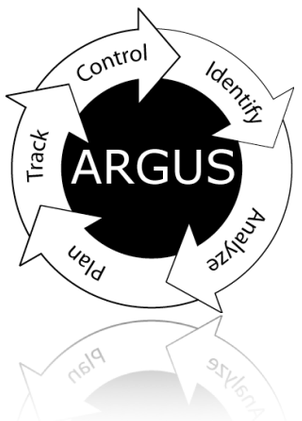
Network Flow Metadata

Very Large Scale Processing with Argus

Carter Bullard
QoSient, LLC

carter@qosient.com

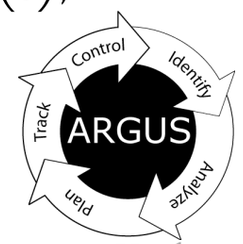
FloCon 2014
Charleston, South Carolina
Jan 13-16, 2014



Carter Bullard carter@qosient.com

- QoSient - Research and Development Company
  - US DoD, IC, DARPA, DISA
    - Very Large Scale Optimization (Operations, Performance, Security)
    - High Performance Network Security Research
    - DARPA CORONET Optical Security Architecture
  - Telecommunications / Performance Optimization
  - FBI / CALEA Data Wire-Tapping Working Group
- QoS / Security Network Management - Nortel / Bay
- QoS / Security Product Manager – FORE Systems
- CMU/SEI CERT
  - Network Intrusion Research and Analysis
  - Principal Network Security Incident Coordinator
- NFSnet Core Administrator (SURAnet)
- Standards Efforts
  - Editor of ATM Forum Security Signaling Standards, IETF Working Group(s), Internet2 Security WG, NANOG

QoSient



Tutorial Objectives

- Define Network Flow Metadata
- Discuss Issues in Large Scale Metadata Generation, Transport, Processing and Storage
- Describe Metadata Support in Argus
  - Strategies for Metadata Generation
  - Methods for Large Scale Metadata Processing
  - Very Large Scale Metadata Storage
- Conclusions

QoSient

Why Network Flow Metadata

"If you can't measure it, you can't improve it."

- W. Edwards Deming

- The purpose of network flow data is to contribute to the cost effective operations, performance and/or security of an IT infrastructure.

- Regardless of intent, this is an improvement process.

- But what is being improved; access control, assurance, dynamic response, resource utilization, transport efficiency, cost, power use, …, none are flow data metrics.

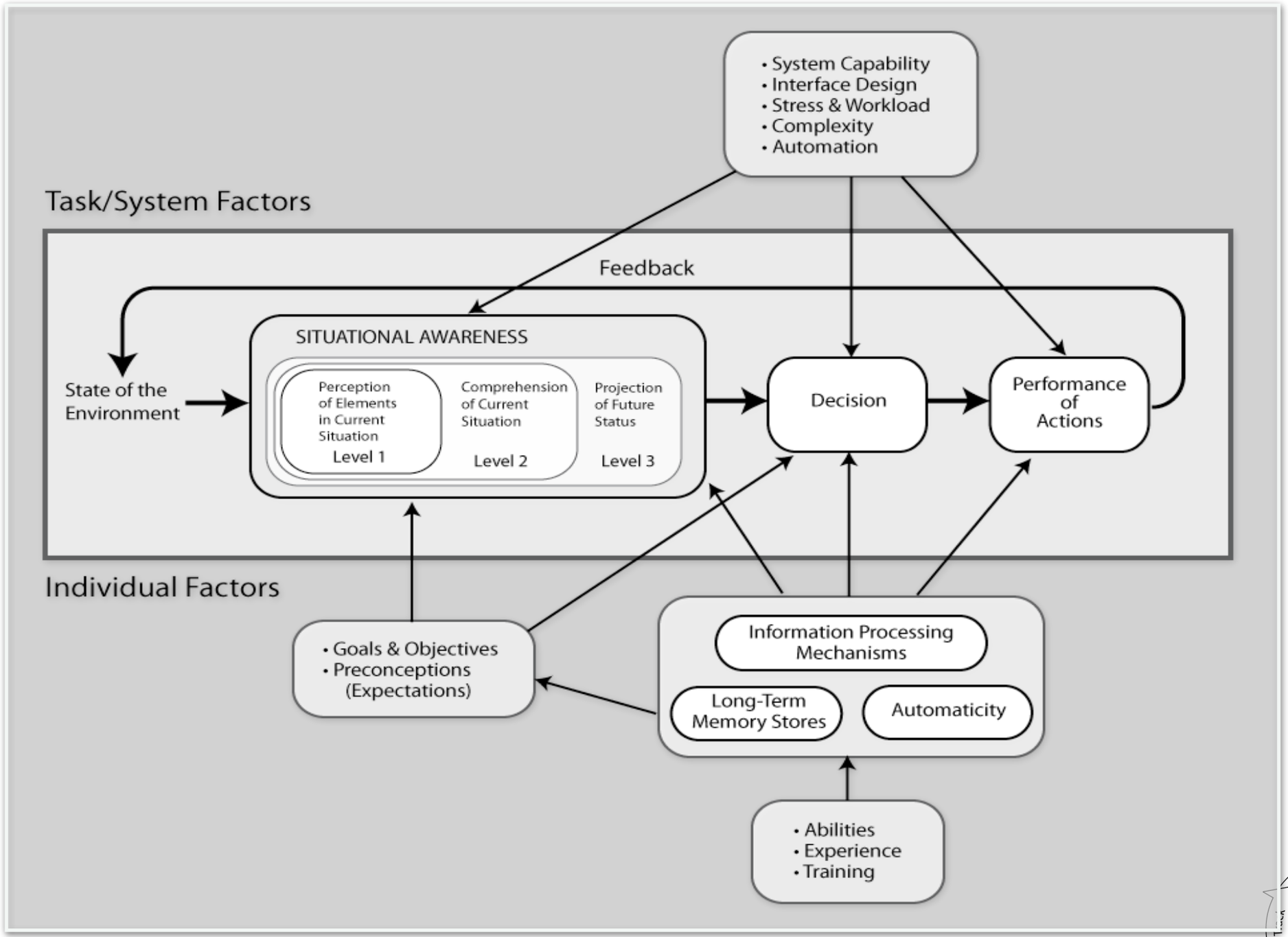
QoSient

Why Network Flow Metadata

- Network flow data provides
  - Presence data - objects and time
  - Load demand information
  - Observation domain identification
- To get to actionable information you need a bit more
- To get to access control policy verification and validation, you need a little more information

QoSient

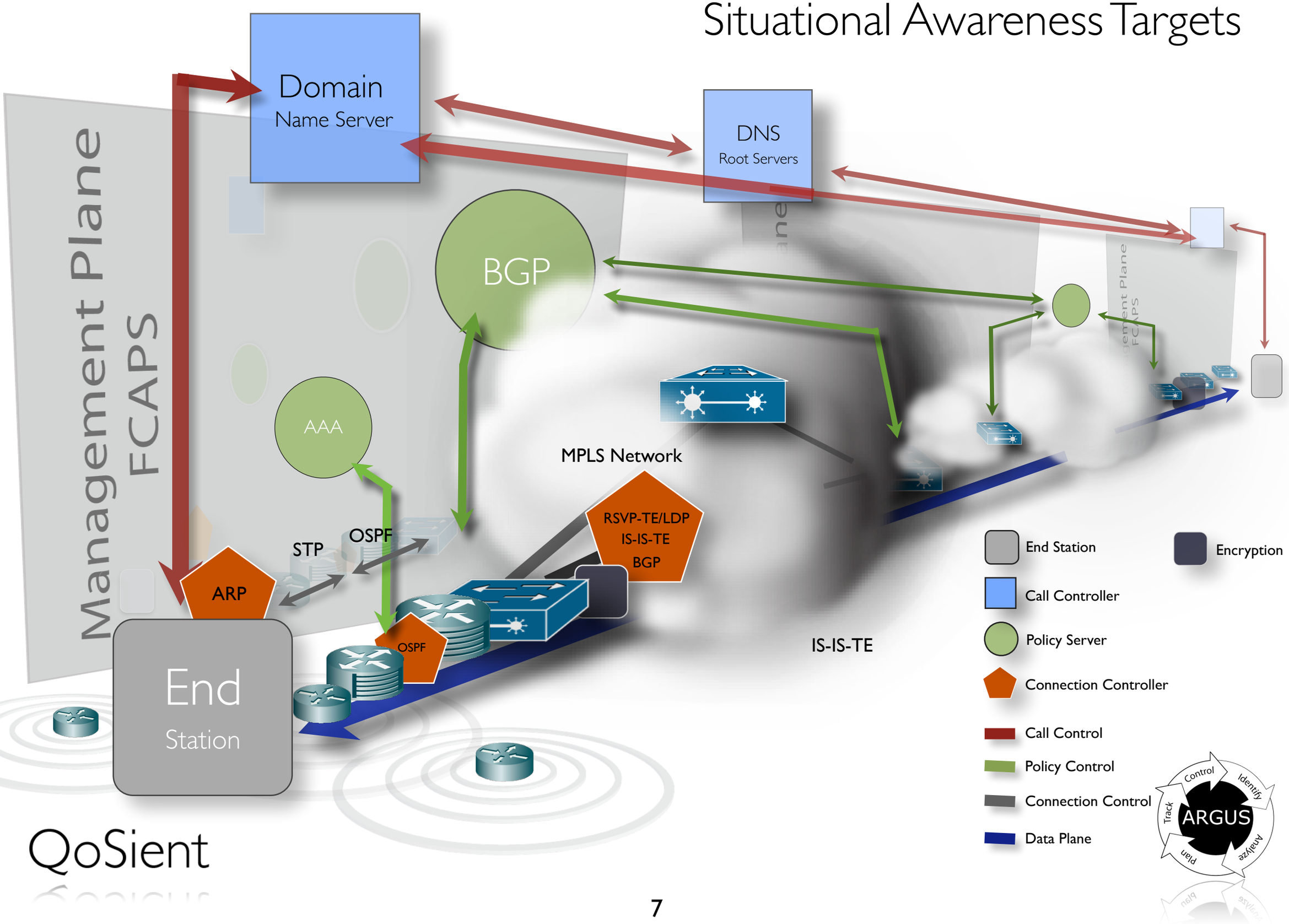
Model of Situational Awareness in Dynamic Decision Making



QoSient

Standard Internet Architecture

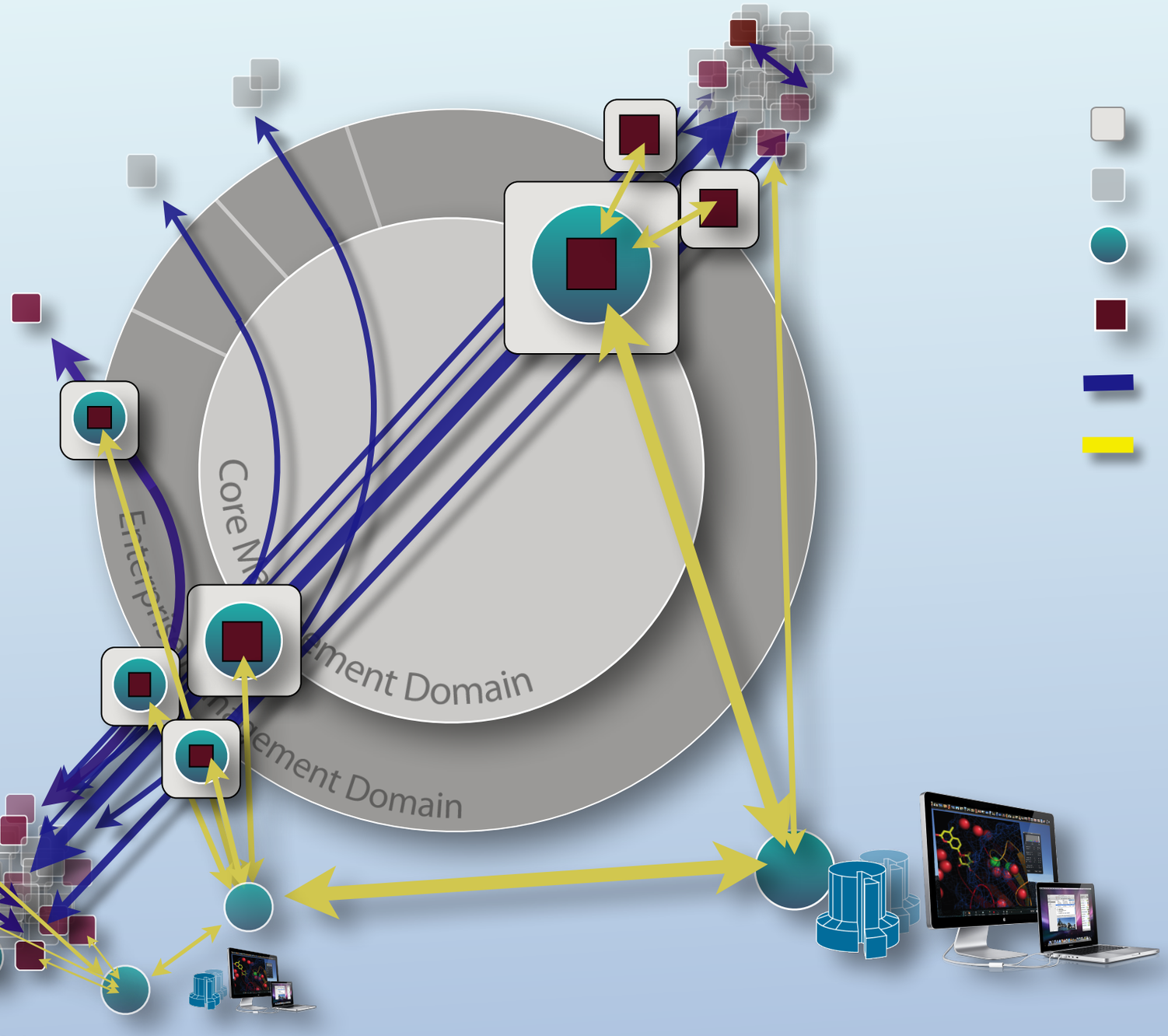
Situational Awareness Targets



QoSient

Complex Comprehensive Awareness

Local and Remote Strategies



White/Visible Node

Black/Non-Visible Node

Metadata Generation Node

Argus Sensor

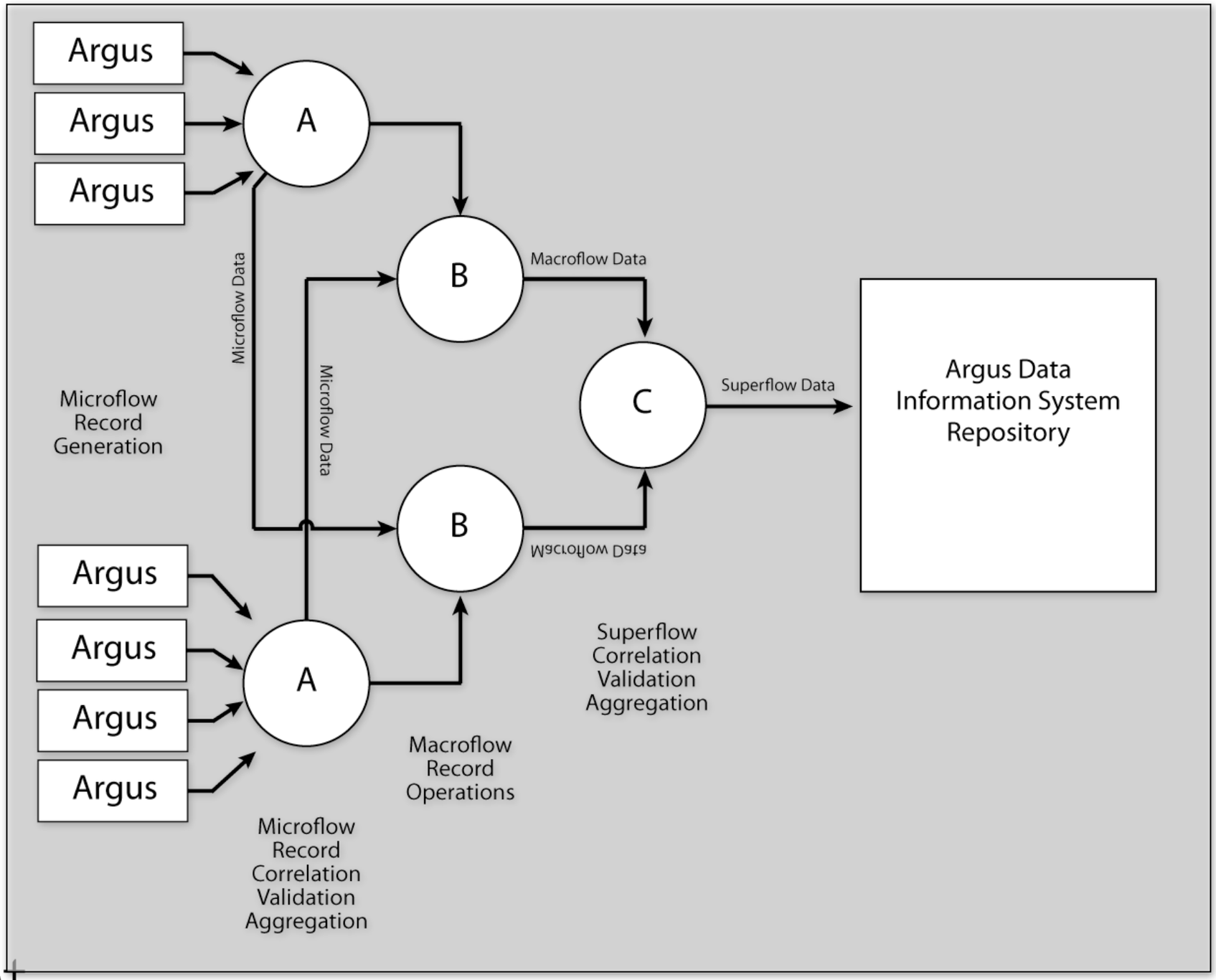
Data Plane

Situational Awareness Data

QoSient

Flow Data Processing Pipeline

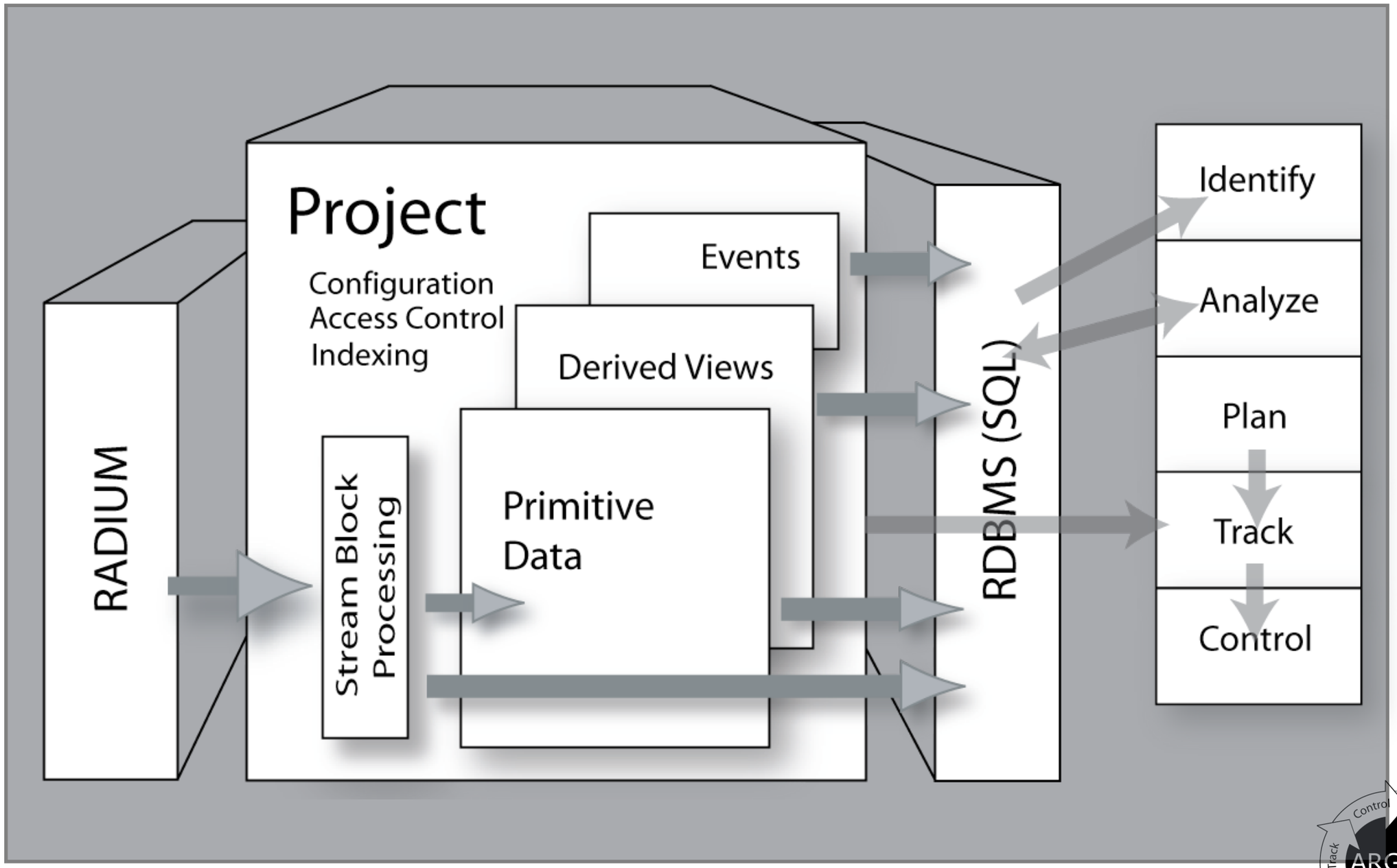
Data Flow Machine Architectures



QoSient

Argus Processing Design

Network Activity Information System (NAIS)



QoSient

Metadata and Network Flow

- Converting Packets into Awareness
  - Network Activity Classification
  - Semantic Enhancement
- Cyber Security Response
  - Identification / Attribution
  - Analysis / Forensics
  - Behavioral Anomaly Discipline
- Knowledge Discovery and Data Mining
  - Statistics, Databases, Pattern Recognition, Machine Learning Data Visualization, Optimization and High Performance Computing

QoSient

Metadata Definition(s)

QoSient

Metadata

- Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use, or manage an information resource.

"Understanding Metadata", National Information Standards Organization (NISO) http://www.niso.org, ISBN: 1-880124-62-9, 2004.

- Data about data / information about information
- Metadata addresses issues in data collections management, sharing and data usability.
- Generally, more metadata makes things easier to find, sort through, arrange, compare with similar items and evaluate. What are they about? How are they related to other things? Who may use them? How and when did we get them? The more we know about some thing, the more we can do with it or let others do.

QoSient

Functions of Metadata

- Discovery
  - Semantic Enhancement / Enrichment
  - Query / Search
  - Analytics
- Management
  - Identification
  - Interoperability
  - Control Intellectual Property Rights
  - Archiving and Preservation
  - Certify Authenticity
- Mark Content Structure
- Indicate Status
- Describe Processes

QoSient

Metadata Types

- Descriptive Metadata
  - The resource for discovery and identification
  - Represents the semantics of data
  - Provides structure for semantic enhancement / enrichment
- Administrative Metadata
  - Information to help manage a digital resource
  - Creation, access controls, rights management, preservation
  - Data provenance
- Structural Metadata
  - Describes the physical and/or logical structure of data
  - Commonly used to facilitate navigation and presentation
  - Addresses issues regarding formats (e.g. ISO 3166 Country Codes, IPFIX)
- Meta-Metadata

QoSient

Metadata Levels

- Discovery Metadata
  - The minimum amount of information that needs to be provided to convey to the inquirer the nature and content of the data resource.
  - Answers the "what, why, when, who, where and how" of data.
- Exploration Metadata
  - Provides sufficient information to enable an inquirer to ascertain that data, fit for a given purpose, exists.
  - Thus, after discovery, more detail is needed about individual data sets, and more comprehensive and more specific metadata is required.
- Exploitation Metadata
  - Properties required to access, transfer, load, interpret and apply the data where it is exploited.

