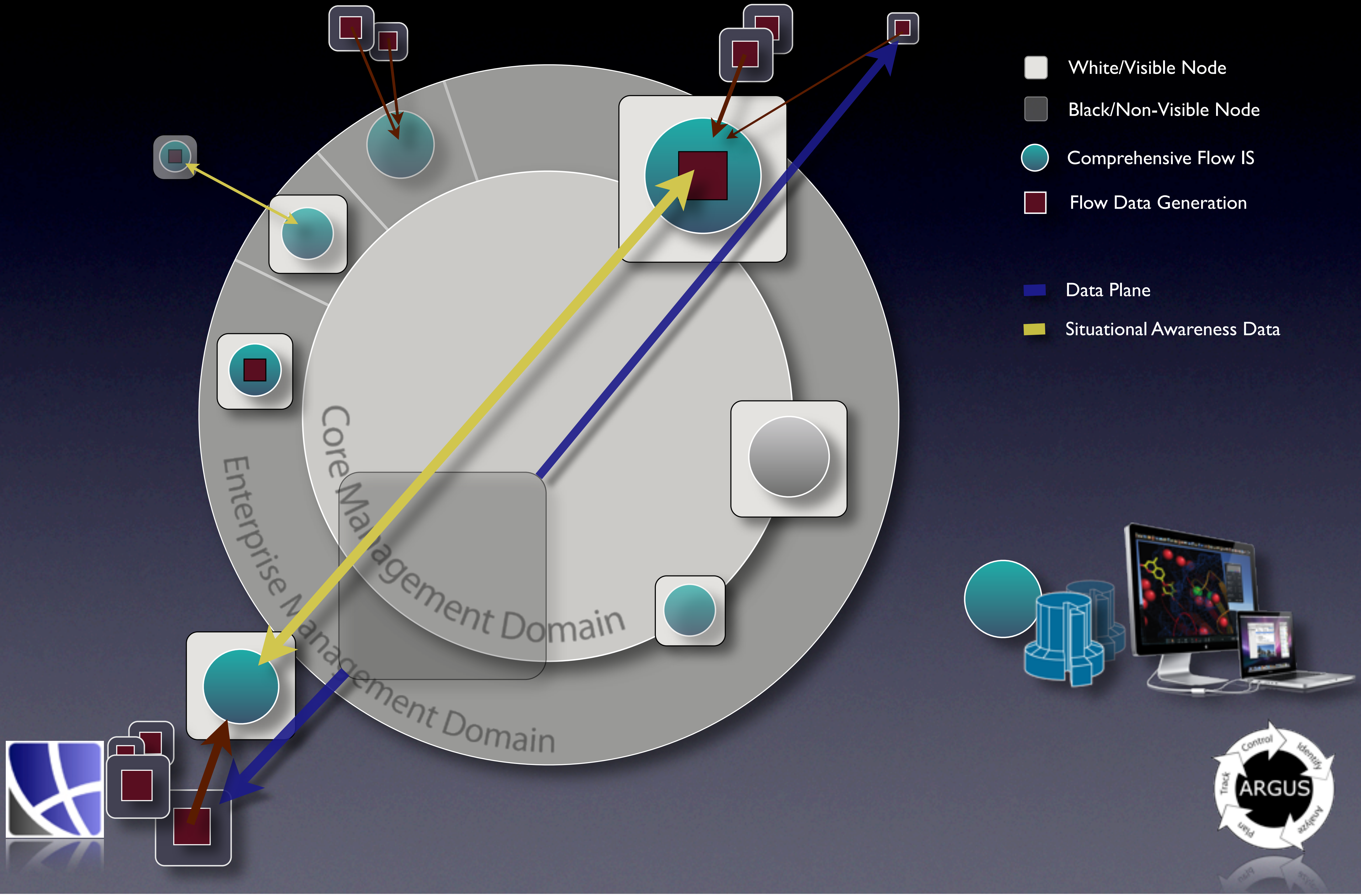
Mixed Black-box White-box Approach
Local Area Network Implementation



- □ White/Visible Node
- ■ Black/Non-Visible Node
- ● Comprehensive Flow IS
- ■ Flow Data Generation
- ▬ Data Plane
- ▬ Situational Awareness Data