Network Flow Metadata

- Metadata is a relativistic concept
  - Network flow data is, by definition, packet metadata
  - Flow data is Intrusion Exploit reporting metadata
- Information not derived from packet contents
  - Primary - Hostnames, DHCP lease assignments, AS Numbers
  - Secondary - GeoSpatial, NetSpatial, User, Application Data
  - Tertiary - Regional Weather Information, IDS classifications
  - Flow status / metrics, packet dynamics are generally not metadata
  - Sessionization Summaries / Aggregations
  - Inter-Flow Dynamics
  - Behavioral Classifications
    - This flow is an internal attack using external addresses
    - This flow contains 8 keystrokes

QoSient

Network Flow Metadata

- Descriptive Metadata
  - Flow Data Objects - discovery and identification
  - Flow Metrics - semantics
  - Argus labels, events, behaviors
- Administrative Metadata
  - Observation domains - managing a digital resource
  - Argus MAR - Creation, access controls, rights management,
  - Argus labels - Data provenance
- Structural Metadata
  - Flow Formats - the physical and/or logical structure of data
  - ??? - Commonly used to facilitate navigation and presentation
  - Data formats (e.g. ISO 3166 Country Codes, String)
- Meta-Metadata

Argus Descriptive Metadata

- Complex Argus flow models
  - P1 / P2 Bi-Directional Multicast Unicast Tracking
  - Availability / Reachability / Connectivity Status indications
- Argus Metrics and Analytics
  - Interpacket arrival and jitter metrics
  - Data Aggregation metrics - COI membership and demands
  - Frequency domain classifications for status / health
- Argus Labels
  - GeoSpatial labeling - Lat/Lon, Zip Code, Country Codes
  - Address based labeling - COI, organizational labeling
  - Flow based labeling - Free form label assignments
- Argus Events and Data Correlation
  - Cross domain flow semantic enhancement
- Argus Behavioral Classifications
  - Producer / Consumer
  - Keystroke detection

QoSient

Metadata Standards

QoSient

Metadata Standards

- Text Encoding Initiative (TEI)
- Metadata Encoding Transmission Standard (METS)
- Metadata Object Description Schema (MODS)
- MPEG-21: Digital Item Declaration Language (DIDL)
- Data Documentation Initiative (DDI)
- Digital Object Identifier (DOI)
- Dublin Core
- Common Warehouse Meta-model (CWM)
- Learning Objects Metadata (IEEE LOM)
- GeoSpatial Metadata
    - ISO 19115:2003 (Geographic Information - Metadata)
    - GILS (Global Information Locator Service)

Data Document Initiative V3.1 (DDI)

Metadata for the Combined Life Cycle Model

QoSient

Descriptive Metadata

- Knowledge Discovery
  - The resource for discovery and identification
- Identification
  - Source, creation dates, times, URL, DOI
- Interoperability
  - Cross system search - Z39.50
  - Metadata harvesting - Open Archives Initiative
- Annotation
  - Hyperlinked relationships between resources
  - Annotations by users
  - Metadata for record keeping systems

QoSient

Descriptive Metadata

- Example ONIX - ONline Information eXchange

```
<Title>
    <TitleType>01</TitleType>
    <TitleText textcase = "02">British English, A to Zed</TitleText>
</Title>
<Contributor>
    <SequenceNumber>1</SequenceNumber>
    <ContributorRole>A01</ContributorRole>
    <PersonNameInverted>Schur, Norman W</PersonNameInverted>
    <BiographicalNote>
        A Harvard graduate in Latin and Italian literature, Norman Schur attended
        the University of Rome and the Sorbonne before returning to the United
        States to study law at Harvard and Columbia Law Schools. Now retired
        from legal practice, Mr Schur is a fluent speaker and writer of both British
        and American English
    </BiographicalNote>
</Contributor>
```

QoSient

Descriptive Metadata

- Example ONIX (cont)

```
<othertext>
    <d102>01</d102>
    <d104>
        BRITISH ENGLISH, A TO ZED is the thoroughly updated, revised, and expanded third
        edition of Norman Schur's highly acclaimed transatlantic dictionary for English speakers.
        First published as BRITISH SELF-TAUGHT and then as ENGLISH ENGLISH, this
        collection of Briticisms for Americans, and Americanisms for the British, is a scholarly yet
        witty lexicon, combining definitions with commentary on the most frequently used and
        some lesser known words and phrases. Highly readable, it's a snip of a book, and one that
        sorts out – through comments in American – the "Queen's English" – confounding as it
        may seem.
    </d104>
</othertext>
<othertext>
    <d102>08</d102>
    <d104>
        Norman Schur is without doubt the outstanding authority on the similarities and
        differences between British and American English. BRITISH ENGLISH, A TO ZED attests
        not only to his expertise, but also to his undiminished powers to inform, amuse and
        entertain. – Laurence Urdang, Editor, VERBATIM, The Language Quarterly, Spring 1988
    </d104>
</othertext>
```

QoSient

Argus Metadata

QoSient

Argus Metadata

- Argus Labels
  - Address based labeling - COI, function, name, organization labeling
  - Port based labeling - Function, service, application, use labeling
  - GeoSpatial labeling - Lat/Lon, Zip Code, Country Codes
  - Flow based labeling - Free form label assignments
- Argus Events and Data Correlation
  - Cross domain flow semantic enhancement
    - Merging data from one source with another
- Argus Behavioral Classifications
  - Sensor based classification
    - Producer / Consumer
    - Keystroke detection
    - Protocol Non-conformance

QoSient

Argus Labels

- ralabel(), radium(), ratop(), rasqlinsert()
- String Object
  - Colon separated <obj = attr[, attr, …., attr] >
  - Reserved keywords for ralabel specific operations
    - RALABEL_IANA_ADDRESS metadata
      - saddr=“ ….“, daddr=“ ….“, iaddr=“ ….“
    - RALABEL_BIND_NAME metadata
      - sname=“ ….“, dname=“ ….“, iname=“ ….“
    - RALABEL_IANA_PORT metadata
      - sport=“ ….“, dport=“ ….“
    - GeoIP City labeling metadata
      - scity=“ ….“, dcity=“ ….“, icity=“ ….”
    - RALABEL_ARGUS_FLOW metadata
      - [flow=]“ ….“

QoSient

IANA Address Labels

- Supports IANA IPv4 address based labeling
  - IANA IPv4 Address Space Registry Syntax
    - http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xml
    - Designation or Name Field is the label
  - Argus Address Specification - Used 1^o to specify address ranges
    - Syntax: <Address Specification Label>
      - Address Specification = Address [- Address]
        - Address
          - Specific Address - X.Y.Z.W
          - Address Range - 1.2.3.4 - 3.4.1.2
          - CIDR Addresses - 2.1.5.0/17
          - CIDR Address Range - 192.168.3.0/24 - 223.0.0.0/8
      - Label = String
- Operations
  - Multiple Address labels separated by ' , 's

QoSient

IANA Address Labels

- Example Configuration
  - IANA IPv4 Address Space Registry + Site Specific Argus Address File

```
#RALABEL_IANA_ADDRESS=yes
#RALABEL_IANA_ADDRESS_FILE="/usr/local/argus/iana-ipv4-address"
#RALABEL_IANA_ADDRESS_FILE="/usr/local/argus/iana-address-file"

# /usr/local/argus/iana-address-file
#
0.0.0.0/8-192.167.255.255/32    Internet
192.168.0.0/16                  QoSient
192.168.0.0/24                  Wired
192.168.0.67                    SMTP
192.168.1.0/24                  Switzerland
192.168.2.0/24                  Wireless
207.237.36.98                   QoSient.com
192.168.3.0/24-223.0.0.0/8      Internet
```

- Valid Values

```
1.24.4.12          saddr = "APNIC, Internet"
192.168.0.67       daddr = "Administered by ARIN,QoSient, Wired, SMTP"
```

BIND Address Labels

- Inserts domain names as metadata for flow addresses
  - Supports FQDN, Domain labels, or Local hostname labels
- Example Configuration
  - RALABEL_BIND_NAME = yes, all, saddr, daddr, iaddr
  - RALABEL_PRINT_DOMAINONLY = yes
  - RALABEL_PRINT_LOCALONLY = yes
- Valid Values

66.171.230.6	dname = "akamaiedge.net."	[n1e9.akamaiedge.net.]
74.125.226.224	dname = "google.com."	[sb.l.google.com.]
192.168.0.67	sname = "ptah"	[ptah.newyork.qosient.com.]

QoSient

IANA Port Labels

- Supports IANA transport port number based labeling
  - IANA Service Name and Transport Protocol Port Number Registry
    - http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml
    - The Service Name Field is the label
- Example Configuration

```
#RALABEL_IANA_PORT=yes
#RALABEL_IANA_PORT_FILE=“/usr/local/argus/iana-port-numbers"

# /usr/local/argus/iana-port-numbers
# Keyword          Decimal      Description                    References
# ----------       ----------   ---------------                --------------
myPortName         34512/tcp    Made this up for the example
                   34512/udp    Made this up for the example
#                               Carter Bullard <carter@qosient.com>
```

- Valid Values

  32512 sport = “myPortName”

QoSient

Argus Flow Labels

- Fall through filter style configuration
  - Flexible flow matching expressions
    - 149 argus supported objects, metrics, fields
    - Arithmetic comparisons
    - Regular expression matching from labels and user content
- Example configuration

```
#RALABEL_IANA_PORT=yes
#RALABEL_IANA_PORT_FILE="/usr/local/argus/iana-port-numbers"

# RALABEL Flow Configuration
filter="port domain and pcr gt 0.0"      label="dnsFault"              cont
filter="ploss gt 1.0"                    label="lossFault"             cont
filter="not co US and srcid interior"    label="InteriorAccessFault"   cont
grep="tivo"                              label="Tivo"
```

- Valid Values

  flow = "dsnFault, InteriorAccessFault, Tivo"

Argus GeoIP Labels

- Geo Location based labeling
  - Support focused on MaxMind's GeoIP API and databases
    - Can provide country codes, name, region ,city ,postal code, latitude, longitude, metro_code, area_code, continent_code, netmask value
    - Can be applied to either source, destination or intermediate addresses
- Example configuration

```
#RALABEL_GEOIP_CITY="saddr, daddr: lat/lon"
#RALABEL_GEOIP_CITY="saddr, daddr,inode: off,cont,lat,lon"
#RALABEL_GEOIP_CITY_FILE="/usr/local/share/GeoIP/GeoIP.dat"
#RALABEL_GEOIP_V6_CITY_FILE="/usr/local/share/GeoIP/GeoIPv6.dat"
```

- Valid Values

```
dcity=37.441200,-121.990501
scity=40.714298,-74.005997:dcity=42.287300,-71.352402
```

QoSient

Argus Events

- Argus event is a non flow data information element that represents attributes applicable to an observation domain, at a specific moment in time
- Events schema primarily designed to support syslog style messaging between argus components, but extended to support general purpose messaging of any type.
- Complex Argus Data Type
  - Argus Record Header
  - Argus Transport Header
  - Argus Event Time Header
  - Argus Event Data Structure
    - Type, cause, status, target, facility, severity, message and metadata
    - Currently supports 4 targets: Database, Syslog, File and Terminals
    - Metadata object is generally an XML data document, defined as a BLOB

Argus Events

```
event[49241]=  2013/01/04.12:47:16.733468:srcid=192.168.0.68:prog:/usr/local/bin/argus-lsof
<ArgusEvent>
  <ArgusEventData Type = "Program: /usr/sbin/lsof -i -n -P">
    COMMAND     PID           USER   FD   TYPE             DEVICE SIZE/OFF   NODE NAME
    mDNSRespo    53 _mdnsresponder   56u  IPv4 0xbb72da10d0ede5ff      0t0    UDP *:50451
    awacsd       69           root  241u  IPv4 0xbb72da10db05dce7      0t0    TCP 192.168.0.68:57367->17.172.208.94:443 (CLOSED)
    apsd         71           root   10u  IPv4 0xbb72da10d2b73327      0t0    TCP 192.168.0.68:53556->17.149.32.65:443 (ESTABLISHED)
    blued        72           root    4u  IPv4 0xbb72da10cd45d23f      0t0    UDP *:*
    ntpd         75           root   20u  IPv4 0xbb72da10d0f08abf      0t0    UDP *:123
    radium      110           root   10u  IPv4 0xbb72da10d128ee77      0t0    TCP 192.168.0.68:49166->192.168.0.68:561 (ESTABLISHED)
    radium      110           root   11u  IPv6 0xbb72da10d137878f      0t0    TCP [::1]:562->[::1]:49171 (ESTABLISHED)

    [snip]

    JavaAppli 16726         carter   57u  IPv4 0xbb72da10e47bb41f      0t0    TCP 127.0.0.1:53004->127.0.0.1:53002 (ESTABLISHED)
    JavaAppli 16726         carter   69u  IPv4 0xbb72da10e48505af      0t0    TCP 127.0.0.1:53005->127.0.0.1:53003 (ESTABLISHED)
    iTunes    19801         carter   17u  IPv4 0xbb72da10e4831197      0t0    TCP *:3689 (LISTEN)
    iTunes    19801         carter   20u  IPv6 0xbb72da10d1564b6f      0t0    TCP *:3689 (LISTEN)
    iTunes    19801         carter   45u  IPv4 0xbb72da10e31a15af      0t0    TCP 192.168.0.68:61015->17.171.36.30:80 (CLOSED)
    iTunes    19801         carter   55u  IPv4 0xbb72da10d6a49197      0t0    TCP 192.168.0.68:60968->17.171.36.30:80 (CLOSED)
    Notes     68535         carter   20u  IPv4 0xbb72da10dcbb2e77      0t0    TCP 192.168.0.68:49899->17.172.34.97:993 (ESTABLISHED)
    Keynote   68546         carter    8u  IPv4 0xbb72da10e2c34327      0t0    TCP *:49901 (LISTEN)
    raevent   69821         carter    5u  IPv6 0xbb72da10d1a78fcf      0t0    TCP [::1]:51255->[::1]:562 (ESTABLISHED)
    perl5.12  69824           root    4u  IPv6 0xbb72da10cea08b6f      0t0    TCP *:561 (LISTEN)
    perl5.12  69824           root    6u  IPv4 0xbb72da10cd45de7f      0t0    UDP *:*
    perl5.12  69824           root    8u  IPv6 0xbb72da10d132bfcf      0t0    TCP 192.168.0.68:561->192.168.0.68:49166 (ESTABLISHED)
    perl5.12  69824           root    9u  IPv6 0xbb72da10d1a793af      0t0    TCP [::1]:561->[::1]:58040 (ESTABLISHED)
  </ArgusEventData>
</ArgusEvent>
```

QoSient

Argus Event Correlation Labels

- Event correlators like radium, ratop, rasqlinsert
  - Flexible flow matching expressions
    - 149 argus supported objects, metrics, fields
    - Arithmetic comparisons
    - Regular expression matching from labels and user content
- Example .rarc configuration

  RA_CORRELATE_EVENTS="yes"

- Valid Values

  pid=136:usr=root:app=radium

  pid=18845,68044,18871,18898,18927,18954,18986:usr=root:app=perl5.16,argus

  pid=216:usr=carter:app=Mail

QoSient

Argus Services Labels

- Captured user data processing
  - raservices.1 + rauserdata.1
    - Complex application fingerprint matching
    - Matches application based on persistent syntax markers
    - Provides exact matches and best guesses
- Example raservices.dat file

```
Service: https           tcp port 443 n = 233603 src = "160300005D010000590300          "  encrypted
Service: microsoft-d     tcp port 445 n = 316171 src = "000000    534D4272000000001853C8"  dst = “000000B6FF534D4272000000009853C8"

Service: synoptics-trap  tcp port 412 n =  41573 src = "244D794E69636B206A6F6D626C65727C"  dst = "244D794E69636B206864696E6765727C"
Service: synoptics-trap  udp port 412 n =  75315 src = "24535220  64                   “  dst = "2450696E67200000"
Service: synoptics-trap  udp port 412 n =  27341 src = "24535220               20      "  dst = “2450696E67200000"
Service: synoptics-trap  udp port 412 n =  11783 src = "24535220                    65"  dst = "2450696E67200000"
Service: synoptics-trap  udp port 412 n =  10110 src = "24535220  61             2053  "  dst = "2450696E67200000"
Service: synoptics-trap  udp port 412 n =   9787 src = "24535220              61      "  dst = "2450696E67200000"
Service: synoptics-trap  udp port 412 n =   4634 src = "24535220        64    6E      "  dst = "2450696E67200000"
Service: synoptics-trap  udp port 412 n =   1132 src = "24535220          20          "  dst = “2450696E67200000”

Service: ntalk           udp port 518 n = 627312 src = "0101000000000000000020000000000000"  dst = "010101  00000000                "
Service: router          udp port 520 n =  52906 src = "0101000000  0000000000000000000"
```

- Valid Values

  srv=ntalk

QoSient

Argus Metadata Label

- The aggregate metadata label is :
  - A string
  - colon ':' separated list of attributes
  - User is not restricted in any way to structure or syntax of configured labels, however, ':' separation and aggregation processing may generate unexpected results.

- Example Labels

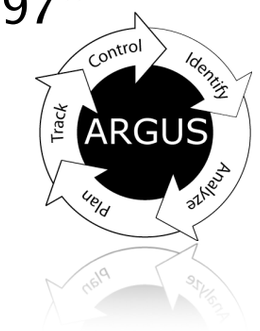
```
label = "saddr=QoSient,Wired,SMTP: daddr=QoSient,Wired,MySQL:
         sname=ptah: dname=osiris: dport=mysql: srv=mysql:
         flow=normal: pid=216: usr=carter: app=Mail"

label = "saddr=Internet,ARIN: daddr=QoSient: flow=DarkSpace"

label = "pid=216: usr=carter: app=Mail"

label = "scity=40.714298,-74.005997: dcity=37.304199,-122.094597"
```

QoSient



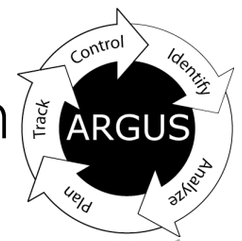
Metadata Generation

QoSient

Argus Metadata Lifecycle

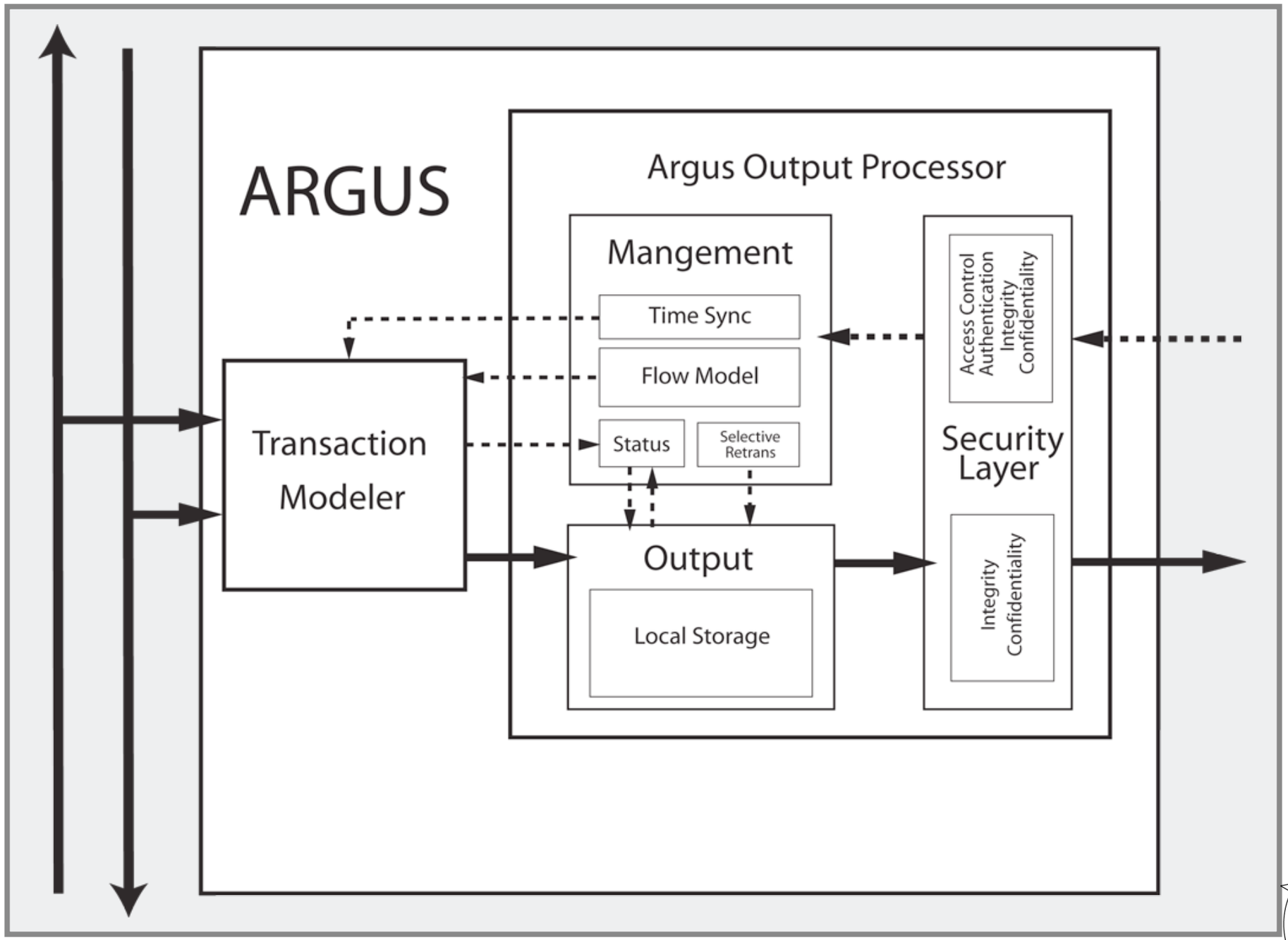
- Argus-3.x generates packet dynamics metadata
  - Wire-line objects, metrics and behaviors
  - Intended for terminal system consumption
- Ra* metadata generation
  - At any stage along the flow system pipeline
    - Multi-scope correlation and reduction
    - Micro, Macro and Superflow Aggregation
    - Multi-dimensional correlation and enhancement
- Metadata can be processed at each stage
  - Some metadata represents pipeline messaging
    - Metadata consumer is next stage in the pipeline
    - Metadata pruned, filtered, aggregated or rejected
  - Most metadata intended for unknown user / application

QoSient

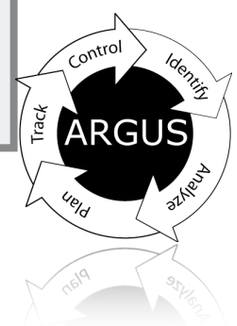


Argus Sensor Design

Packets to Flows



QoSient



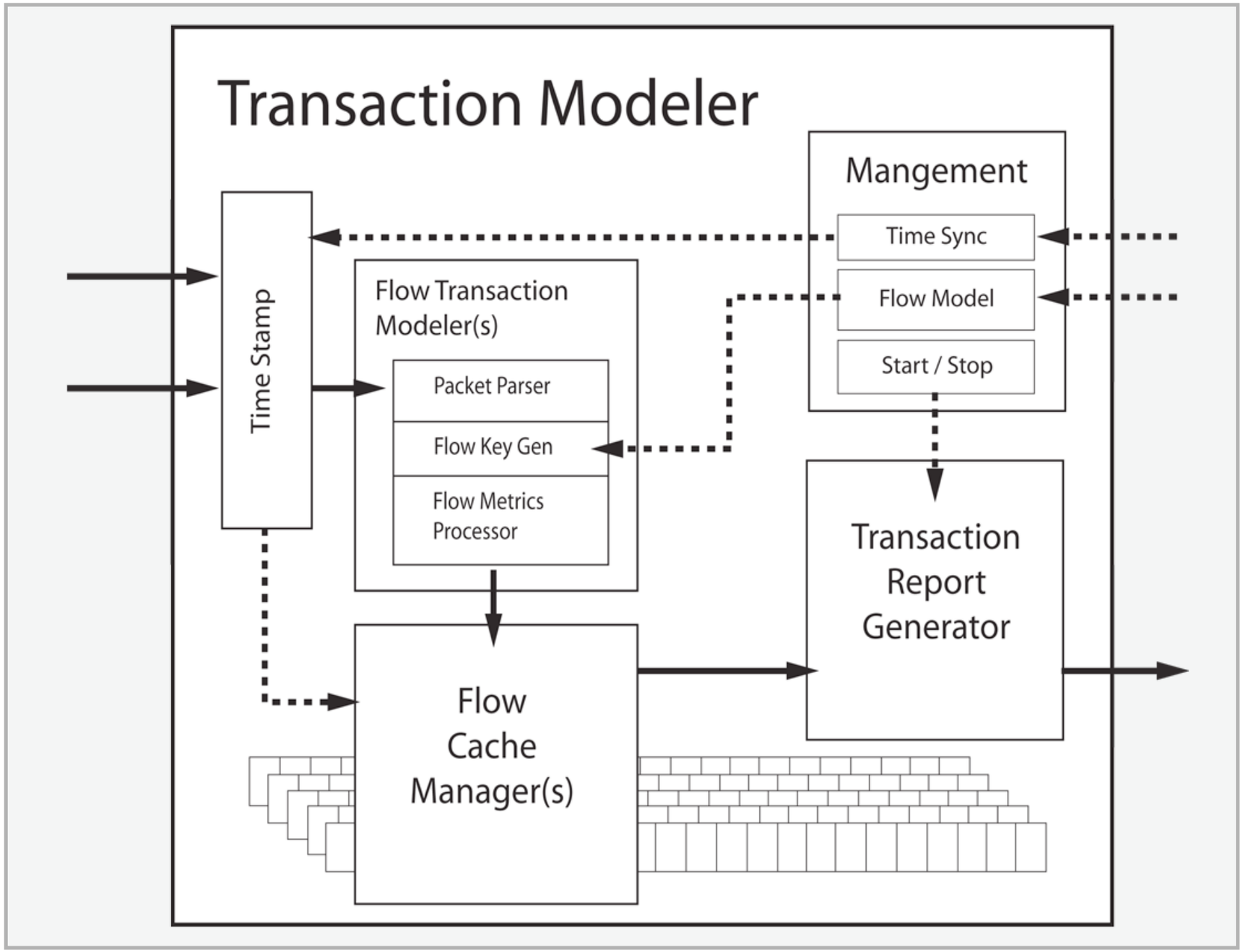
Argus Sensor Metadata

- Data that is not derived from packet contents
  - Complex P1/P2 flow models
    - Broadcast / unicast flow tracking
    - ICMP event mapping
  - Transactional state information
    - Bi-directional flow tracking (connectivity, availability metrics)
    - TCP connection establishment time
  - Packet dynamics metrics
    - Loss detection
    - Packet size reporting
    - Inter-packet arrival and jitter values
    - Keystroke detection metrics
- This is NOT DPI / IDS style classification

QoSient

Argus Sensor Design

Transactional Processor



QoSient

End-to-End Situational Awareness

Network Optimization - Black Core Mesh

SNMP RMON Element Statistics/Traps
Interface Status / Transitional Events
Bulk Link Statistics

System Layer 2-7 Flow Data
System Communication Efficiency
Connectivity / Availability
Offered Load / Loss / Jitter
One-Way Delay (GPS synchronization)
Round Trip Delay

Comprehensive Layer 2-7 Flow Data
Site Communication Efficiency
Enterprise Communication Efficiency
Site Offered Load / Loss / Jitter
Network Transit Times

Comprehensive Layer 2-7 Flow Data
End-to-End Communication Efficiency
Reachability / Connectivity
Received Load / Loss / Jitter
Network Transit Times

Comprehensive Layer 2-4 Flow Data
ISP Communication Efficiency
Ingress Available Capacity / Loss / Jitter
One-Way Delay (GPS Synchronization)
Network Path Status

Comprehensive Layer 2-4 Flow Data
Network Path Assurance / Status
Reachability / Availability Assessment
One-Way Delay (GPS Synchronization)

Core Service Provider Management Domain

Enterprise Management Domain

Comprehensive Flow Monitor

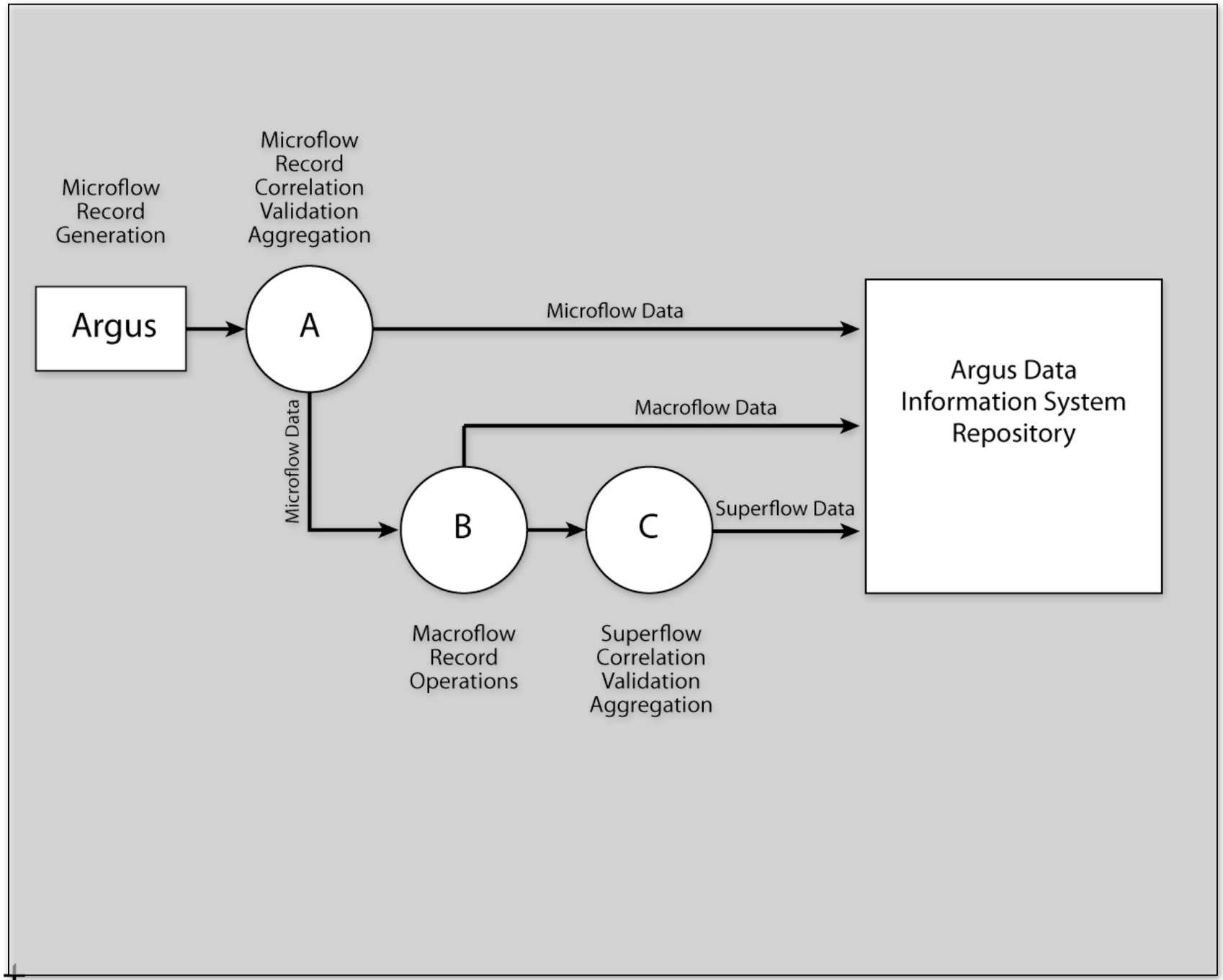
SNMP RMOM Style Monitor

Information System Repository

ARGUS
Control
Identify
Analyze
Plan
Track

Flow Data Processing Pipeline

Data Flow Machine Architectures



QoSient

Metadata Generation Process

- Classification based metadata generation
  - Object and metrics matching
  - Set operations to provide semantic enhancement
    - This address is in this group (Community of Interest identification)
    - This amount of data is classified as "exceeded contract".
- Cross domain correlation metadata
  - Correlate records from multiple observation domains
    - Correlation specific semantics
  - Select-Join like operations between information systems
    - DNS name mapping
    - DHCP based user assignments

QoSient

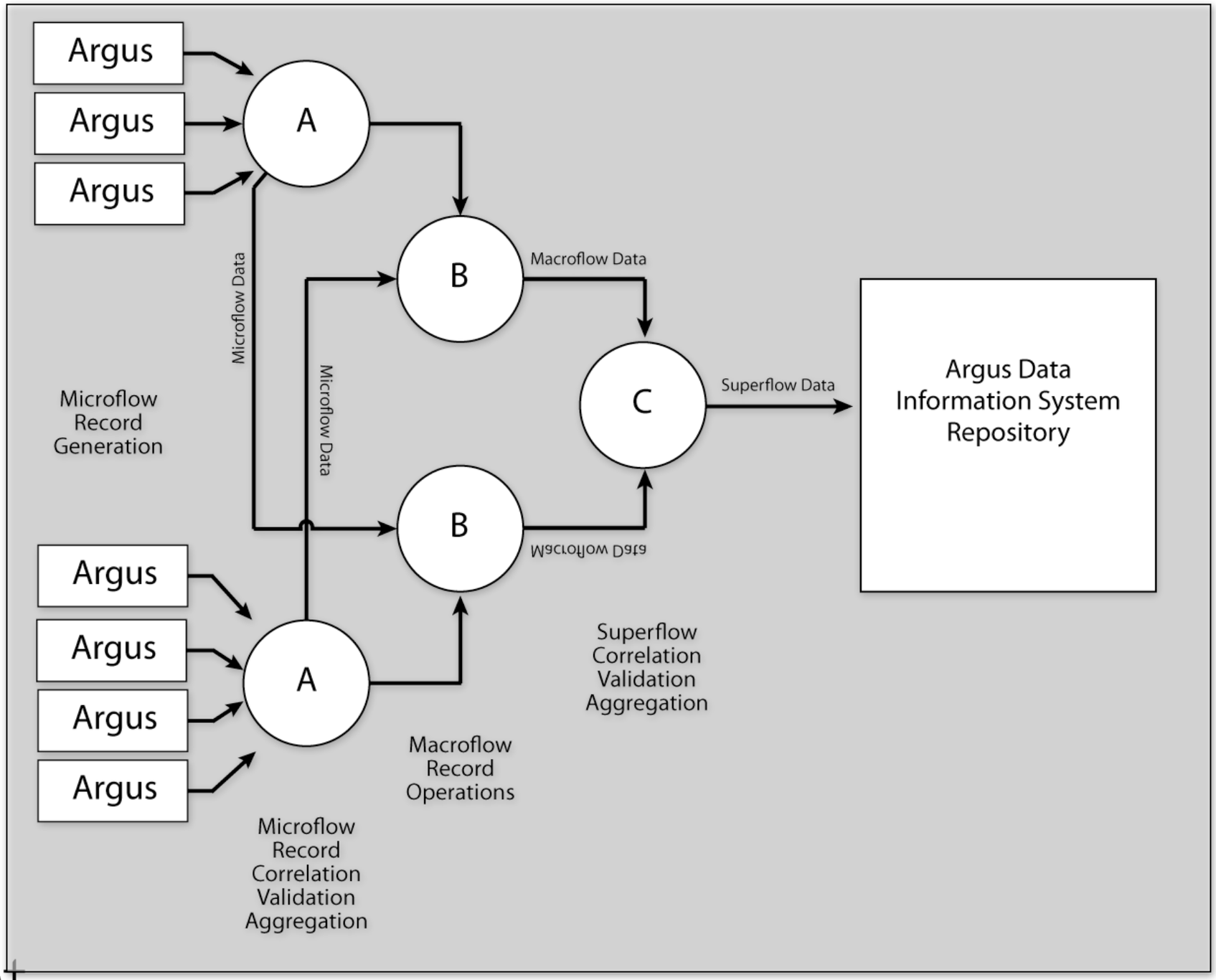
Data Document Initiative V3.1 (DDI)

- DDI Design Rules apply to argus metadata management
  - Persistent sections should be separate from dynamic information.
  - Information modules should follow life cycle paths
  - Discovery information should in non-specialized modules
  - Links should be unidirectional to avoid loops
  - Links should point back in time
  - All comparisons are pair wise, comparing source with target
  - Groups inherit down the tree unless clear override provided
  - Metadata will be expressed in ways which support both human-readability and machine-processing

QoSient

Flow Data Processing Pipeline

Data Flow Machine Architectures



QoSient

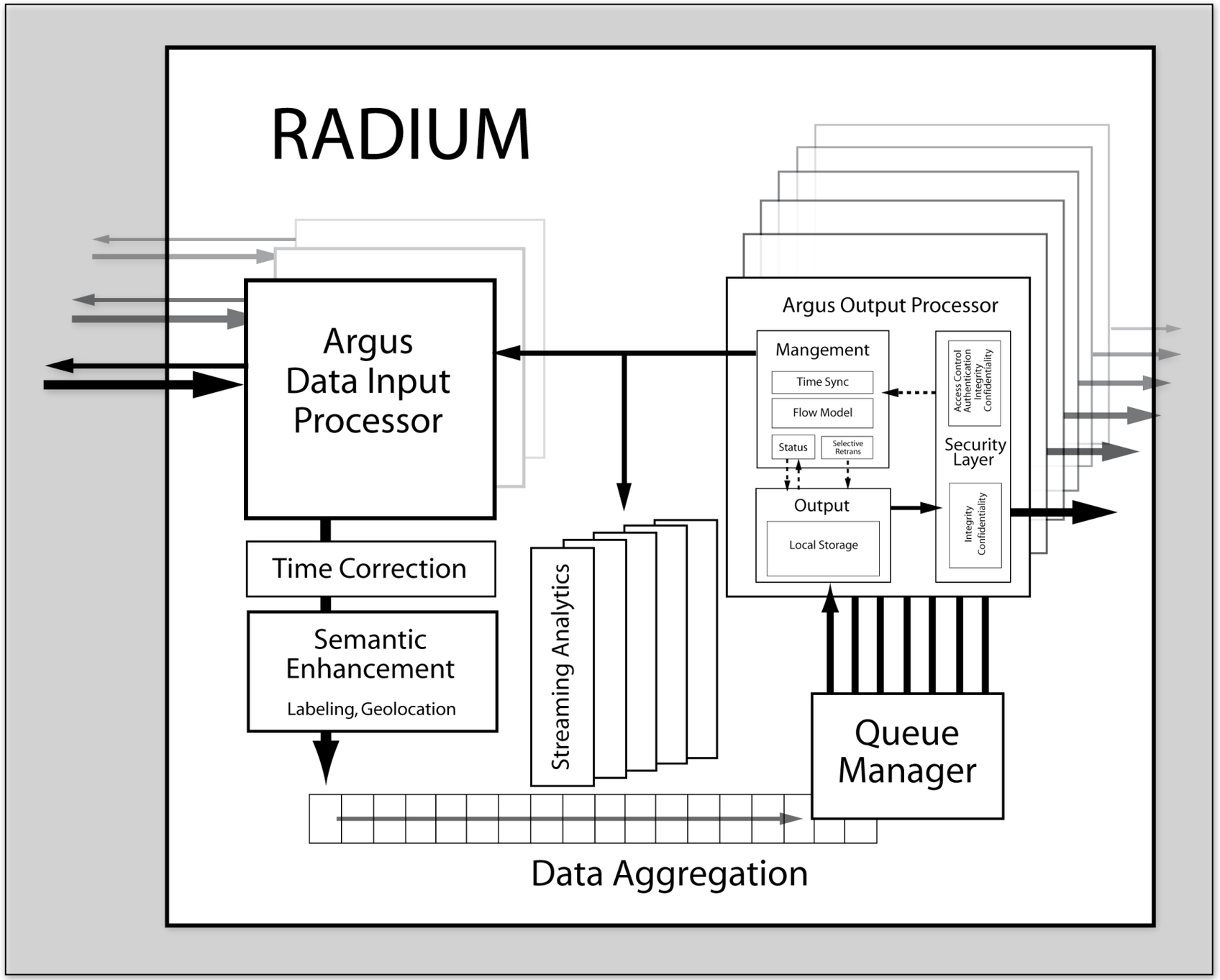
Metadata Generation

- Domain Specific
  - Origin / Observation Domain Scoping
  - Time
  - Data generation / collection methods
  - Data Compression
- Scope Specific
  - Services, application and protocol specific metadata
- Target Processing Issues
  - Cross domain keying - to support real time correlation
  - Near real-time streaming

QoSient

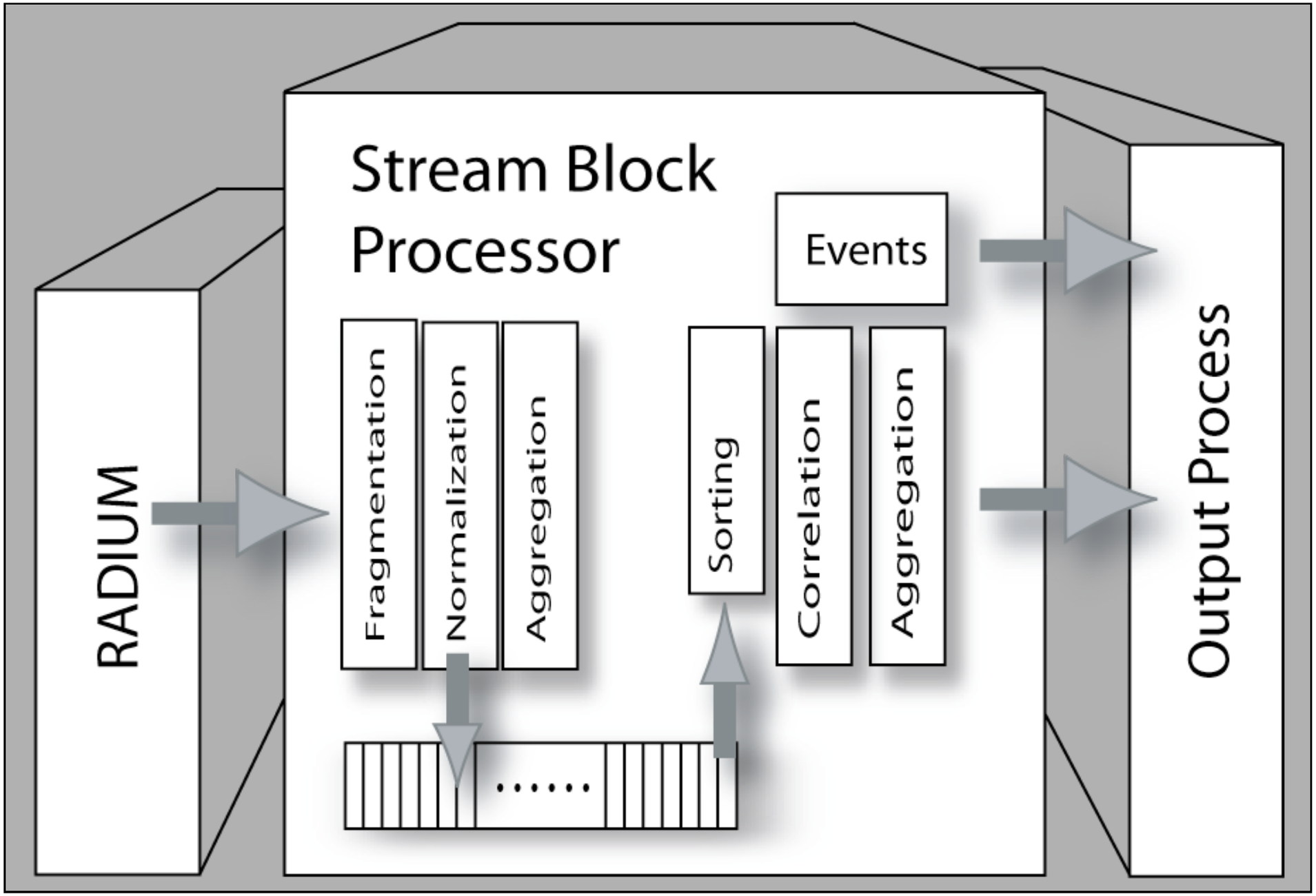
Argus Collection Design

Radium Process



Argus Processing Design

Stream Block Processor



Network Flow Metadata Transport

QoSient

Argus Metadata DSR

- Metadata transported in Argus Label DSR
  - Label is currently an ASCII string.
  - Current length limits are 1024 bytes
  - Extending that to 32K bytes in 3.0.10 and providing optional compression.

```
struct ArgusLabelStruct {
   struct ArgusDSRHeader hdr;
   union {
      char *svc;
      char *label;
   } l_un;
};
```

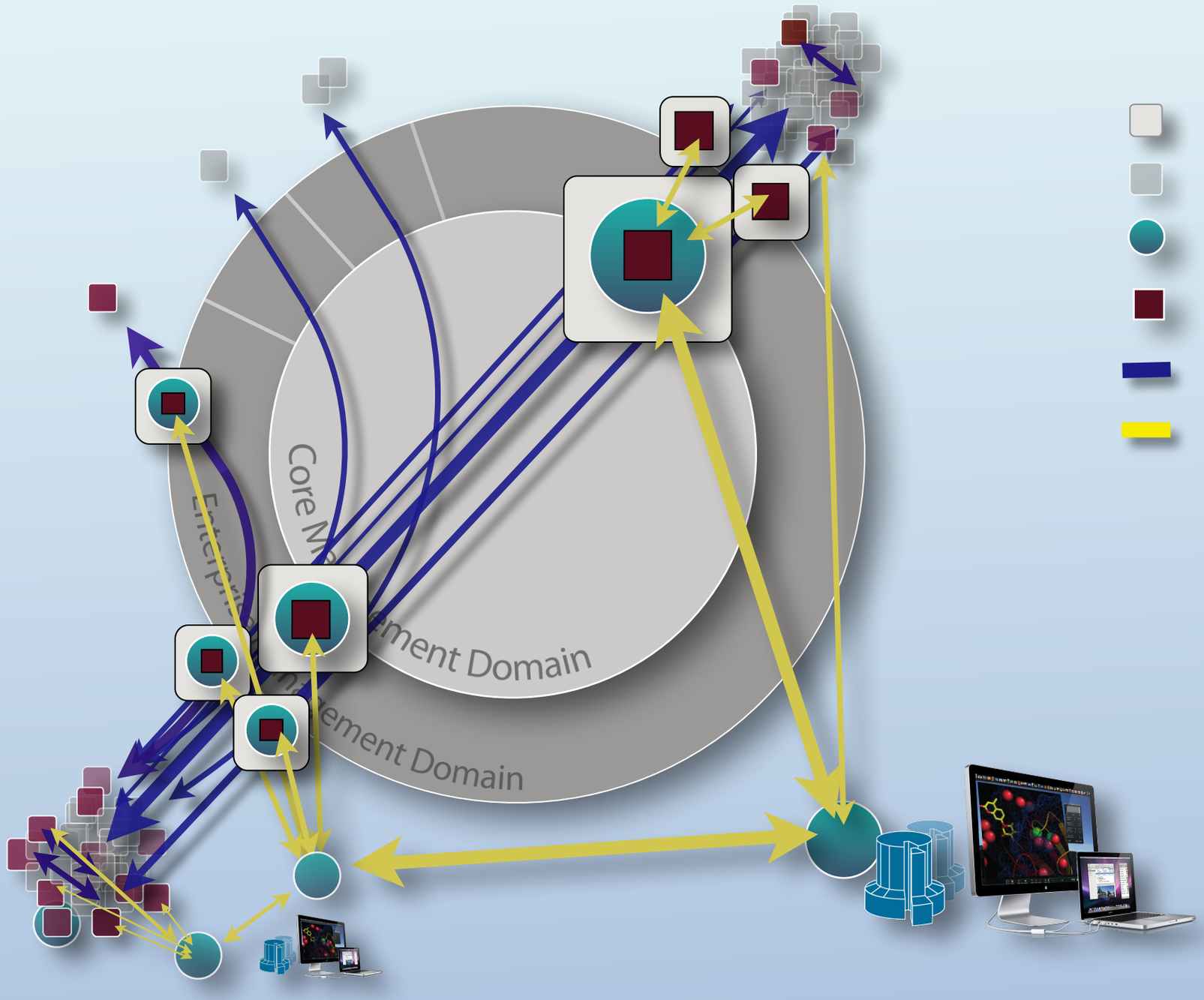
```
struct ArgusDSRHeader {
   unsigned char type;
   unsigned char subtype;
   union {
      struct ArgusDSRfixLen fl;
      struct ArgusDSRvar8bitLen  vl8;
      struct ArgusDSRvar16bitLen vl16;
   } dsr_un;
};
```

Network Flow Metadata Processing

QoSient

Complex Comprehensive Awareness

Local and Remote Strategies



White/Visible Node

Black/Non-Visible Node

Metadata Generation Node

Argus Sensor

Data Plane

Situational Awareness Data

QoSient

Argus Metadata

- Argus metadata must meet the minimum requirements for argus data generation, transport, processing and storage.
  - Filtering
  - Stripping / Removal
  - Aggregation
  - Anonymization
  - Printing / Display

QoSient

Argus Metadata
Filtering

- Flow Model based metadata
  - ra - icmpmap or intpkt gt 0.12 or src jit gt 1.0
- Label based metadata
  - Regular expression matching
    - ra -M label="saddr=.*ARIN.*:"
- Behavioral metadata
  - ra - nstrokes gt 0

QoSient

Argus Metadata
Stripping / Removal

- All ra* programs support metadata stripping
  - Strip specific data on input
    - ra -M dsrs="-label,-behavior, -jitter"
  - Indirectly remove dsrs on input
    - ra -M dsrs="time,trans,mac,flow,metrics"
- Custom client programs use DSR library support
  - Simple strategies to control DSR use.
    - argus->dsrindex - active DSR bitmap

QoSient

Argus Metadata Aggregation

- Merging 2 argus records with labels
- Operations apply to object sets

$L_1[obj_1=value_{1,1},value_{1,2}:obj_2=value_{2,1},value_{2,2}:obj_3=value_{3,1},value_{3,2}]$

$L_2[obj_1=value_{1,1},value_{1,2}:obj_2=value_{2,1},value_{2,2},value_{2,3}]$

  - Equivalence rejection
    - retain if all obj values are equal
  - Union - without redundancy
    - retain all objects, and all values, but no repeat values
    - Limit number of objects. Limit number of values per object.
  - Intersection
    - retain only shared objects and shared values.

QoSient

Argus Metadata Anonymization

- General anonymization of metadata
  - Strip label metadata by default on output
    - Complex, custom objects not anonymizeable
    - Equivalent to user data anonymization
- Anonymizing Geo-spatial data
  - Remove the data
  - Random / fixed offset, area aliasing, de-resolution, regional distortion
- Custom anonymization
  - Requires original label rejection with rules based regeneration.
  - Rejection of unknown semantics
  - Extended anonymization configuration

QoSient

Argus Metadata
Printing / Display

- Simple String, XML
- Display Strategies
- Timeline Event
- Clustering
- Geospatial Data Mapping

QoSient

Metadata

- Network Flow Metadata Processing
  - Data Strategy
    - Embedded vs Relational Model
  - Aggregation
  - Selection / Filtering

QoSient

Data Correlation Strategies

- Flow Attribute Matching
  - Flow Identifiers
  - Protocol specific identifiers
  - Packet Dynamics
    - Inter-packet arrival times
    - Packet Size
  - Transactional Dynamics
    - Duration
- Non-flow Attribute Matching
  - Non Flow Key Identifiers
  - Cross Domain Transactional Keys
  - Time

QoSient

Flow - Flow Correlation

- Time Synchronization
- Packet dynamics (PD) can be used to detect stepping stone techniques.
- New understanding of packet dynamics can provide additional awareness needed for successful network path assurance, man-in-the-middle detection, stepping stone detection, replay and attribution.