GeoSpatial Situational Awareness System

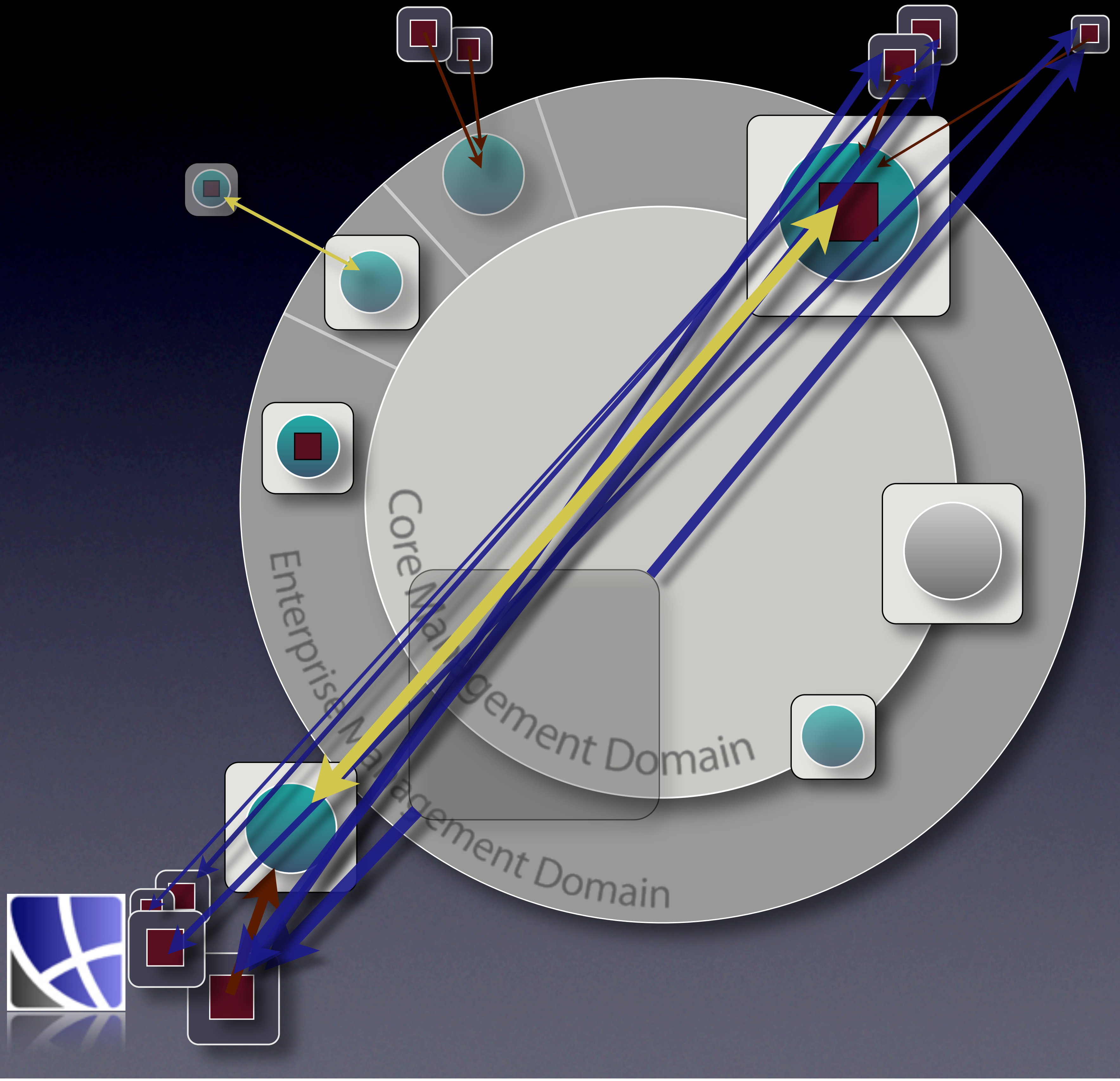
Mixed Black-box White-box Approach



- ⬜ White/Visible Node
- ⬛ Black/Non-Visible Node
- ● Comprehensive Flow IS
- 🟥 Flow Data Generation
- 🟦 Data Plane
- 🟨 Situational Awareness Data