QoSient

Network Flow Metadata Storage

QoSient

Metadata Storage

- Intra flow label storage
  - Keep labels in flow records as they were received / processed.
    - Retains information integrity - critical for evidence
    - Maintains information granularity
    - Supports ad-hoc data mining strategies
      - Extends metadata utility
- Inter flow label storage
  - Labels stored in a separate data structure
    - Cross domain indexing
    - Provides opportunity for data reduction
    - Enhances structured data mining
- Suggest Both Strategies

QoSient

Tutorial Objectives

- Define Network Flow Metadata
- Discuss Issues in Metadata Generation, Transport, Processing and Storage
- Describe Metadata Support in Argus
  - Strategies for Metadata Generation
  - Methods for Metadata Processing
    - Transport Issues
    - Aggregation
    - Correlation
  - Metadata Storage
- Conclusions

QoSient

Argus Metadata Support

QoSient

Flow - Non Flow Correlation

- Replay attack detection
  - Bi-Directional Protocol Time Uncoupling
- Stepping stone detection
  - Two completely independent flows, that share the same instantaneous burst behavior and packet size frequency distribution (shifted for encapsulations)
- Man vs Machine detection
  - Interactive vs Non-Interactive Session Detection
  - Packet, transaction and session jitter analysis
- Man-in-the-middle detection
  - Pass Thru - Detectable one-way latency, hop count, path resource modifications
  - Proxy - Connection setup time modifications, header attribute changes
- Performance as an Asset that needs Protection
  - Path Availability, Bandwidth, Latency, Jitter, MTU, ...
  - Continuous One-Way latency determinations

QoSient

Using Metadata

QoSient

Situational Awareness

QoSient

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

QoSient

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

Situational Awareness System

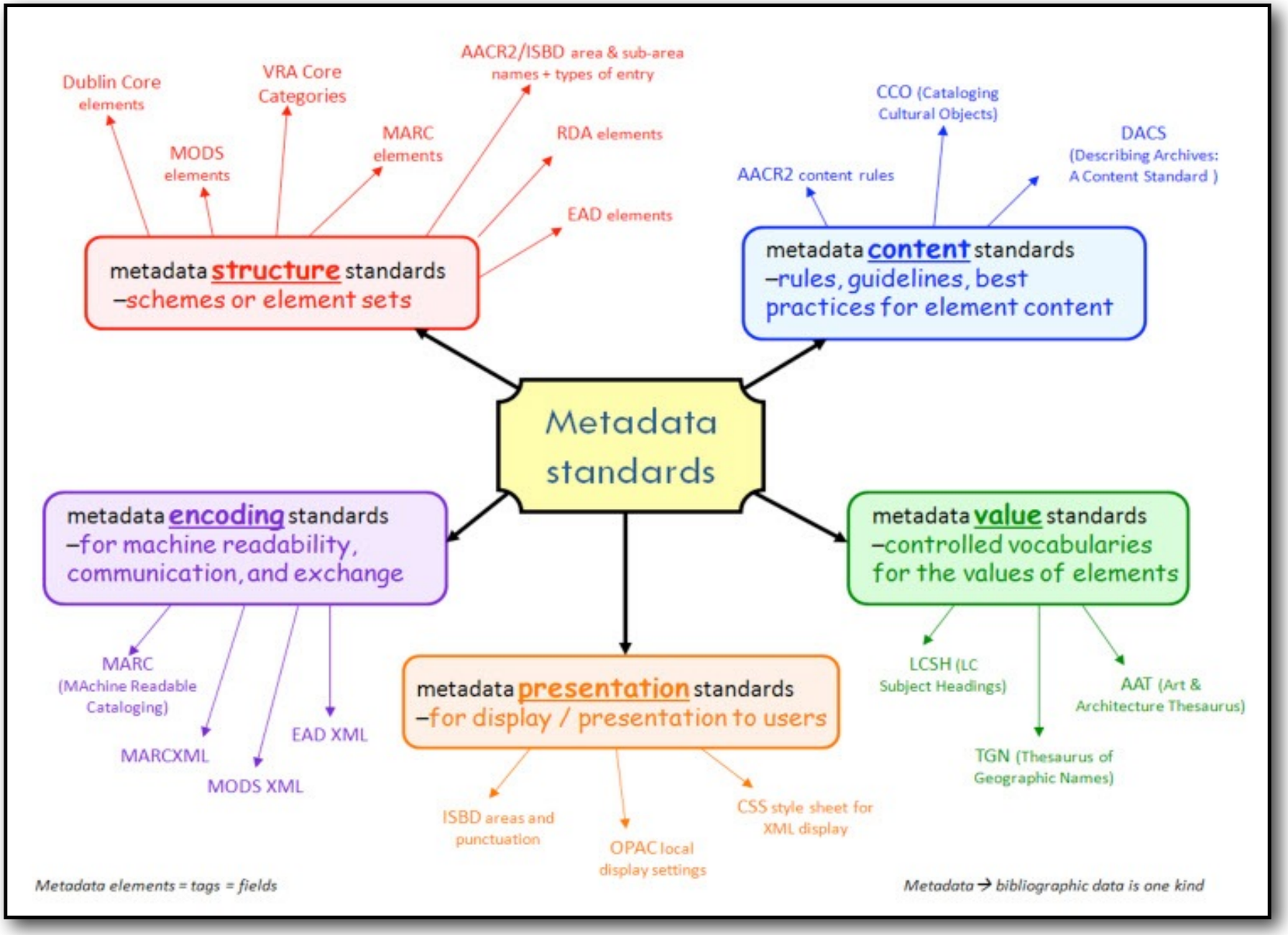
Basic design is local sensing, data collection and management, with local near real time data processing and large scale data sharing to support multi-dimensional control plane comprehension.

- Federated Database Model
  - Access controlled by local administrative domain (scoping)
  - Cloud-like distributed processing and query support
  - Flexible data management strategies
  - Large numbers of simultaneous users
- Near real-time information availability
  - Register for information of interest
  - Complex data processing / aggregation / enhancement
  - Large scale data correlation processing
  - Anonymization

QoSient

Supporting Slides

Typology for Metadata Standards



Miller, Steven J. , *Metadata for Digital Collections: A How-To-Do-It Manual.* New York: Neal-Schuman, 2011. ISBN: 9781555707460

QoSient

Introduction to Argus

- Discuss the problem space
- Describe Argus design and implementation
- In the context of approaching some real problems
  - Cyber Security
    - Insider Threat protection through Non-Repudiation
  - Degradation of Service
    - Identification
    - Attribution
    - Mitigation

QoSient

Argus

http://qosient.com/argus

- Argus is a network activity audit system

  Argus was officially started at the CERT-CC as a tool in incident analysis and intrusion research. It was recognized very early that Internet technology had very poor usage accountability, and Argus was a prototype project to demonstrate feasibility of network transactional auditing.

- The first realtime network flow monitor (1989)
- Top 100 security tools used in the Internet today
  - Generates detailed network resource usage logs
  - Source of historical and near realtime data for the complete incident response life cycle
- Designed to provide useful data for network
  - Operations - Service availability and operational status
  - Performance - End-to-end assessment of user traffic
  - Security - Audit / Non-Repudiation

QoSient

Argus History

- Georgia Tech (1986)

  Argus was the first data network flow system. Started at Georgia Tech, Argus was used as a real-time network operations and security management tool. Argus monitored the Morris Worm, and was instrumental in monitoring the "Legion of Doom" hacking incident.

- CERT/SEI/Carnegie Mellon University (1991)

  Argus was officially supported by the CERT as a tool in incident analysis and intrusion research. Used to catalog and annotate any packet file that was provided to the CERT in support of Incident Analysis and Coordination, it was a focal point for research in intrusion analysis and Internet security.

- Argus Open Source (1995 - Present)

  Transitioned into public domain in 1995. Supported by CMU and CERT/SEI at many levels including the current argus developers mailing list.

  Used now by a very large number of educational, commercial and governmental sites for network operations, security and performance management.

  Top 100 Security Tools worldwide

Who's using Argus?

- U.S. Government
  - DoD Performance/Security Research - Gargoyle
    - https://software.forge.mil/projects/gargoyle
    - JCTD-Large Data, CORONET, NEMO, JRAE, Millennium Challenge
  - Tactical Network Security Monitoring / Performance Analysis
    - Naval Research Laboratory (NRL), DISA, General Dynamics, IC
- Network Service Providers
  - Operational/Performance Optimization
  - Acceptable Use Policy Verification
- Educational (1000's of sites world-wide)

Carnegie Mellon University	Enterprise wide near realtime network security audit
Stanford University	Distributed security monitoring
University of Chicago	Network security research
New York University	Acceptable use policy verification

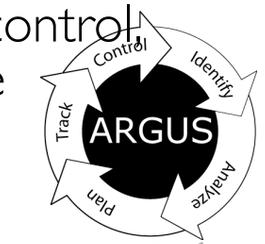
- ISPs, Enterprises, Corporations, Individuals

QoSient

Network Situational Awareness

- Argus is designed to be THE network SA sensor
  - Ubiquitously deployable DPI traffic sensor
  - Comprehensive (non-statistical) traffic awareness
  - Provides engineering data, not business intelligence
    - Detailed network transactional performance
    - Network fault identification, discrimination and mitigation
    - Customer gets the primitive data, not just reports/alerts
  - Near realtime and historical capabilities
  - Packet capture replacement
- Supporting a large number of SA applications
  - Advanced Network Functional Assurance (Operations)
    - End-to-End transactional performance tracking (data and control plane)
    - Network component functional assurance (NAT, reachability, encryption)
    - Policy enforcement verification/validation (Access control, path, QoS)
  - Advanced Network Optimization (Security and Performance)
    - Network entity and service identification, analysis, planning tracking and control, including baselining, anomaly detection, behavioral analysis and exhaustive forensics

QoSient



Problem Space

QoSient

US Cyber Security Focus

- US Cybersecurity focus is shifting
  - Shifting from cyber warfare, back to cyber
- Structured around 4 basic themes
  - Designed-in Security - inherent resistance to attack
  - Tailored Trustworthy Spaces - flexible, adaptive, distributed trust
    - Focus → Wireless Mobile Networks
  - Moving Target - dynamism as a protection mechanism
    - Focus → Deep Understanding of Cyberspace
    - Focus → Nature-Inspired Solutions
  - Cyber Economic Incentives
- Supporting National Priorities
  - Health IT, Smart Grid, Financial Services, National Defense, Transportation, Trusted Identities, Cybersecurity Education

QoSient

Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program, US NSTC, Dec 2011

DHS Cybersecurity Strategy

- Protecting Critical Information Infrastructure
  - Reduce Exposure to Cyber Risk
  - Ensure Priority Response and Recovery
  - Maintain Shared Situational Awareness
  - Increase Resilience
- Strengthening the Cyber Ecosystem
  - Empower Individuals and Organizations to Operate Securely
  - Make and Use More Trustworthy Infrastructure
  - Build Collaborative Communities
  - Establish Transparent Processes
- Strategy refers to real-time and near real-time mechanisms
  - "... to collect and exchange information in real-time ..." - situational awareness
  - "... capabilities will be communicated in near real-time ..." - resilience
  - "... near real-time machine-to-machine coordination ..." - strengthening
  - "... acting collectively in near real-time to anticipate ..." - collaboration

QoSient

The Cybersecurity Strategy for the Homeland Security Enterprise, US Department of Homeland Security, Nov 2011

DISA Convergence Strategy

Long Term Security Components

- Network Normalization
  - Reduce Network Classifications to Two
  - Formal Security Boundaries
- Shift Protection Strategy Framework
  - Perimeter to Transactional Information Protection
  - Granular End-2-End Security Controls
  - Protected Information Exchange
    - Ensure Confidentiality, Integrity and Availability
- Enterprise Service Management Portfolio
- Mission Assurance Services Portfolio

QoSient

Theoretical Security Threats and Countermeasures

Countermeasures		Threat				
		Unauthorized			Degradation of Service	Repudiation
		Use	Modification	Disclosure		
Authentication	Cryptographic	x		x		
Integrity			x			
Confidentiality				x		
Access Control		x	x	x	x	
Non-Repudiation (audit)		x	x	x	x	x

Derived from ITU-T Recommendation X.805
Security Architecture for Systems Providing End-to-End Communications

Primary Security Countermeasure

Secondary Security Countermeasure

QoSient

Non-Repudiation

- Most misunderstood countermeasure *
  - ITU-T Recommendation X.805 security dimension
    - *Prevent ability to deny that a network activity occurred*
- Principal source of true deterrence
  - Non-repudiation provides comprehensive accountability
  - Creates concept that you can get caught
- Argus approach to network non-repudiation
  - Generate data to account for all network activity
    - Comprehensive Network Transactional Audit
    - Mechanism specified by DoD in NCSC-TG-005
      - The Red Book - Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria (1987)
  - Focus on all X.805 Security Planes
    - User, Control and Management network activity

QoSient

* Crypto-technical redefinition of non-repudiation by Adrian McCullagh in 2000 to apply only to digital signatures has created a great deal of confusion. While you can have repudiation of a signature, it's not the only thing you can repudiate.

Non-Repudiation Concepts

ITU X.813

Information Technology

Open Systems Interconnection

Security Frameworks in Open Systems: Non-repudiation Framework

"The Non-repudiation service involves the generation, verification and recording of evidence. ..... Disputes cannot be resolved unless the evidence has been previously recorded."

The service provides the following facilities which can be used in the event of an attempted repudiation:

- generation of evidence
- recording of evidence
- verification of generated evidence
- retrieval and re-verification of the evidence

QoSient

Why Non-Repudiation?

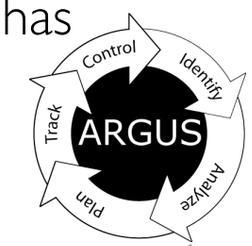
When it exists and structured well, you get

- Effective information for incident response
  - Fundamental ground truth - if its not there, it didn't happen
  - Classical forensics support
  - Evidence suitable for criminal and civil complaints
- Enhanced network situational awareness
  - Network Service Behavioral Baselining
    - Who is really using my DNS servers?
    - What is generating Email in my enterprise?
    - How much data did he transmit last night?
  - Network Policy Enforcement Assurance
    - Are my IPS / IDS / Firewall protections working?
  - Network Fault Attribution
    - Is it an attack? Is it real? Is it a bug? Is it Fred?

Enables enhanced analytics, simulation and 'what if' analysis

- This host polls this email server every 60.0023 +/- 0.0004231 seconds and has been doing that for 17.6243 months, with only 27 outages lasting .....
- Will this new access control policy, break anything?

QoSient



Achieving Non-Repudiation

- Comprehensive Activity Accountability
  - Complete Activity Sensing and Reporting
  - Develop Information System with Formal Properties
    - Fundamental ground truth (if its not there, it didn't happen)
- Accurate and Efficient Activity Representation(s)
  - Stored data must represent actual activity
    - Attribute verifiability
      - Must be unambiguous with regard to object identification
      - Must have a relational algebraic correctness
    - Time synchronization and precision
      - Must convey correct order of events
- Fundamental Data Utility
  - Formal and Mature Data Model
  - Useful Data Availability Properties
  - Effective Storage and Retention Strategies

QoSient

Real world issues

Non-Repudiation systems must support addressing real world issues

- Must capture adequate forensics data for incident response
  - Enterprise focused on contemporary security issues
  - Policy enforcement verification validation
  - Provide high level of semantic capture/preservation
  - Support complex behavioral analysis through packet dynamic awareness
- Should support real time awareness
  - Data presence information - access control verification
  - Contribute large scale multi-level hierarchical distributed situational awareness

Provide real deterrence

In a perfect world, you would have a single source for all your network forensics data

- Support near real-time and historical requirements
- FISMA continuous network monitoring role

QoSient

Real world issues

Incident Response

- NASA calls. One of your machines attacked a satellite launch
  - Very important military mission
  - Concerned that you may have done it on purpose.
  - Cost the US Gov't $357M
  - 7.5 months ago
  - FBI is coming over in a few minutes

In a perfect world, you would

- Review enterprise network activity audit logs as first step
  - Single location for entire enterprises network logs
    - Query for any activity to NASA network or host
    - Pinpoints local hosts involved
    - Now begins the forensics examination
      - Was the attacking machine broken into?
      - If so, (hope so), where did it come from?
  - With multiple internal non-repudiation systems
    - You should be able to identify external / internal attack progression
      - Attack methodologies
      - Identify stepping-stone hosts

QoSient

Real world issues

Xerox machines intellectual property loss

News story reveals problems with Xerox machines

- Photocopy machines don't delete copy images
- Hospitals have lots and lots of Xerox machines

What can you do?

With single enterprise border non-repudiation system

- You would know if anyone from the outside ever discovered your Xerox machines in a scan
- You would know if anything directly accessed your Xerox machines from the outside

With non-repudiation system at the Xerox LAN border

- You would have logs of all network accesses to machine
- You would know what accesses extracted data rather than presented data to the printer
- You would have the content visibility needed to identify what images were extracted.

QoSient

Real world issues

Intrusion Detection Behavioral Anomalies

- Access from user X to supercomputer A account
  - Authenticated, acceptable
  - No apparent system log deviations
  - But came from a host outside the normal COI
- Human analyst noticed the network inconsistency
- User was on vacation
- First indication of significant US Gov't problem with Stakkato

QoSient

Real world issues

Unintended/Unexpected data exposure

- Symptom - Poor application performance
  - Database application exhibiting very poor performance
  - Each transaction taking 0.3-0.4 seconds to complete.
  - All software components running on a single machine
  - Absolutely no clues from debugging information
  - Wasn't this bad last week
  - Very, very, very sensitive medical information
- Network flow monitoring revealed problem
  - All IPC messaging was being transmitted onto the network
    - Data was being transmitted to the internal software process using network
    - Network turned it back around, after it left the LAN
  - One software component poorly configured
    - Using server's external name (NAT'ed environment)
  - Very, very, very, very bad

QoSient

Degradation of Service

- A primary design goal of Argus is DoS identification
  - Argus used in DDoS research papers (1996-2010)
  - CERT Advisory CA-1996-01 UDP Port Denial of Service
  - Many commercial DDoS products are flow data based
- Degradation is an attack on Quality of Service
  - QoS sensitive situational awareness is critical
    - QoS anomaly detection
    - QoS fault management
    - QoS intentional assignments
  - DoS protection really needs to be a part of QoS optimization
    - Can't discriminate QoS degradation when there is poor QoS
- Argus data specifically designed to support:
  - QoS Fault identification/discrimination/mitigation/recovery
    - Pre fault QoS Characterization and Optimization
    - Realtime fault detection and QoS anomaly characterization
    - Post fault recovery, forensics and impact assessments
    - Formal QoS optimization processes

QoSient

Security and Performance

- Security and performance are tightly coupled concepts
  - Network performance is an asset that needs protection
    - DoD GIG Information availability assurance (DoDD 8500.1)
    - Commercial product delivery dependent on network performance
    - Performance is being specifically attacked
  - Security and performance contribute directly to QoS
  - Security and performance are both optimizations
    - Many times at odds with each other
- Performance awareness data is security awareness data
  - Presence with identifying information is much of the forensics story
- Performance as a leading security indicator
  - Exfiltration and spam generation consume resources
  - Classic "man in the middle" and "traffic diversion" detection
    - Scenarios create measurable end-to-end performance impacts
  - [D]DoS detection is a performance anomaly problem

QoSient

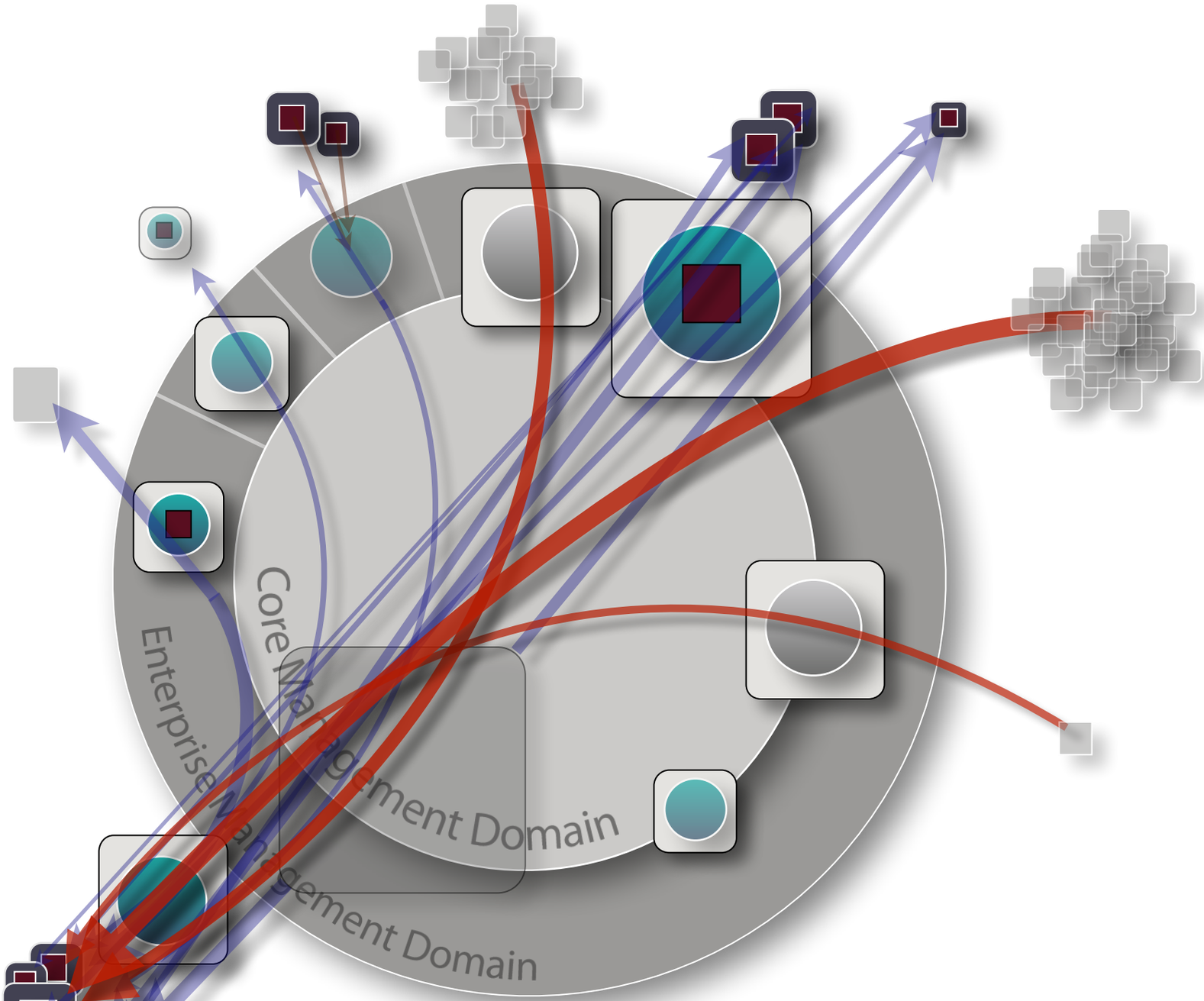
Degradation of Service (cont)

- QoS Fault Discrimination
  - Traditional QoS fault detection and mitigation
    - End-to-End oriented QoS tracking capability
      - Availability, demands, path, latency and efficiency modifications
      - Host vs Network QoS impact discrimination
      - Distributed sensor strategies provide best "finger pointing" capabilities
    - Historical audit provides baseline analytics for boundary tests
    - Discrimination can involve session dependency analysis
      - Front end network service dependancies
        - ARP, DNS, IP reachability, TCP availability, Service
      - Back end service dependency awareness
  - Discriminating intentional QoS failure
    - Protocol vulnerability exploitations
    - Exclusionary methods for attack designation
      - Flash crowd vs DDoS
      - Indirect attack assessment support

QoSient

Distributed Situational Awareness

Attack Scenarios - External Threats



White/Visible Node

Black/Non-Visible Node

Data Plane

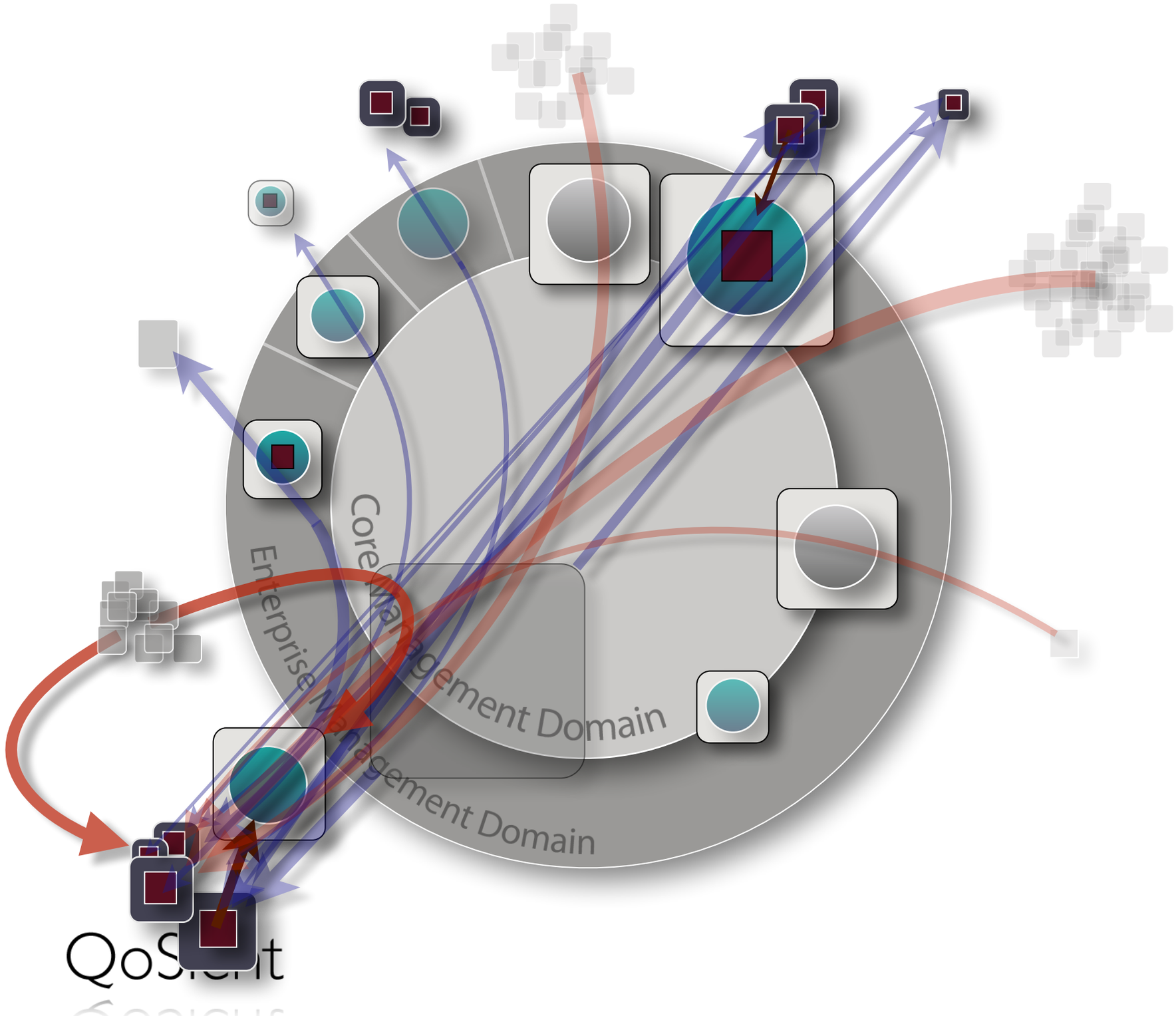
Command and Control
Attack Traffic

QoSient

ARGUS
Control
Identify
Analyze
Plan
Track

Distributed Situational Awareness

Attack Scenarios - Interior Exterior Spoofing



White/Visible Node

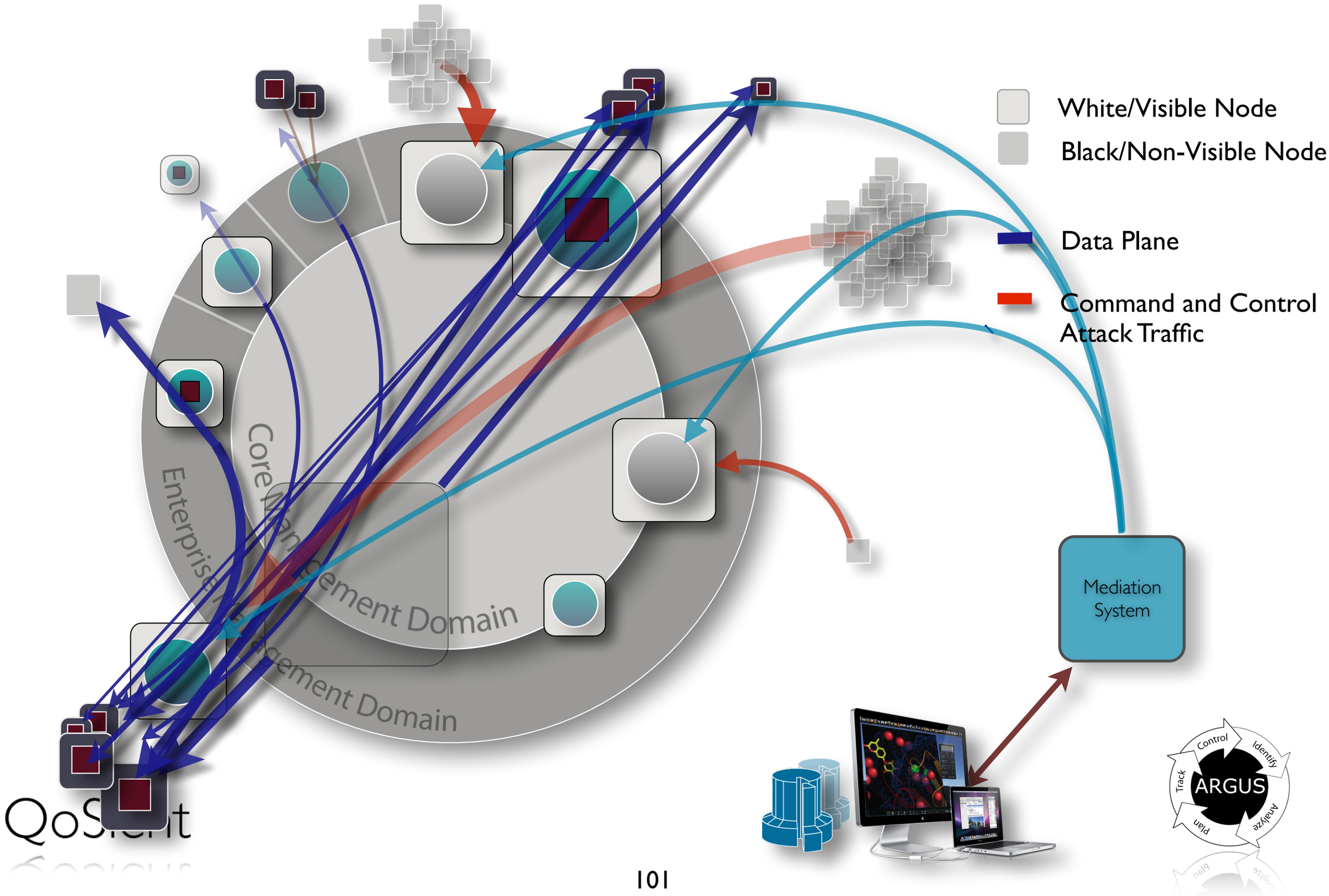
Black/Non-Visible Node

Data Plane

Command and Control
Attack Traffic

Distributed Situational Awareness

Attack Mediation



Degradation of Service (cont)

- Methods used to defeat [D]DoS mitigation
  - Mitigation involves denying access from list of exploit IP addresses
  - IP address spoofing
    - Host along attack path emulates [D]DoS traffic
      - Internal host that can "see" the target can forge 100,000's of simultaneous active connections to/from foreign hosts
    - Routing mediated address spoofing
      - BGP modifications allow near local networks to spoof address space
      - Internal modification to locally support foreign address space
        - Static routes can be setup so that "China" is routed to port 23b
        - Control plane attacks (ARP, RIP, OSPF) to advertise "China" is over here
- Result is that you just can't seem to shake the attack
- Distributed sensing detects this scenario
- Net-spatial data and active traceback strategies

Degradation of Service (cont)

- QoS Fault Mediation
  - Argus can provide information for effective mediation
  - Provide realtime forensics for threat analysis
    - Realize that QoS of critical assets are being affected
    - Provide real-time list of active nodes
    - For web attacks provide recurring URL visits
  - Provide CIDR addresses to block
    - Need to be sensitive to ACL limits of network equipment
    - Need to be cleaver when trying to block 50K IP addresses
  - Provide CIDR addresses to allow
    - Historical Community of Interest (COI) for allowable customers
    - The list of networks active at the initial time of attack
- Argus information to assure mediation worked
  - Network now performing within SLA
  - Track conditions to indicate when to revert, if ever

QoSient

Degradation of Service (cont)

- Methods used to defeat [D]DoS mitigation
  - Mitigation involves denying access from list of exploit IP addresses
  - IP address spoofing
    - Host along attack path emulates [D]DoS traffic
      - Internal host that can "see" the target can forge 100,000's of simultaneous active connections to/from foreign hosts
    - Routing mediated address spoofing
      - BGP modifications allow near local networks to spoof address space
      - Internal modification to locally support foreign address space
        - Static routes can be setup so that "China" is routed to port 23b
        - Control plane attacks (ARP, RIP, OSPF) to advertise "China" is over here
- Result is that you just can't seem to shake the attack
- Distributed sensing detects this scenario
- Net-spatial data and active traceback strategies

QoSient

Formal Non-Repudiation Systems

J-STD-025A

WAI/GT/FuncSpecs v1.0.1 (2000-06)

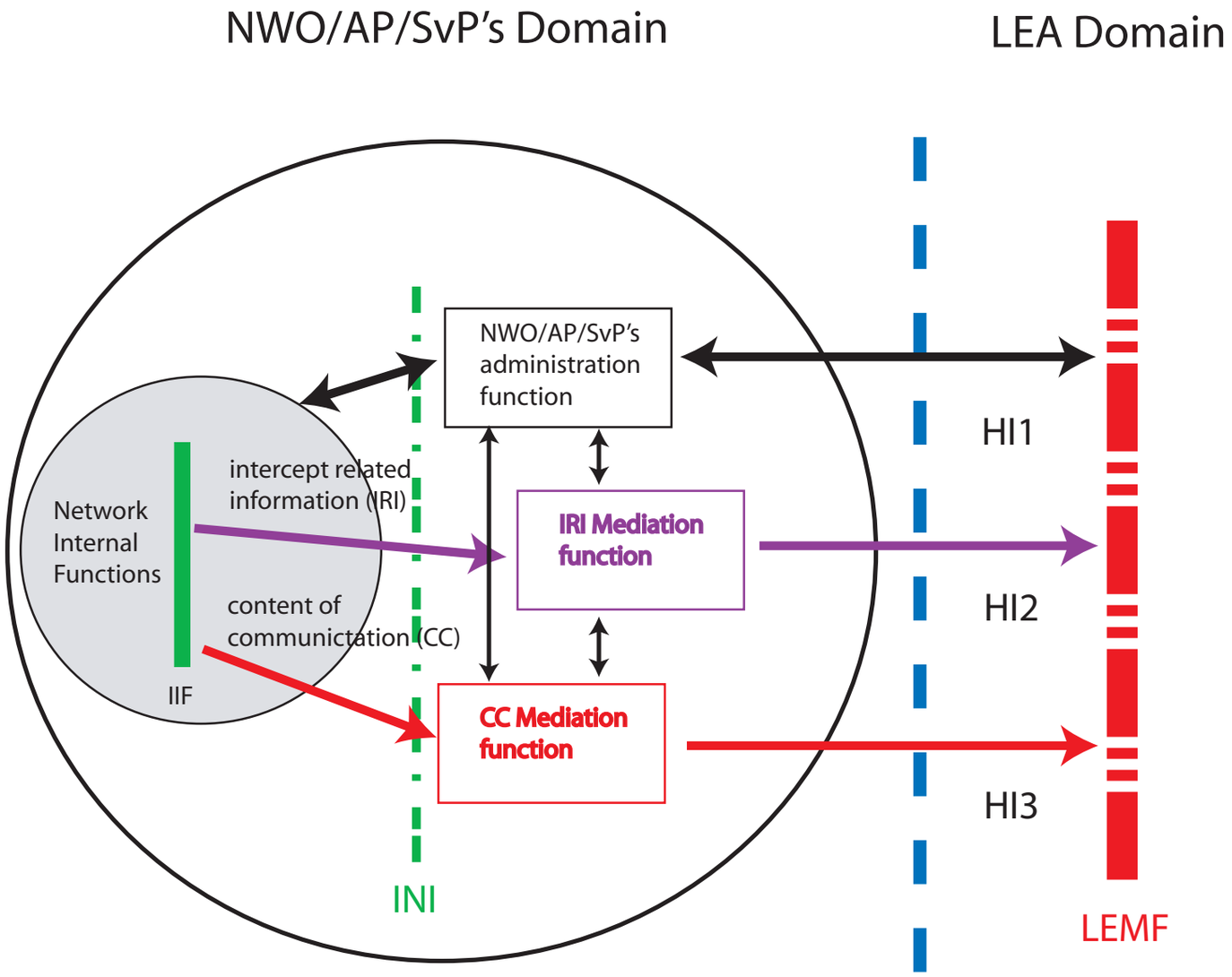
- Telephone Billing Records (retrospective)
- J-STD-025A / ETSI TS 101 671 (prospective)
  - Dialed Number Recorder (DNR/Pen Register)
  - Full Audio Interception (Title III/FISA)
- When concepts applied to data networks:
  - Content capture unencrypted (keys)
  - Information Protection Requirements
  - Geo-Location Information
  - Time Constraints
  - Unchanged State of Service

QoSient

ETSI ES 201 671

Telecommunications Security

Lawful Interception(LI); Handover interface for the lawful interception of telecommunications traffic



LI handover interface HI

IIF: internal interception function
INI: internal network interface

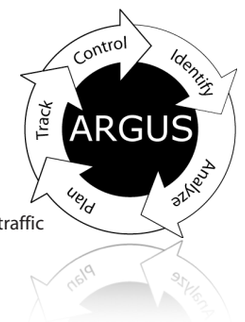
HI1: administrative infomration
HI2: intercept related information
HI3: content of communication

NOTE 1: Figure 1 shows only a reference configuration, with a logical reprsentation of the entities involved in lawful interception and does not mandate separate physical entities.

NOTE 2: The mediation functions may be transparent.

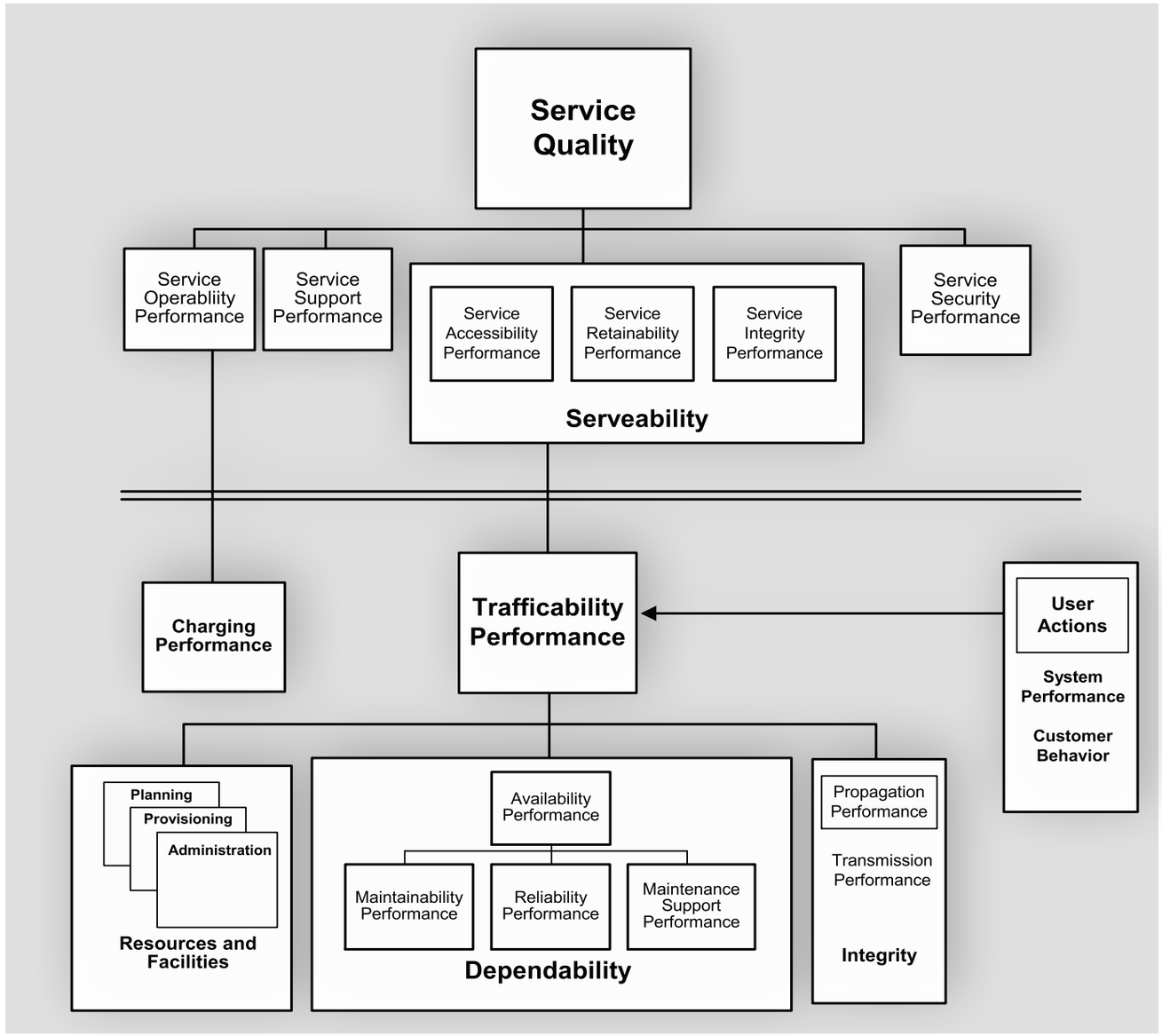
Functional Block Diagram Showing Handover Interface HI

QoSient



What Are CDRs Used For?

- Billing
- Traffic Engineering
- Network Management
- Maintenance
- Marketing
- Product Development
- Security
  - Fraud Detection
  - Forensics Analysis
  - Incident Response
  - Non-Repudiation / Audit



From ITU-T Recommendation E.800 Quality of Service, Network Management and Traffic Engineering

QoSient

Network Auditing

- Specified by DoD in NCSC-TG-005
  - The Red Book - Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria (1987)
- Goal to provide Non-Repudiation
  - Comprehensive audits accounting for all network use
  - Creates real deterrence in formal systems
    - Fear of getting caught is extremely powerful
  - Utility comes from the quality of collected information
- Internet network transaction auditing is emerging
  - Started at the CMU CERT-CC in early 1990's - Argus
  - Directly modeled after the PSTN CDR
  - Aspects of IP network auditing are being standardized

QoSient

Achieving Non-Repudiation

- Comprehensive Activity Accountability
  - Complete Activity Sensing and Reporting
  - Develop Information System with Formal Properties
    - Fundamental ground truth (if its not there, it didn't happen)
- Accurate and Efficient Activity Representation(s)
  - Stored data must represent actual activity
    - Attribute verifiability
      - Must be unambiguous with regard to object identification
      - Must have a relational algebraic correctness
    - Time synchronization and precision
      - Must convey correct order of events
- Fundamental Data Utility
  - Formal and Mature Data Model
  - Useful Data Availability Properties
  - Effective Storage and Retention Strategies

QoSient

Comprehensive Accountability

- Account for all network activity
  - Because any network activity can be associated with a cyber-security activity
    - Generally, if you aren't looking 'there', 'there' is where they will be
    - Hidden variables enable the adversary
  - Observation scope must be relevant
    - Utility of collected information should be very high
    - Using PSTN as guide, ISP can collect anything, but share nothing.

Argus approach to network non-repudiation

- Generate data to account for all network activity
  - Comprehensive Network Transactional Audit
  - Mechanism specified by DoD in NCSC-TG-005
    - The Red Book - Trusted Network Interpretation of the Trusted Computer System Evaluation Criteria (1987)
- Focus on all X.805 Security Planes
  - User, Control and Management network activity