GeoSpatial Situational Awareness System

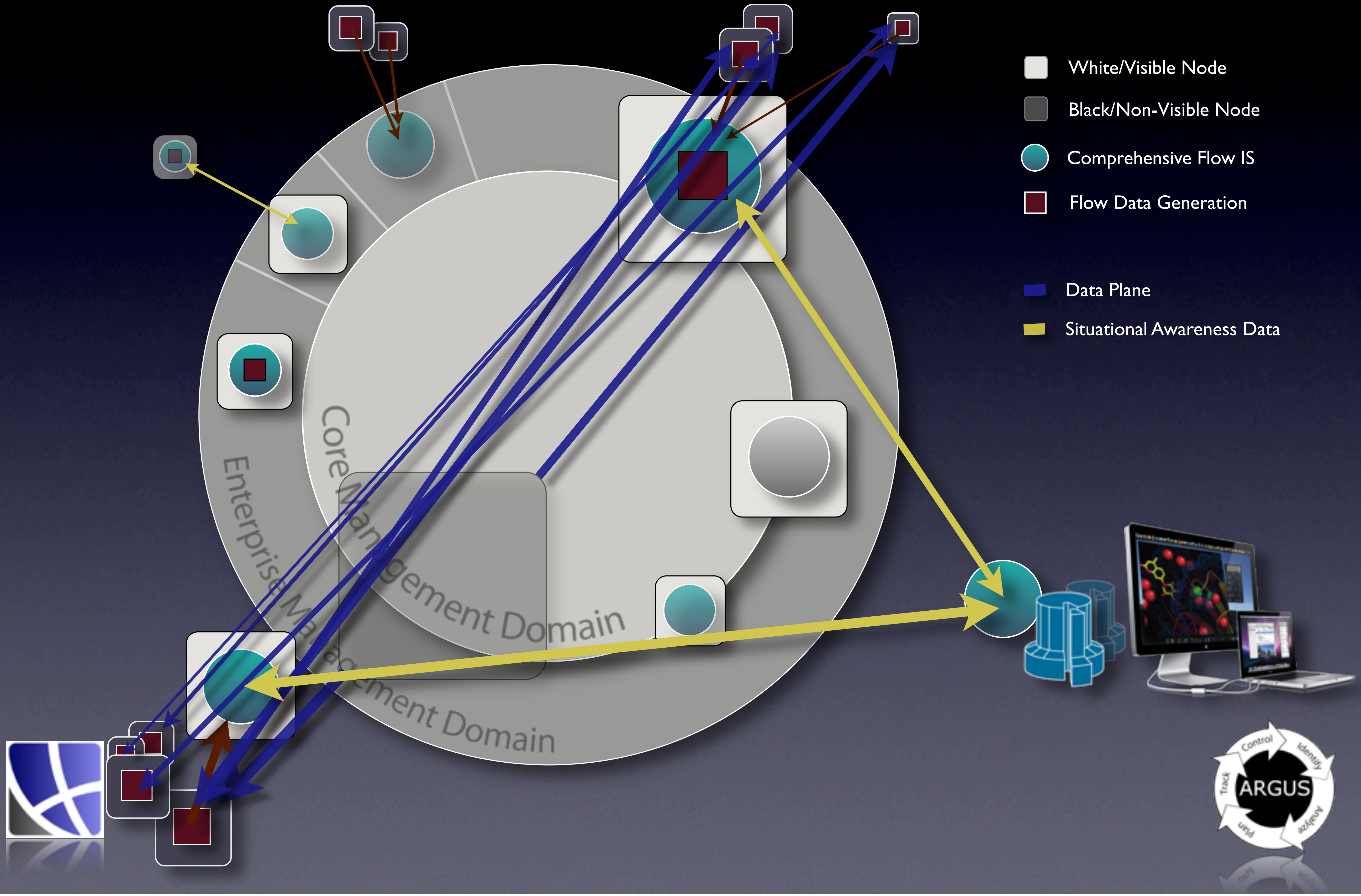
Mixed Black-box White-box Approach



- □ White/Visible Node
- ■ Black/Non-Visible Node
- ● Comprehensive Flow IS
- ■ Flow Data Generation
- ▬ Data Plane
- ▬ Situational Awareness Data