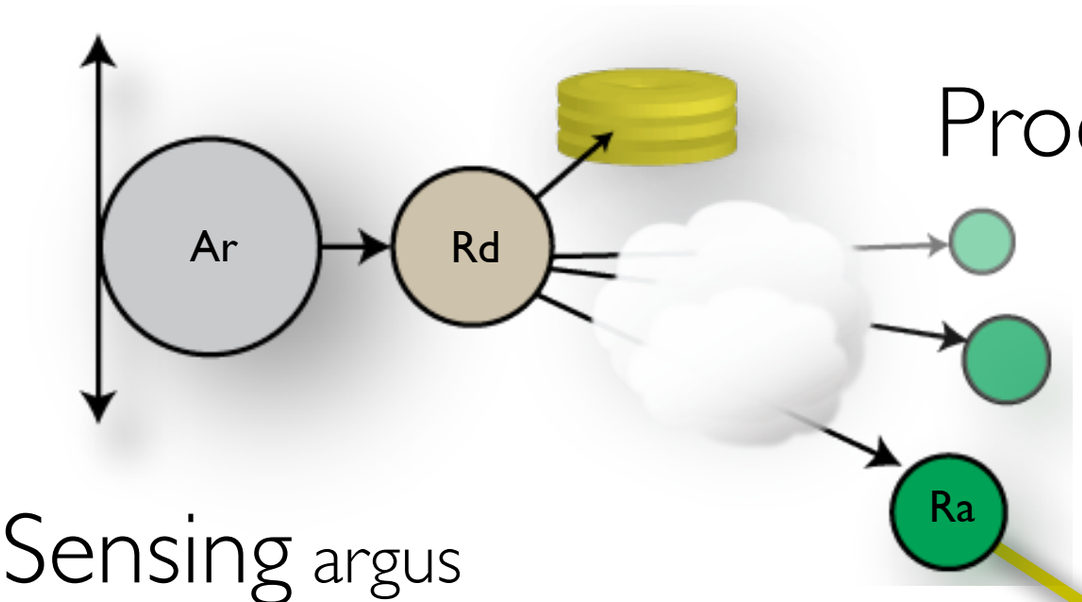
QoSient

Real-Time Argus

QoSient

Argus System Design

Distribution radium



Sensing argus

Processing ra*

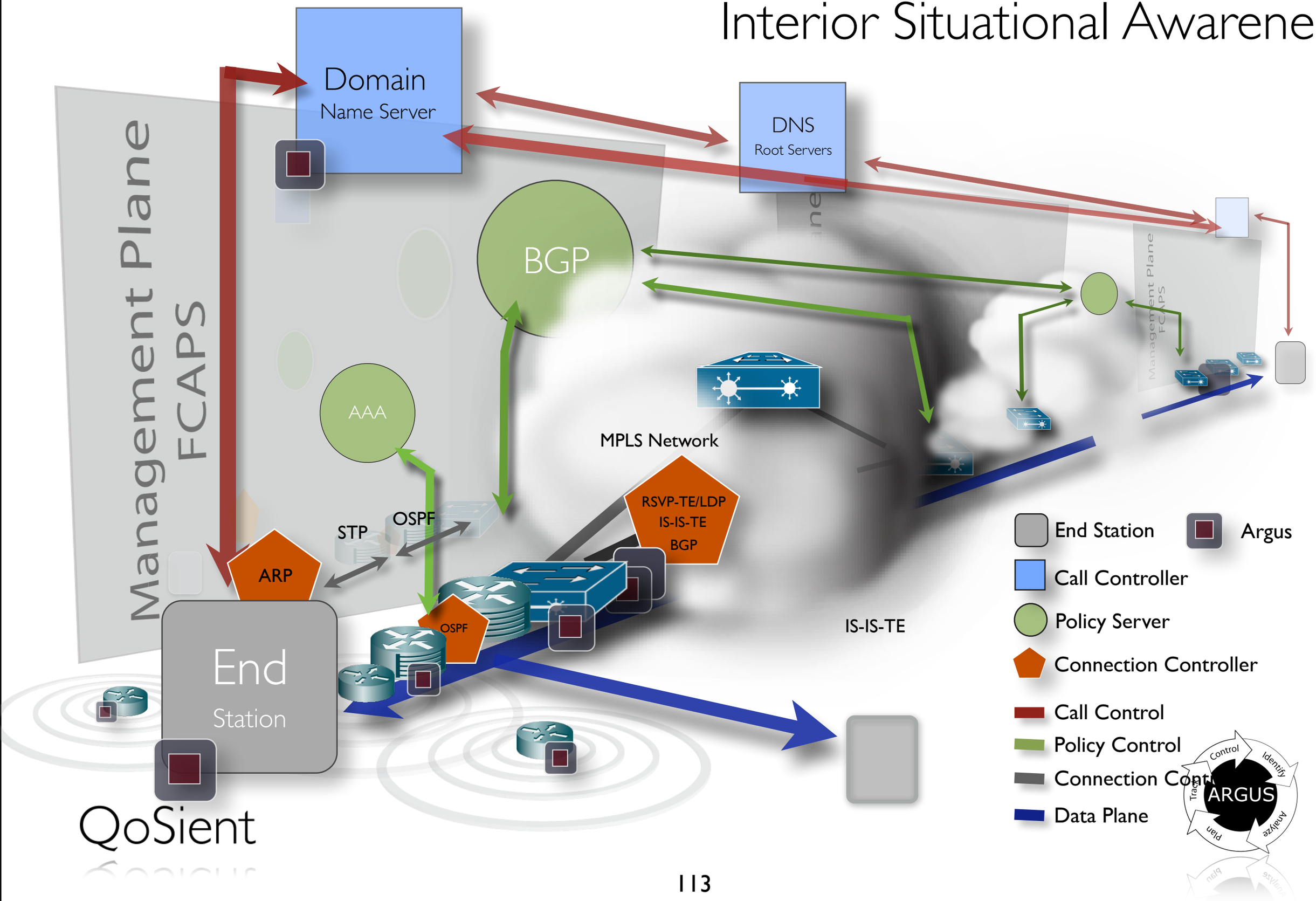
Archival

rasplit, rasql

QoSient

Comprehensive Enterprise Awareness

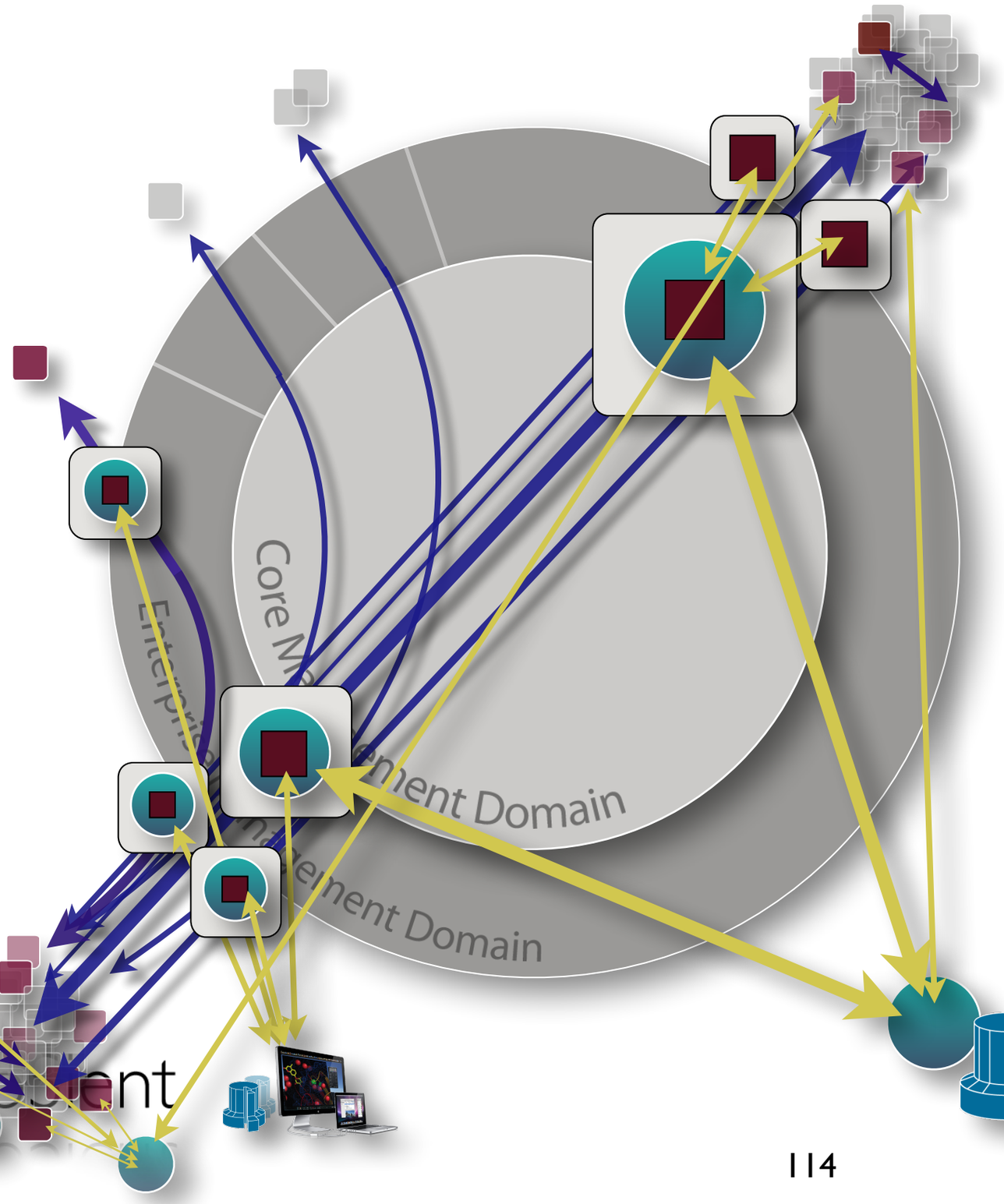
Interior Situational Awareness



QoSient

Complex Comprehensive Awareness

Local and Remote Strategies



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

Comprehensive Flow IS

Argus Sensor

Sensing
Argus Data Generation

QoSient

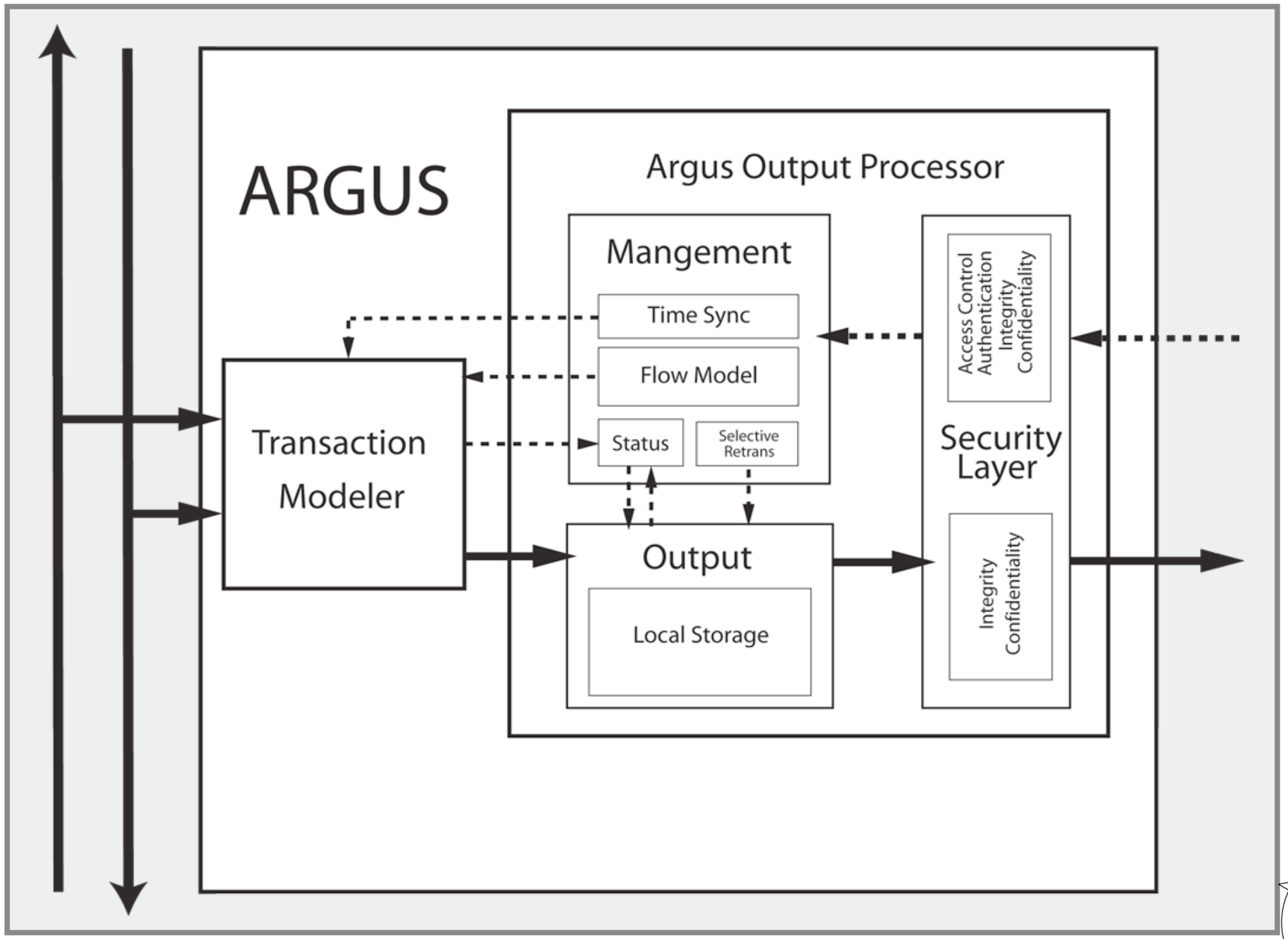
Argus Data Generation

- Packets to Flows
- Getting Started with Argus
- Argus Deployment
- Configuration
- Running Argus

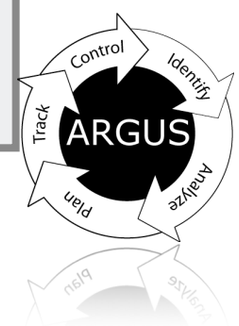
QoSient

Argus Sensor Design

Packets to Flows

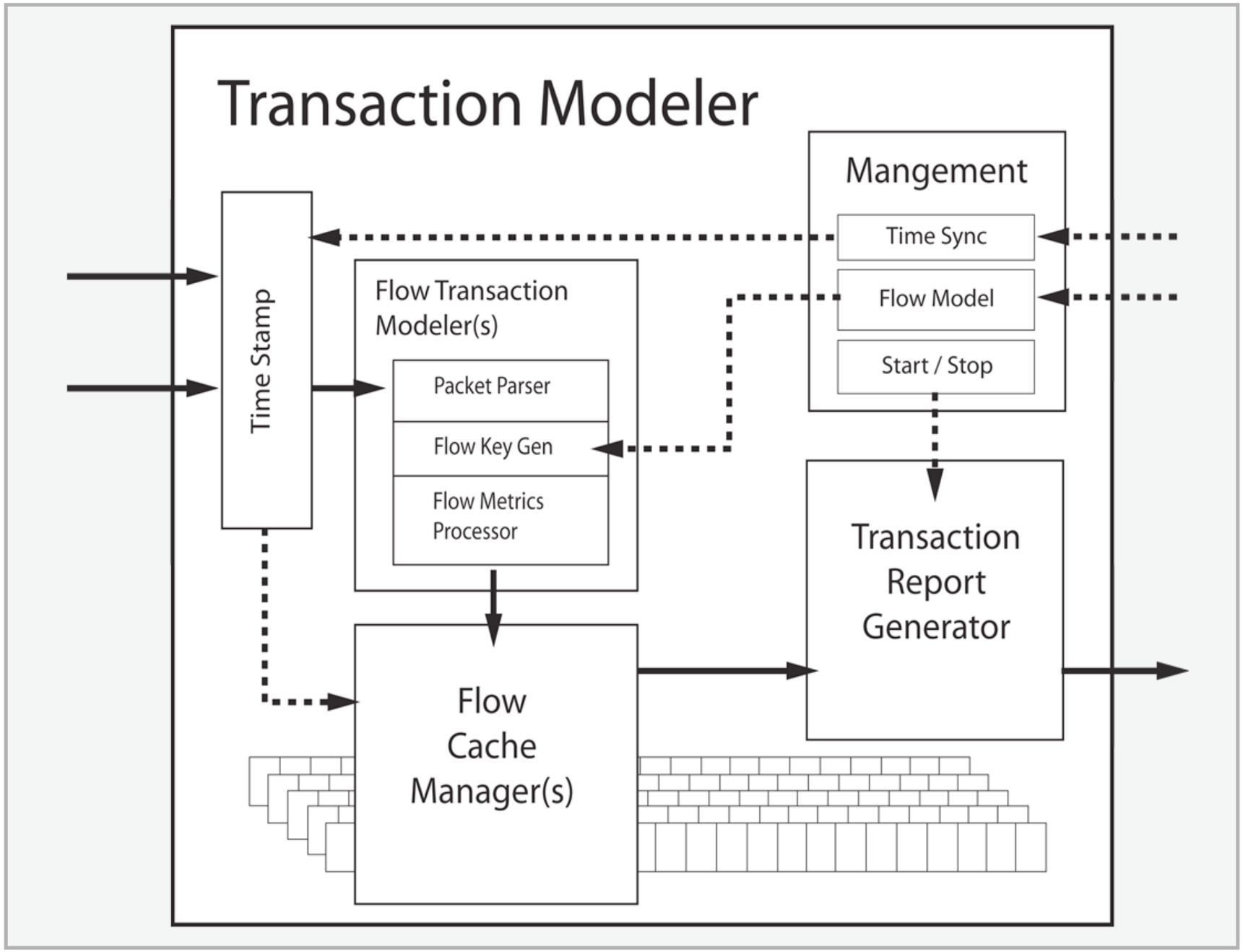


QoSient



Argus Sensor Design

Transactional Processor



QoSient

Network Flow Information

- All types contain IP addresses, network service identifiers, starting time, duration and some usage metrics, such as number of bytes transmitted.
- More advanced types are transactional, convey network status and treatment information, service identification, performance data, geo-spatial and net-spatial information, control plane information, and extended service content.

  - Available IP Flow Information
    - Argus
      - Control and Data Plane network forensics auditing
      - Archive, file, stream formats. (Binary, SQL, CSV, XML)
    - YAF/SiLK - CERT-CC (IP data only)
      - Designed for Cyber security forensics analysis
      - IETF IPFIX stream formats. Binary file format.
    - IPDR - Billing and Usage Accountability (IP data only)
      - ATIS, ANSI, CableLabs, SCTE, 3GPP, Java CP, ITU/NGN
      - File and stream formats (XML).
    - Netflow, JFlow, Sflow (IP data only)
      - Integrated network vendor flow information - statistical/sampled
      - Used primarily for router operations, network management

QoSient

Packets to Flows

- Packet Timestamping
  - Methodology, Time Synchronization and resolution
- Packet Header Parser
  - Multiple flow tracking strategies determines parser
  - Supports OSI, IEEE, IP and Infiniband packet formats
  - Innermost Layer 3 target header (service layer)
    - Complex encapsulation stacking
      - L2 -> L3 -> L2 -> L3 -> L4 -> L2 -> L3
      - Support protocol discovery
  - Limited by packet snap size
    - Argus supports complex packet capture support
    - Privacy issues
    - Control plane vs data plane parsing

QoSient

Packets to Flows

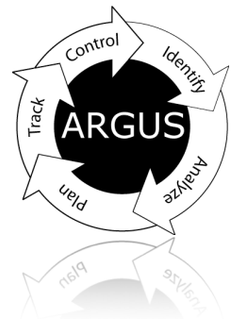
- Flow Key Generation
  - All packets are classified into a flow of some kind
  - Argus supports 14 fundamental flow types
    - Not protocols, flow types (P, P1-P2, Multicast/Unicast, etc....)
    - Bi-directional support for all flow types (when they exist)
    - Bi-direction flow keys for all supported encapsulations
  - Flow Key is "key" to all flow tracking
  - One packet one flow rule
    - Simplify flow machine call structure
    - Control plane is the bending of the rule
      - ICMP packet accounted for in ICMP flow
      - ICMP state mapped to flow identified in contents

QoSient

Packets to Flows

- Flow Metrics Processor
  - Metric and attribute generation
    - Some metrics can be derived from packet itself
      - Packet size, application demand, reachability
    - Others require state
      - connectivity, availability, RTT, rate, loss, jitter, size distribution
  - Flow attribute (re)assignments
    - Flow state machine tracking
    - Dynamic attribute tracking
- Flow Cache Manager
  - Controls reporting of flow status
  - Controls dynamic flow redefinitions/reassignments

QoSient



Getting Started

- http://qosient.com/argus
- 'Using Argus' and 'Getting Argus' Links
- Argus documentation
  - Man pages provided in distribution
  - HOW-TO and FAQ on the web site.
  - Argus developers mailing list
    - argus-info@lists.andrew.cmu.edu.
    - Most questions are answered here
    - Email carter@qosient.com

Getting Argus

- http://qosient.com/argus/downloads.htm
- Current stable version is argus-3.0.6
- Provided as tarball source package
- Ported to 27 platforms
  - Linux, xBSDs, Mac OS X, Windows, HPUX, Solaris, VxWorks, AIX, OpenWRT, Tilera
- Depends on:
  - libpcap - http://tcpdump.org/release
  - flex - http://flex.sourceforge.net
  - bison - http://www.gnu.org/software/bison

QoSient

Making Argus

- Simple installation
  - ./configure; make
- Complex environments
  - Read ./README and ./INSTALL
  - Cygwin/OpenWRT
- Support standard autoconf options
  - ./configure --help
  - Common variations
    - prefix=/your/destination/directory
    - SASL Support
    - Native compiler options

QoSient

Installing Argus

- Simple installation
  - make install
- ./INSTALL describes some complex examples
- /etc/argus.conf
- System startup configuration
  - Linux chkconfig.1 support
  - MacOS X /Library/LaunchDaemons support
- RPM support - ./lib/argus.spec

QoSient

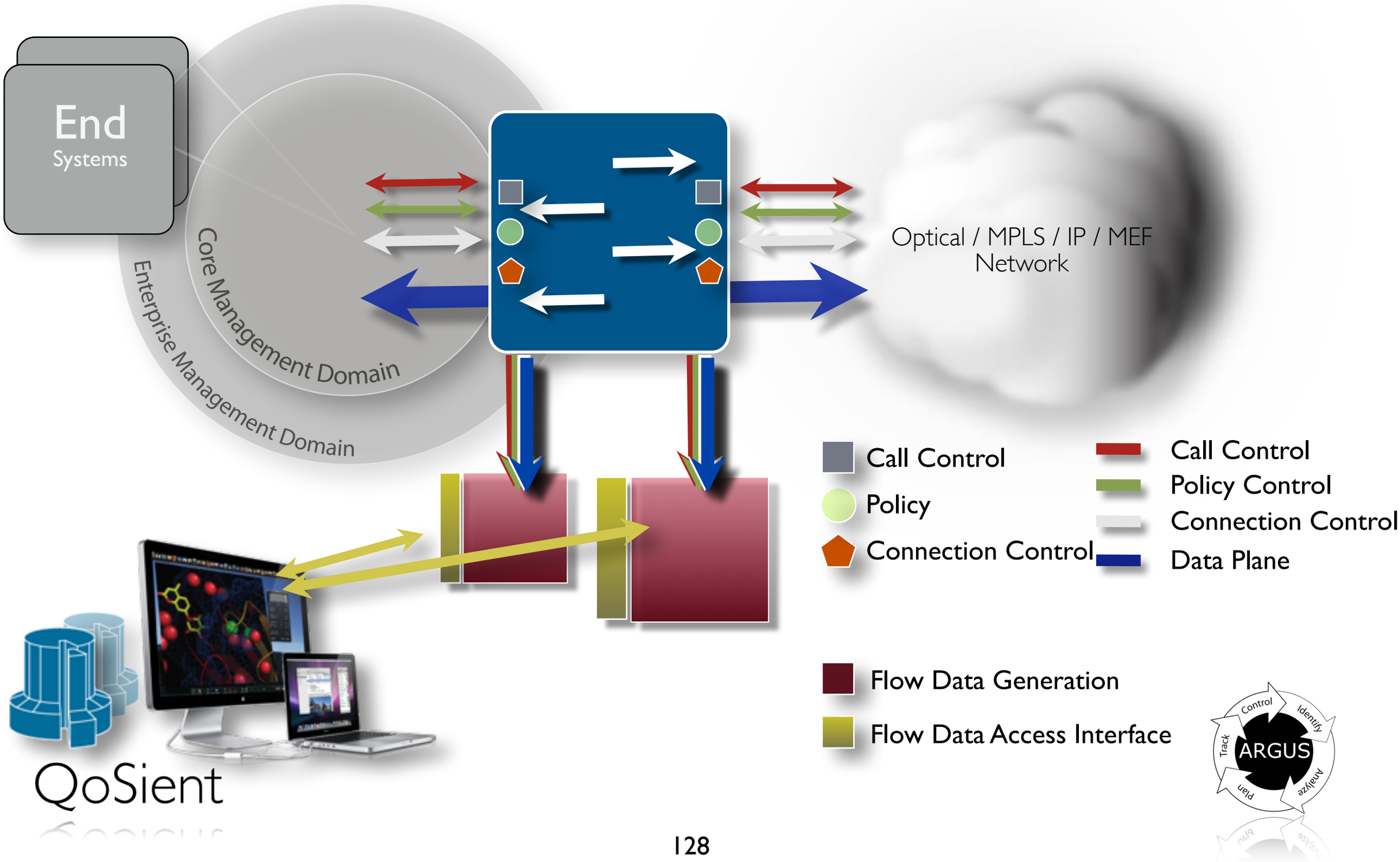
Deployment

- Monitoring Strategies
  - Enterprise Border Monitoring
  - Subnet Monitoring
  - End System Monitoring
  - Complex/Comprehensive Monitoring

QoSient

Enterprise Border Awareness

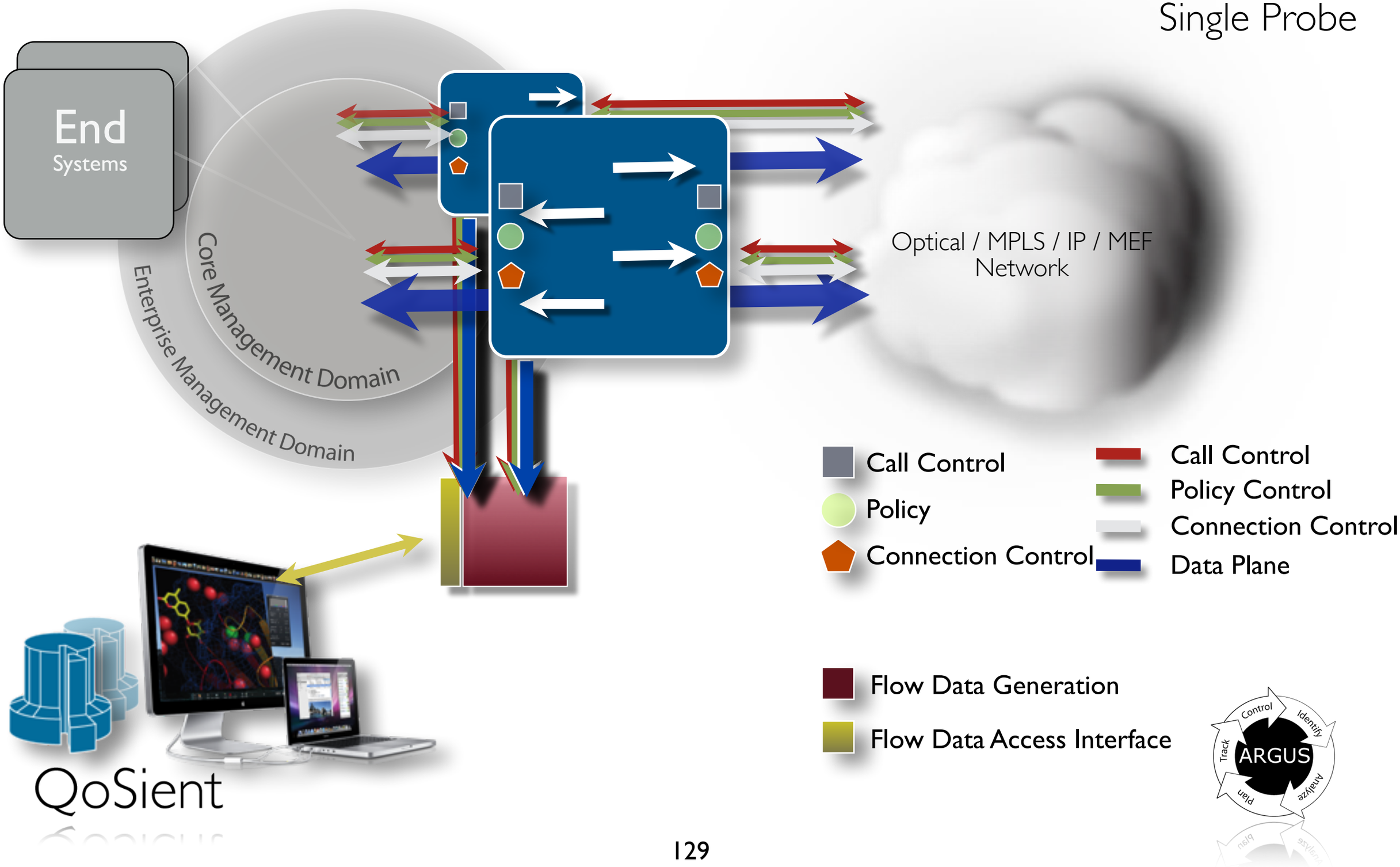
Internal/External Strategies



Enterprise Border Awareness

Asymmetric Routing Strategies

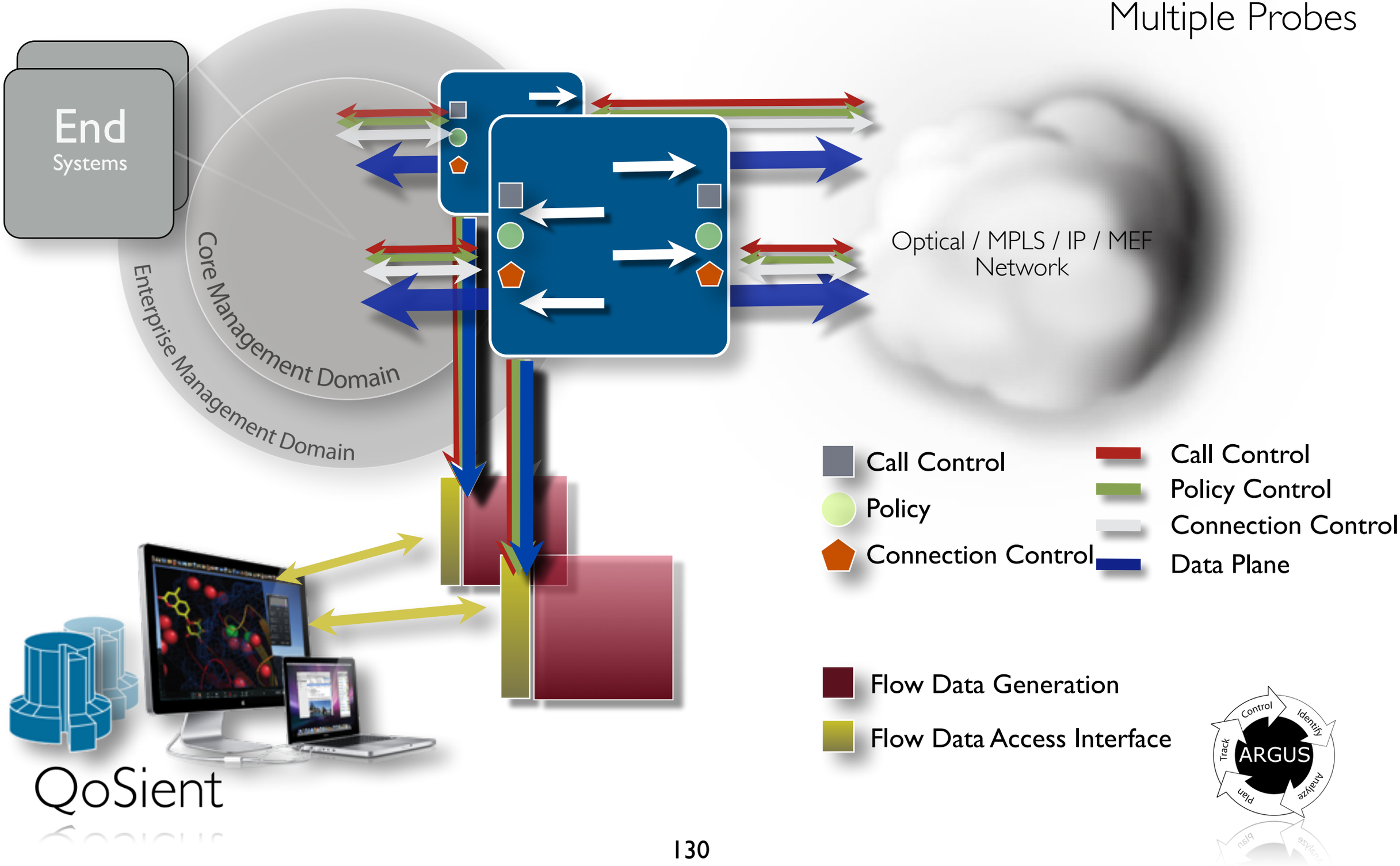
Single Probe



Enterprise Border Awareness

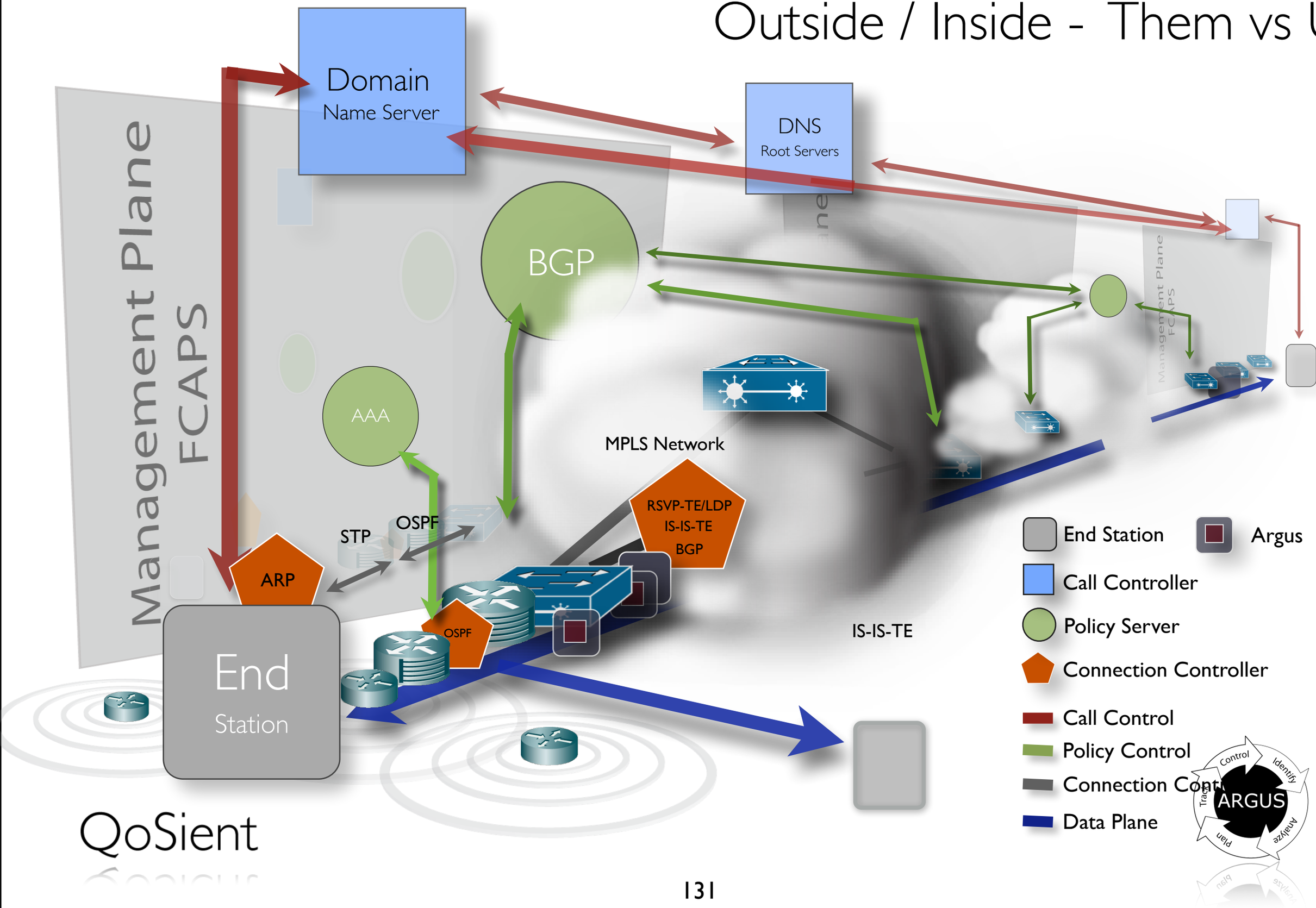
Asymmetric Routing Strategies

Multiple Probes



Comprehensive Enterprise Awareness

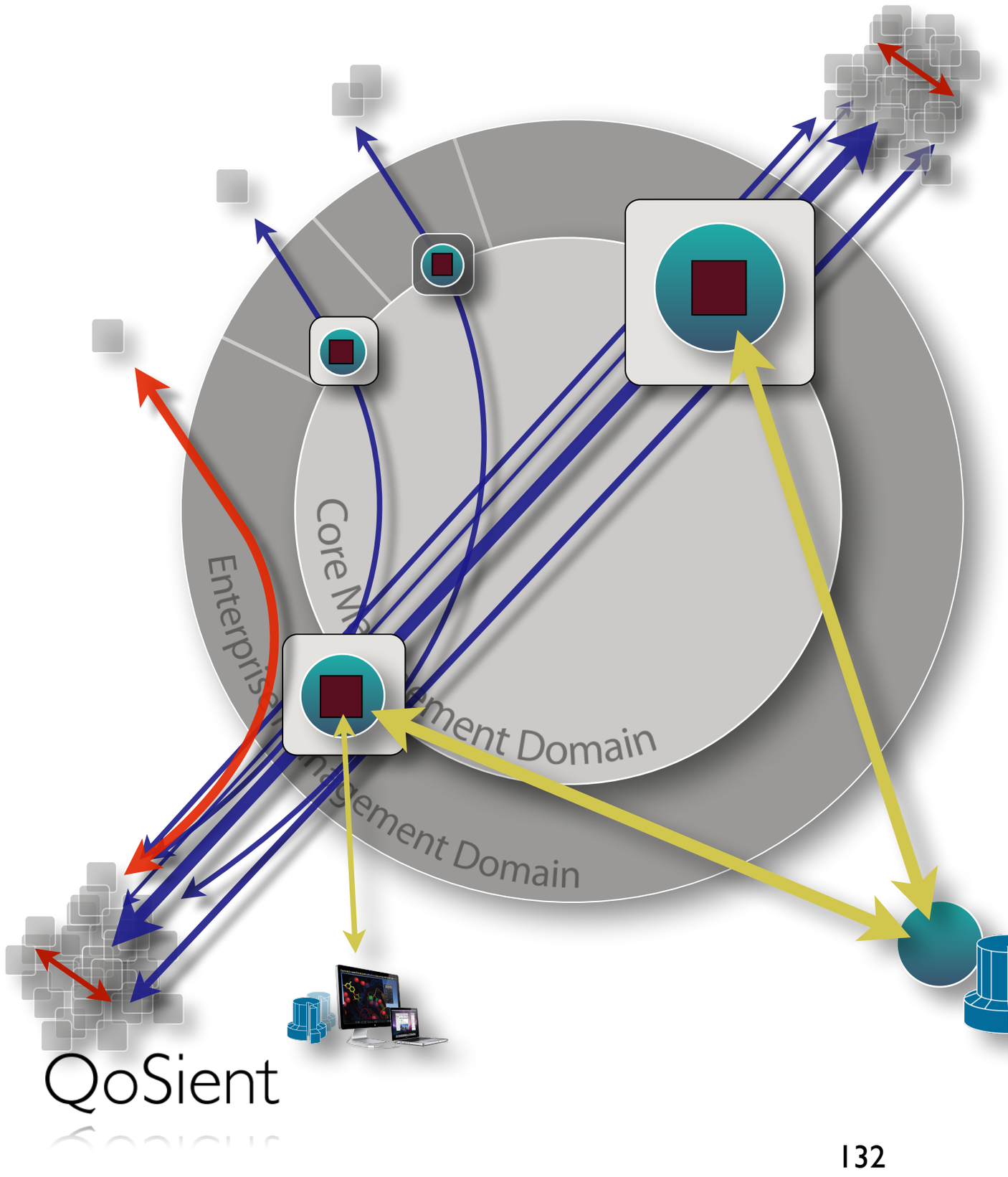
Outside / Inside - Them vs Us



QoSient

Enterprise Border Awareness

Outside Inside / Them vs Us



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

Unobservable Traffic

Comprehensive Flow IS

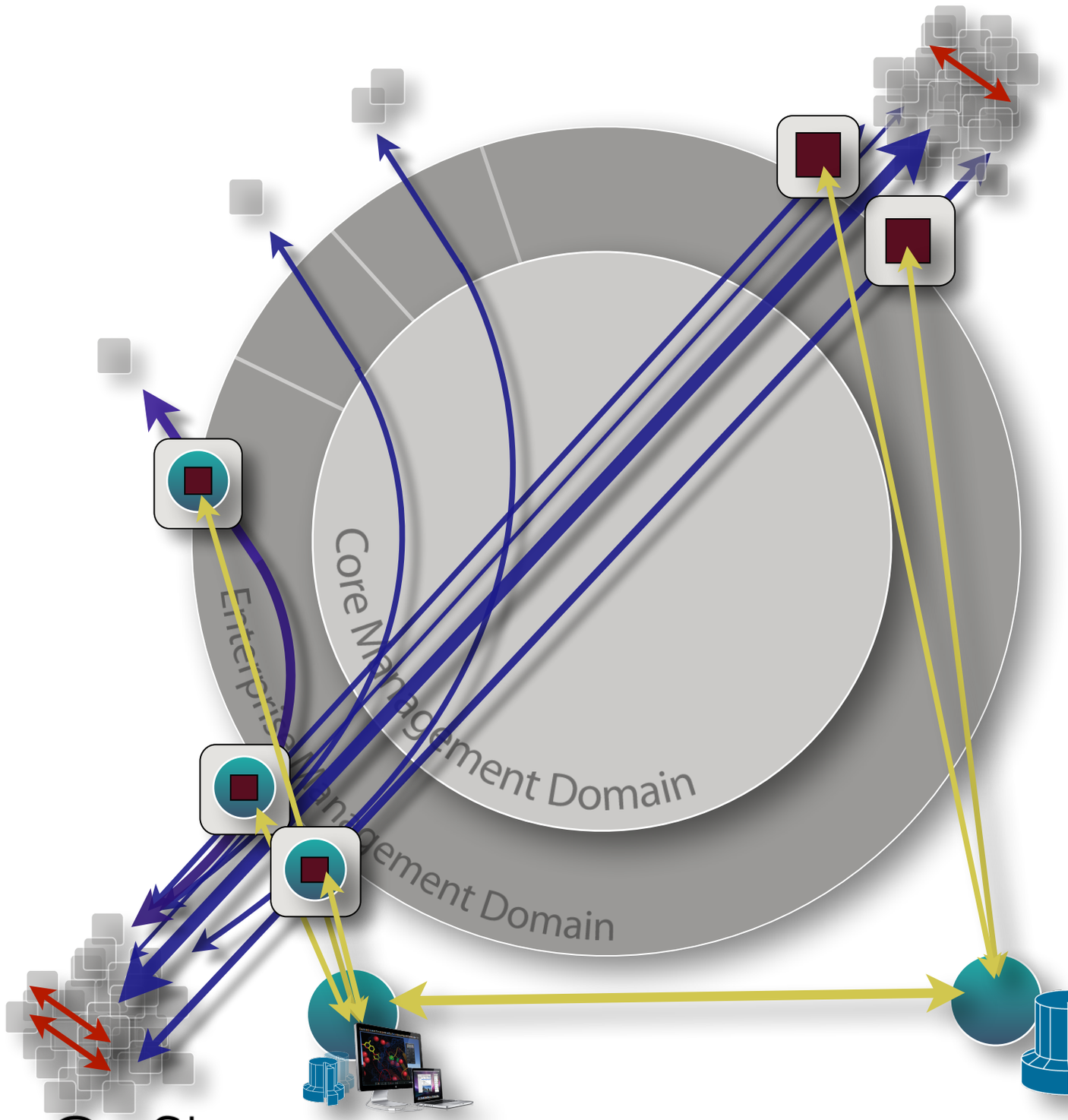
Argus Sensor

QoSient

ARGUS

Subnet Border Awareness

Local and Remote Strategies



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

Unobservable Traffic

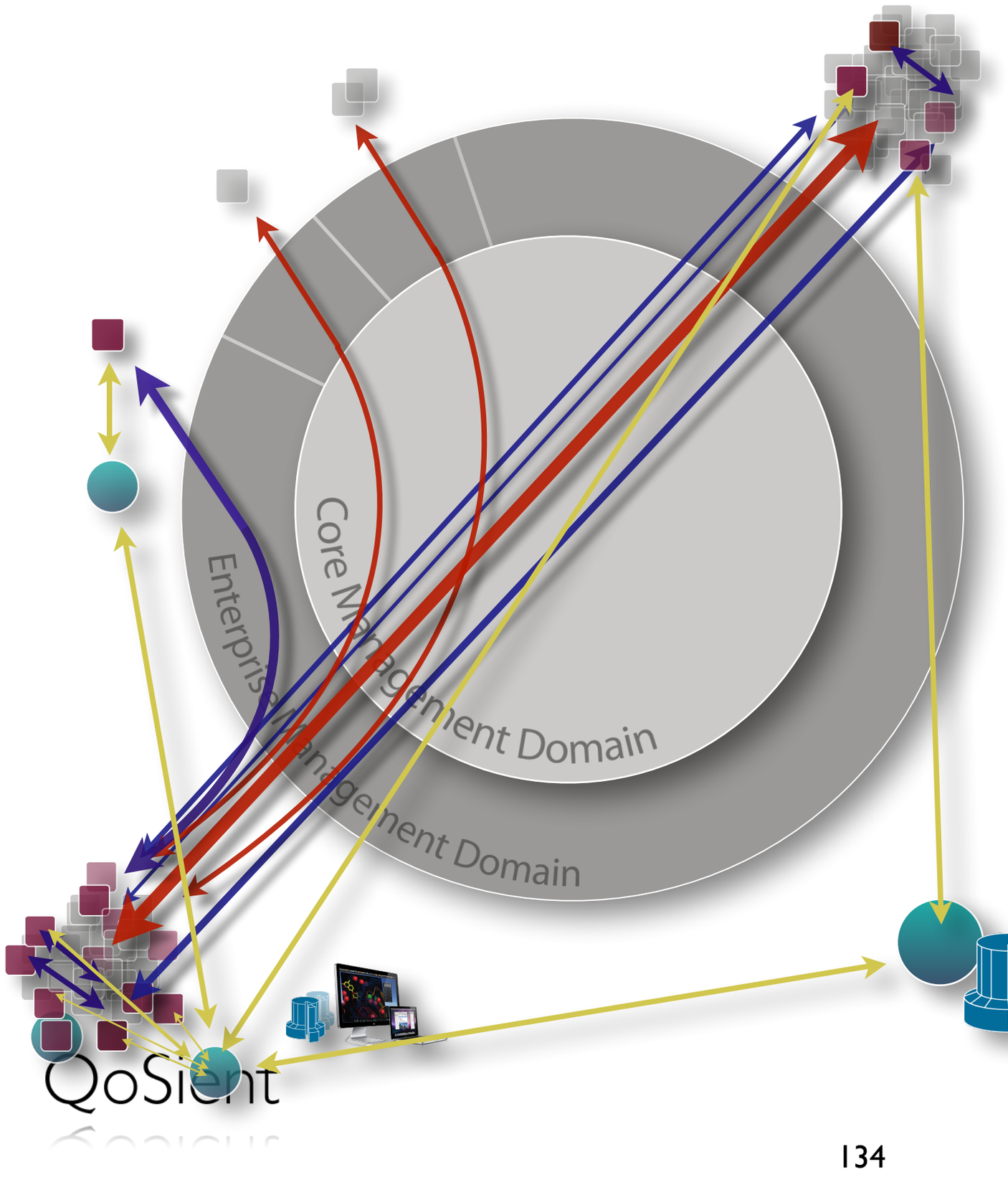
Comprehensive Flow IS

Argus Sensor

QoSient

End System Awareness

Local and Remote Strategies



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

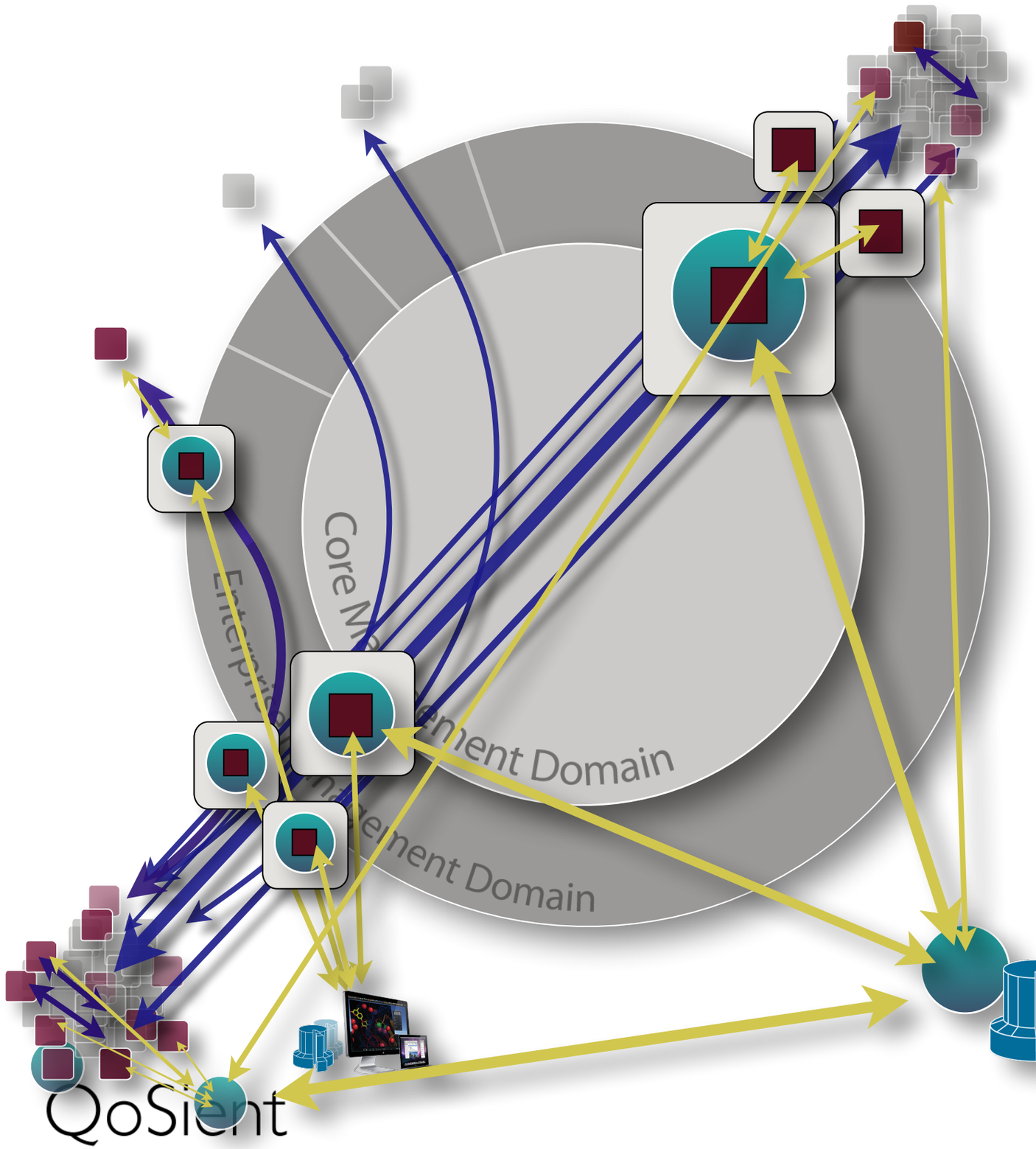
Unobservable Traffic

Comprehensive Flow IS

Argus Sensor

Complex Comprehensive Awareness

Local and Remote Strategies



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

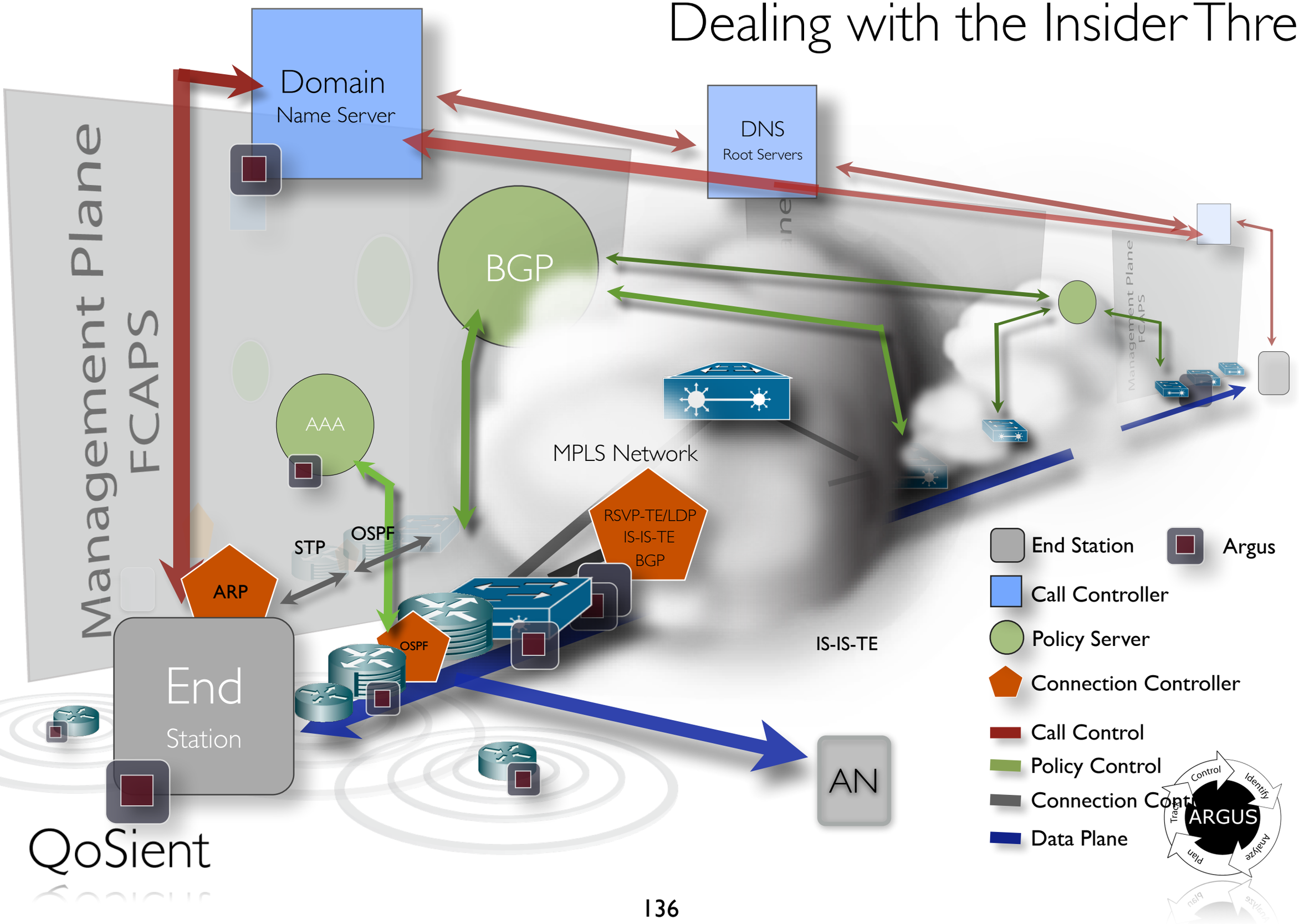
Unobservable Traffic

Comprehensive Flow IS

Argus Sensor

Comprehensive Enterprise Awareness

Dealing with the Insider Threat

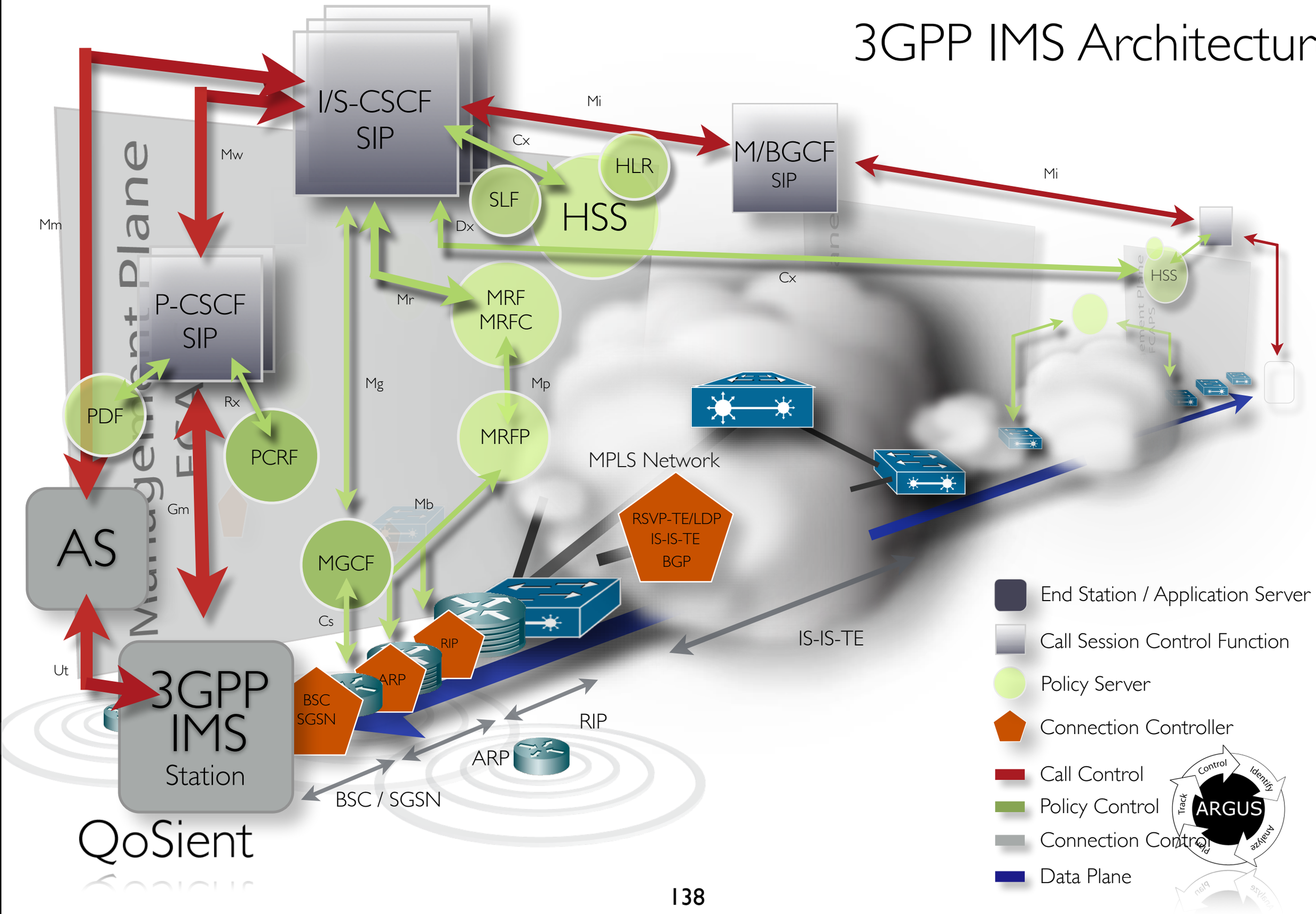


Complex Monitoring

- Critical elements
  - Time synchronization
  - Comparable flow key models
    - If collection system provides complex streaming analytics and aggregation
  - Observation Domain ID Allocations
    - Unique identifiers throughout the complete system
- Real-Time Operation
  - All sensors use same ARGUS_FLOW_STATUS_INTERVAL
  - All intermediate processing operates in the same time domain