GeoSpatial Situational Awareness System

Mixed Black-box White-box Approach

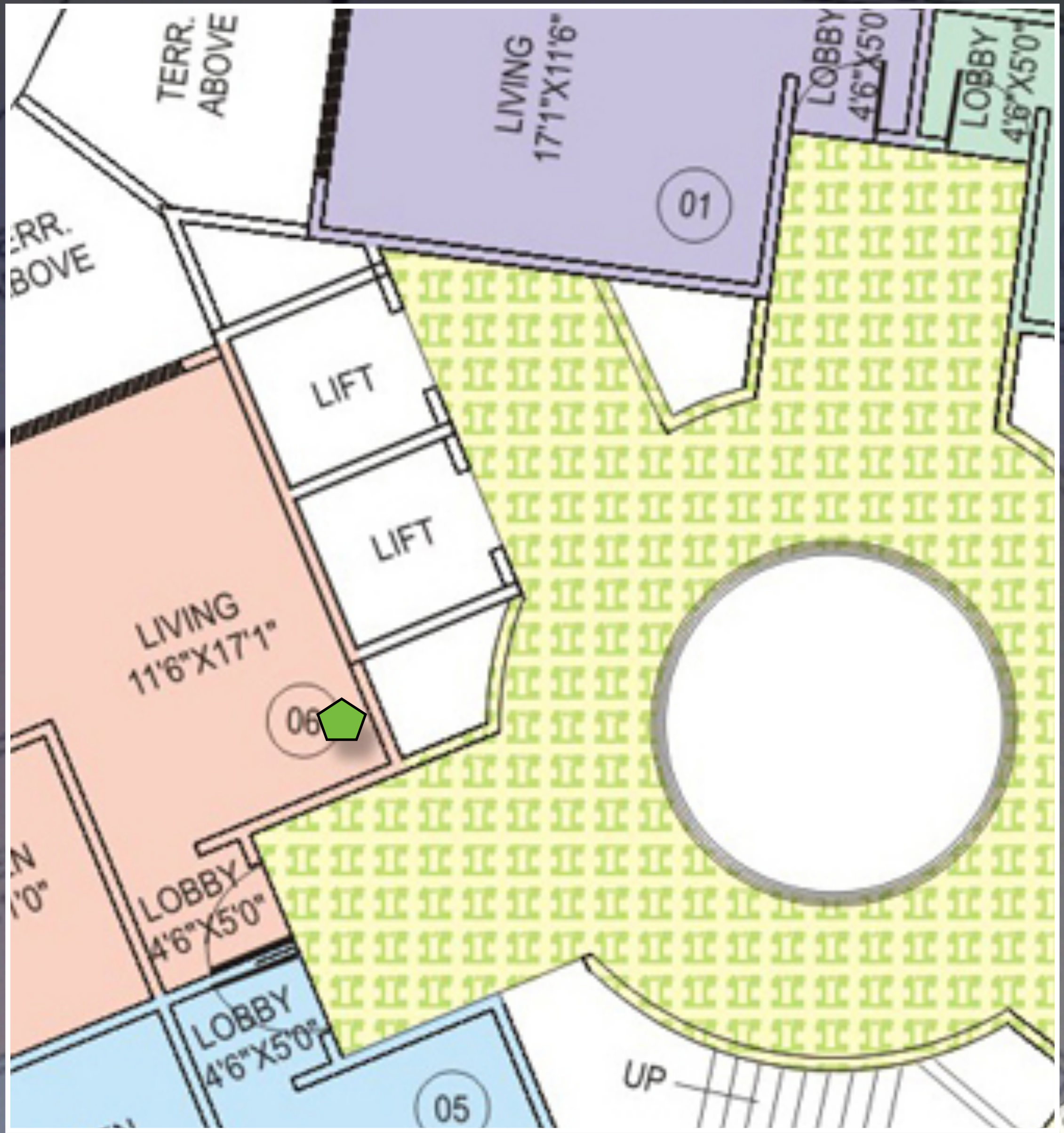


- ⬜ White/Visible Node
- ⬛ Black/Non-Visible Node
- ● Comprehensive Flow IS
- 🟥 Flow Data Generation
- 🟦 Data Plane
- 🟨 Situational Awareness Data

Threat Distance Assessment

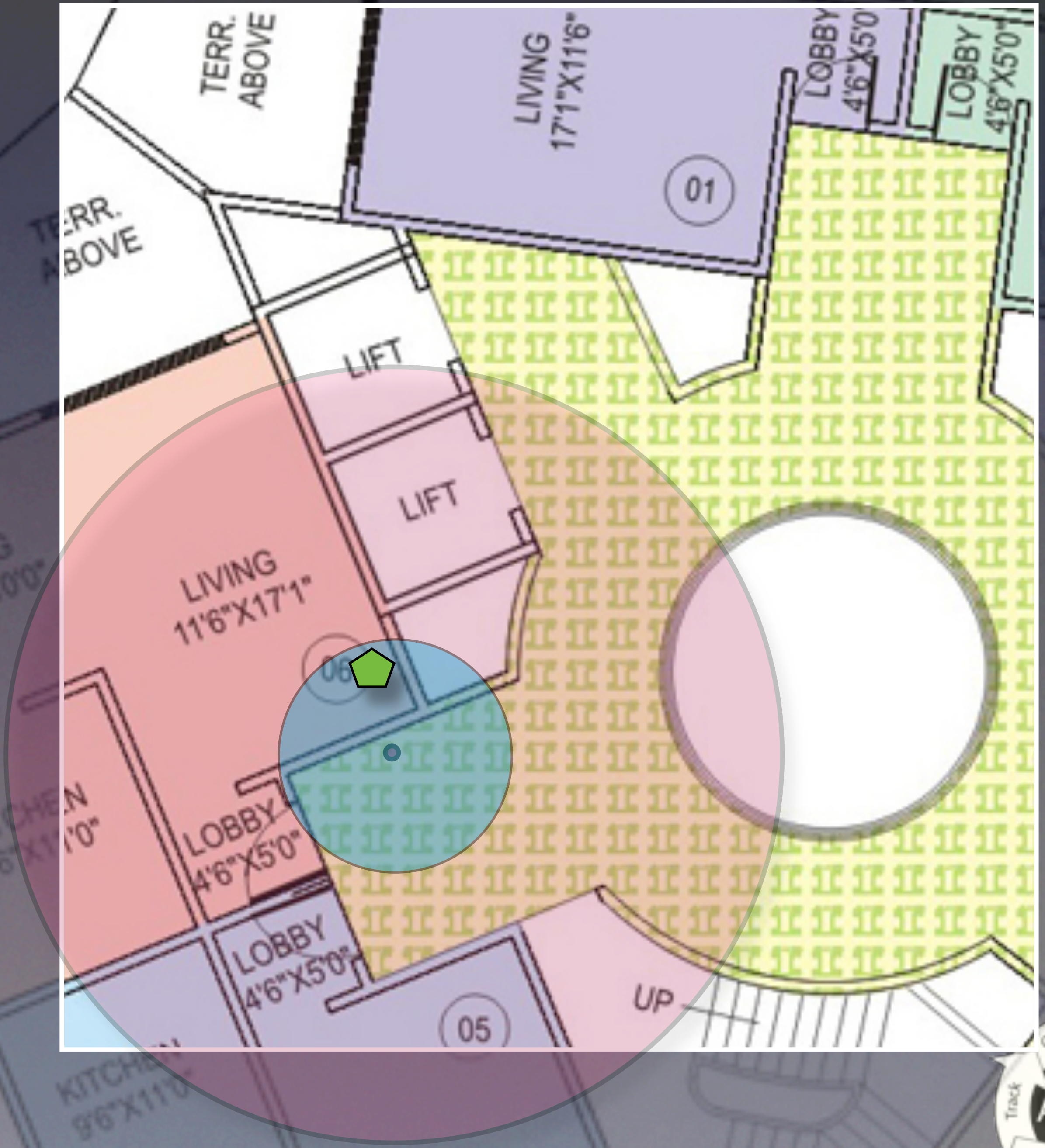
🟥 802.11 a/b/g/n

🟦 Bluetooth



Threat Distance Assessment

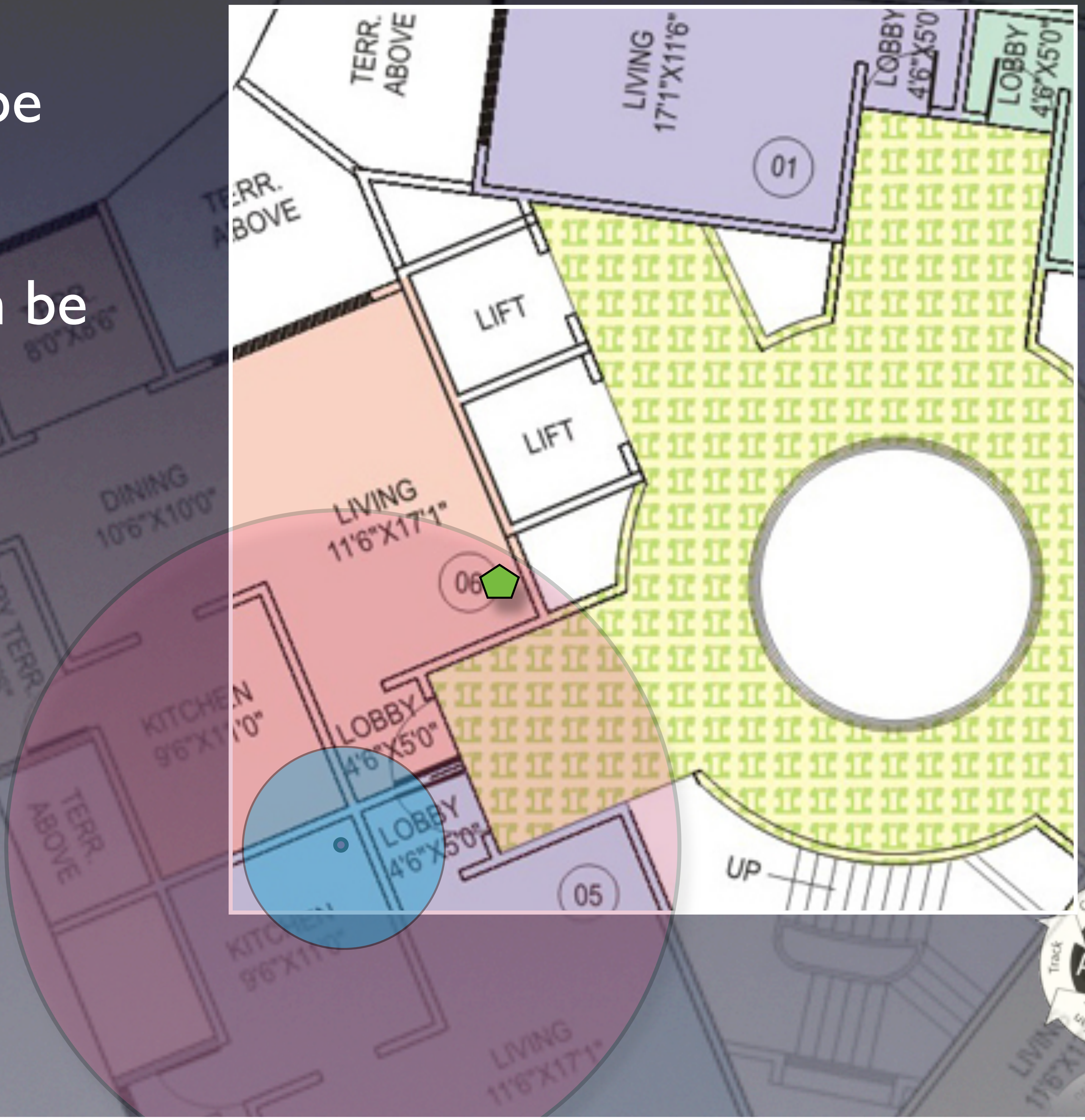
- Attack distances can be a matter of inches



🟥 802.11 a/b/g/n
🟦 Bluetooth

Threat Distance Assessment

- Attack distances can be a matter of inches
- Attack positioning can be rather surprising



🟥 802.11 a/b/g/n

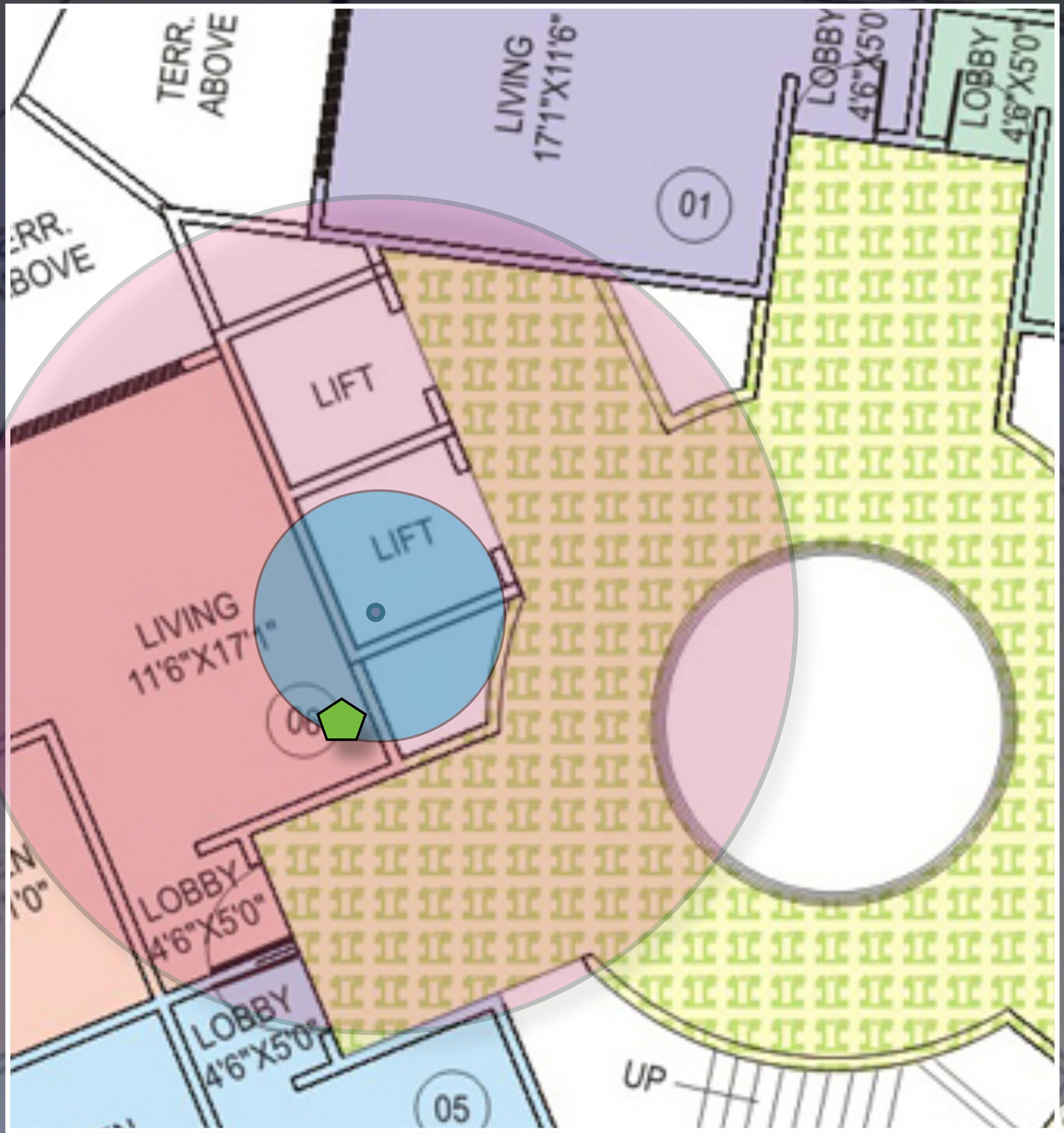
🟦 Bluetooth

Threat Distance Assessment

- Attack distances can be a matter of inches
- Attack positioning can be rather surprising

🟥 802.11 a/b/g/n

🟦 Bluetooth



Now Inches are Important

- How do we get the sensitivity?
  - Flow Data Generation in End Systems
    - Very dynamic situation
    - Commercial GPS is not accurate enough
      - Predominately an indoor problem
    - Unfortunately, the end system may not be cooperative
  - Flow Data Generation in Wireless Routers
    - Need to perform triangulation for 3D space
    - May require 4 signal strength values for complete discrimination.
  - Fuse signal strength information
    - Average/max/min receive signal strength vector over life of the flow.
    - Provide result of triangulation computations
    - Correlate with signal strength of 802.11 abgn beacon information

Conclusions

- Network flow data is a/the primary forensics and security data source for many large scale security systems.
- Fusion of non-packet derived information and flow data can be a very very nice thing.
- Some data requirements do exist
  - Time synchronization is critical
  - Semantic similarity
  - Relational algebraic constraints exist
- Flow data / flow data fusion is a big deal.
- Lots of work to be done in this area!!!