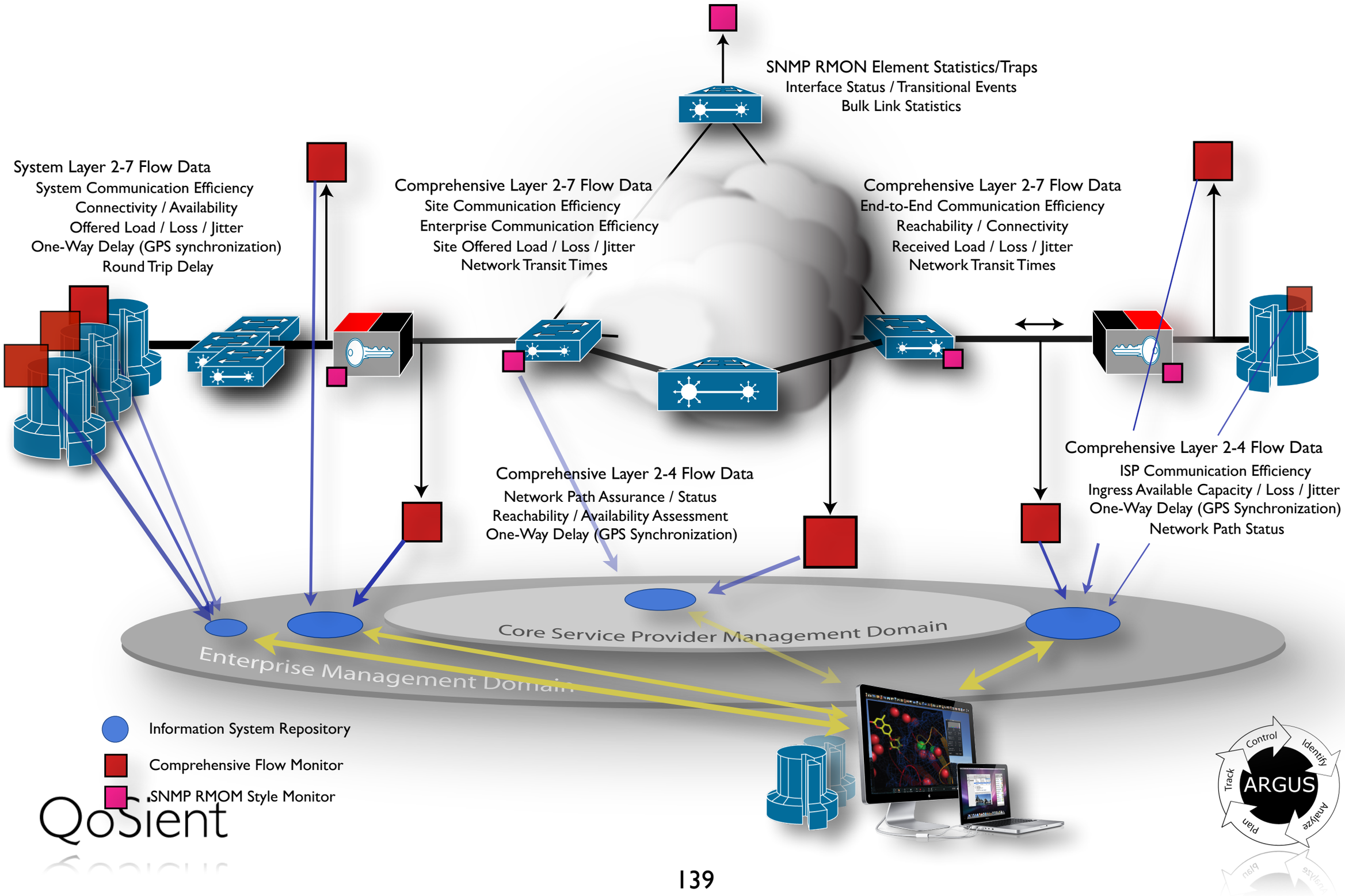
Mobile User Data Networks

3GPP IMS Architecture



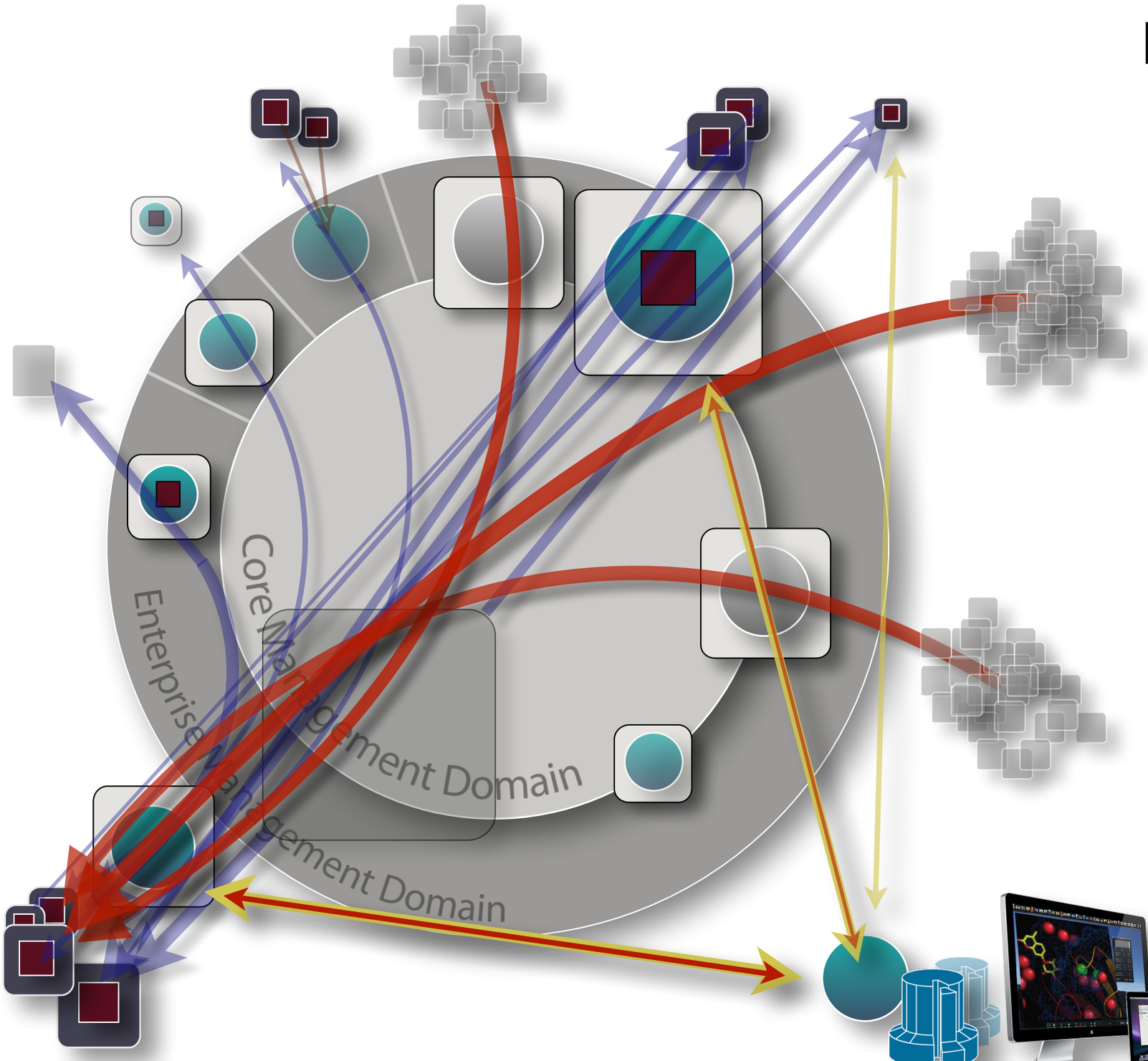
End-to-End Situational Awareness

Network Optimization - Black Core Mesh



Distributed Situational Awareness

Multi-Probe Multi-Site



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

Unobservable Traffic

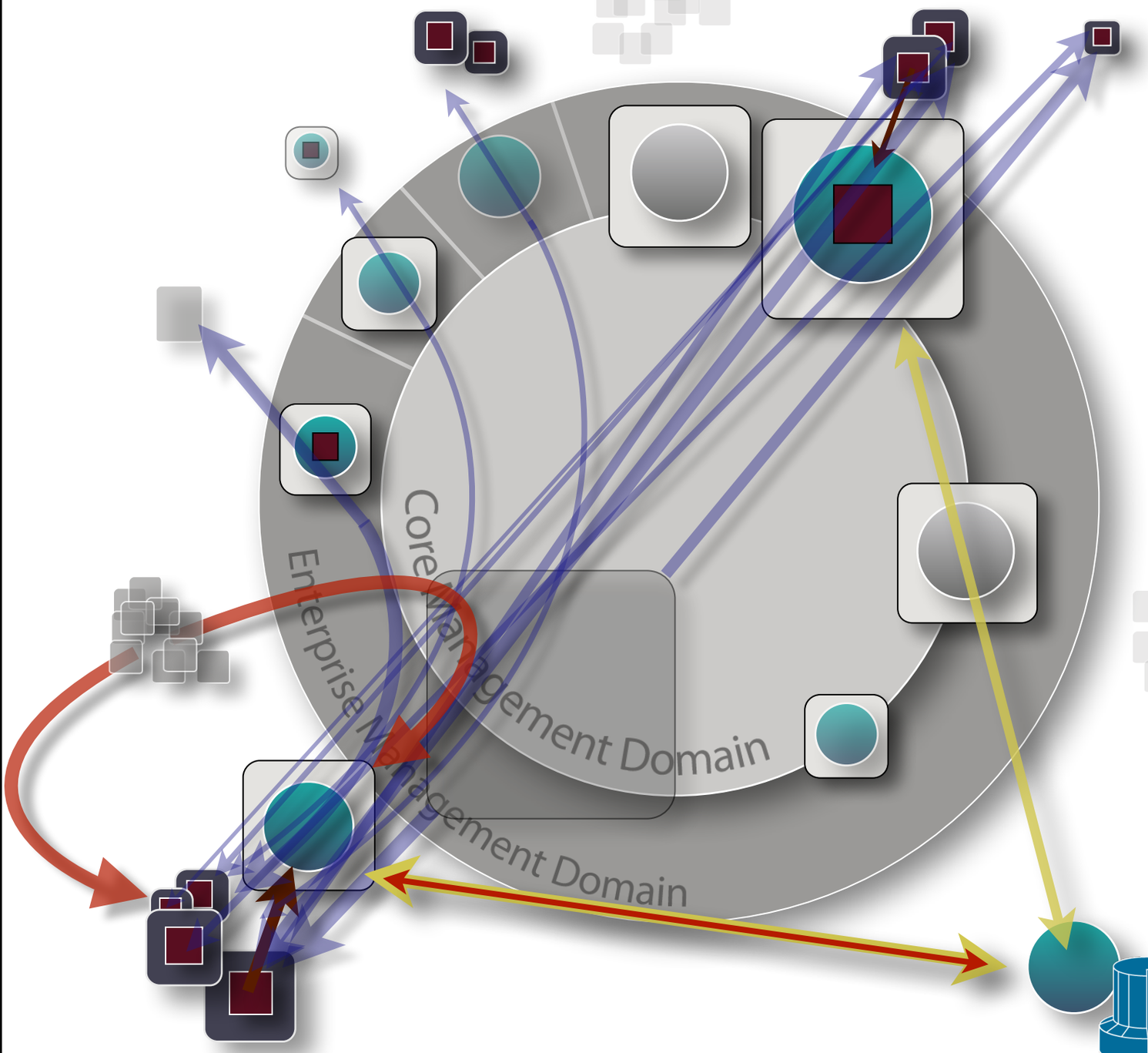
Comprehensive Flow IS

Argus Sensor

QoSient

Distributed Situational Awareness

IP Spoofing Scenarios



White/Visible Node

Black/Non-Visible Node

Data Plane

Situational Awareness Data

Unobservable Traffic

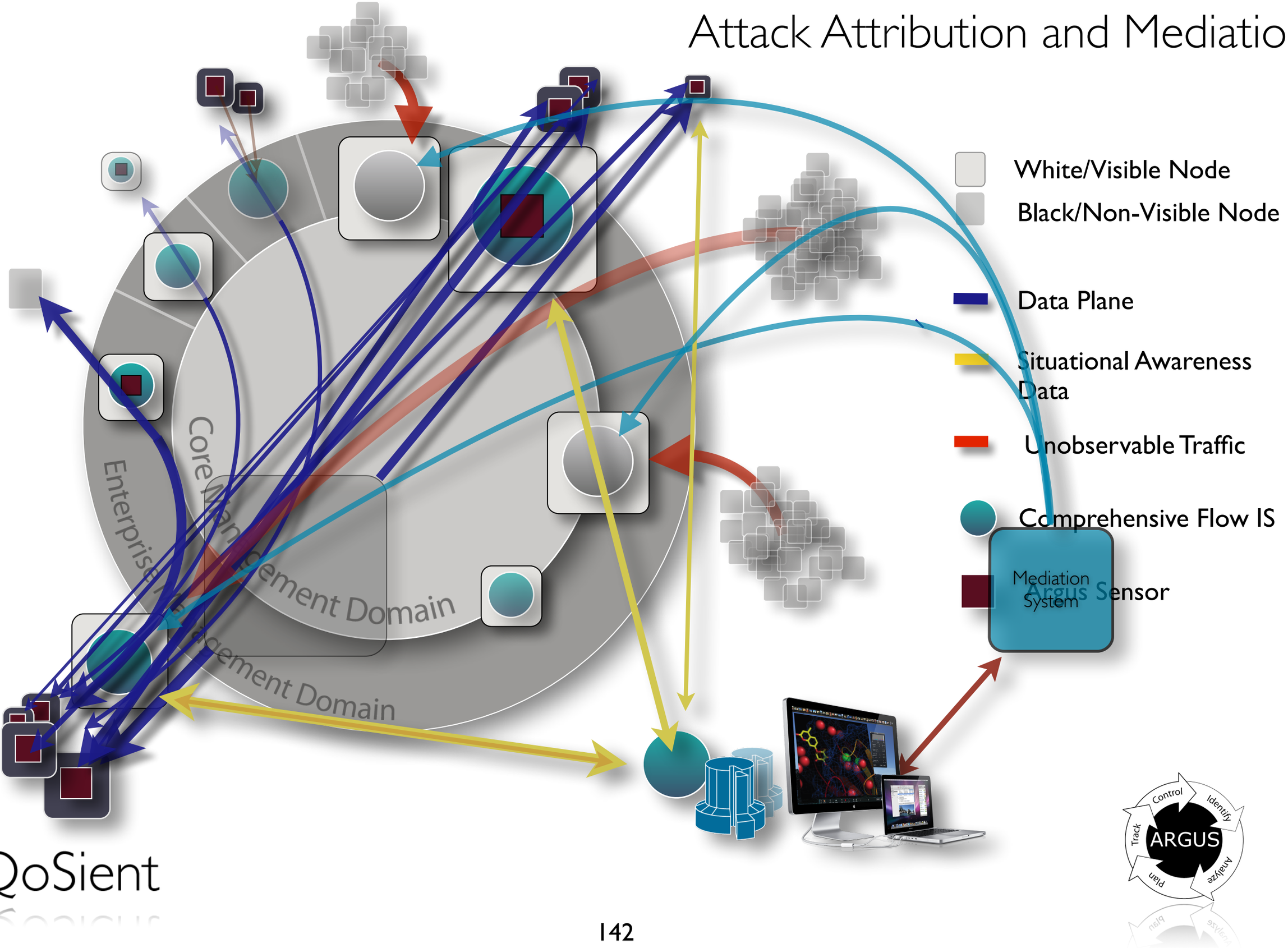
Comprehensive Flow IS

Argus Sensor

QoSient

Distributed Situational Awareness

Attack Attribution and Mediation



QoSient

Data Collection

QoSient

Data Collection

All ra* programs can read data from any Argus data source, files, stream, encrypted, and/or compressed, and can write current file structure.

Making a real-time argus based system needs just a little bit more.

- File Distribution
- Radium Distribution
- Argus Repository Establishment
  - cron
  - rasplit/rastream
  - rasqlinsert/rasql

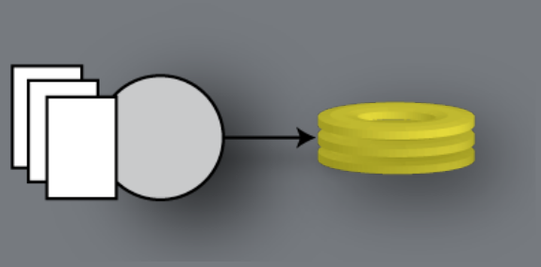
QoSient

Data Collection

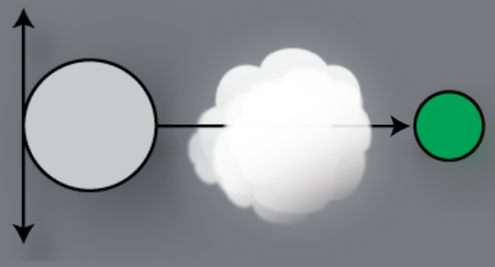
- Argus Data Distribution
  - Real Time Streaming Distribution
    - Data Flow Machine Architecture
      - Stream Processing Pipelines
    - Transport Protocols
      - Push and Pull Reliable and Unreliable Unicast
      - Push Multicast
  - File Based
- Argus Data Collection
  - Simple Collection Strategies
  - Complex Hierarchical Collection and Distribution

QoSient

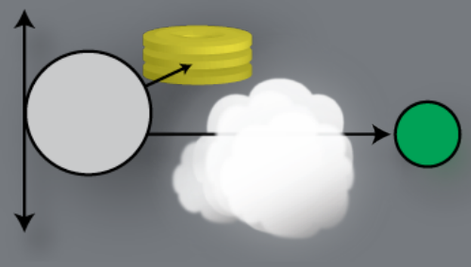
Data Collection



Argus reading from packet files or network and writing directly to disk

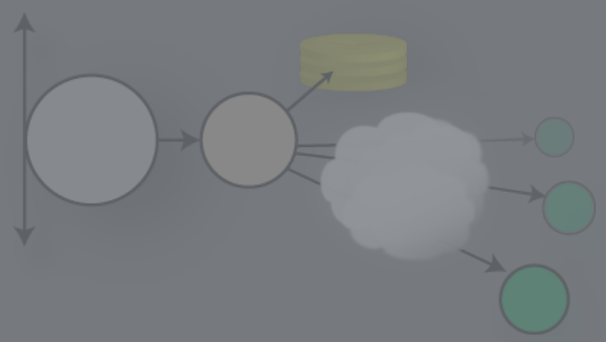


Argus reading from the network and writing directly to network based client

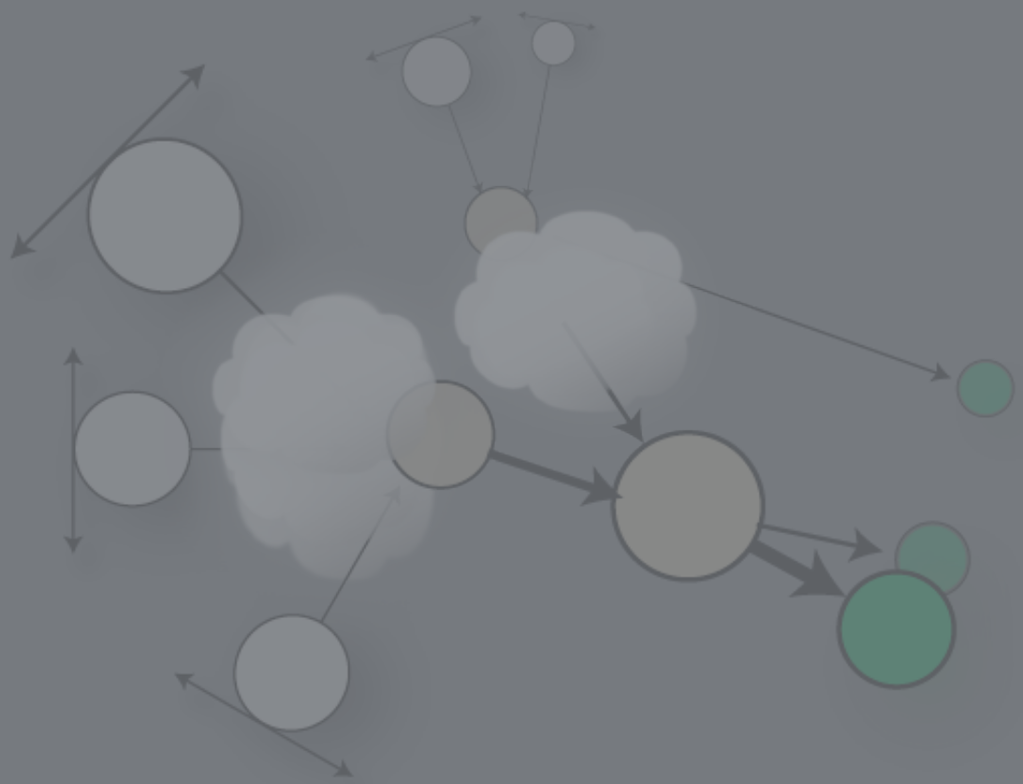


Argus reading from the network and writing directly to disk and network based client

Argus reading from the network and writing directly to a network Radium, writing to a client



Argus writing to local Radium which is writing directly to disk and to network based clients

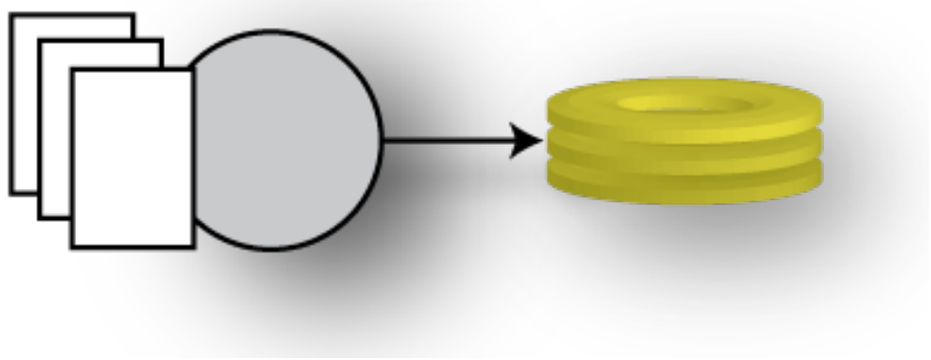


Many Argi writing directly to a Radium based distribution network, which is providing data to a set of clients.

QoSient

Data Collection

- Local Generation and Storage
  - Basis for argus-2.0 argusarchive.sh
  - Direct argus support for renaming files
  - Normally cron mediated
  - Issues with time and record spans
  - System designer has most control !!!

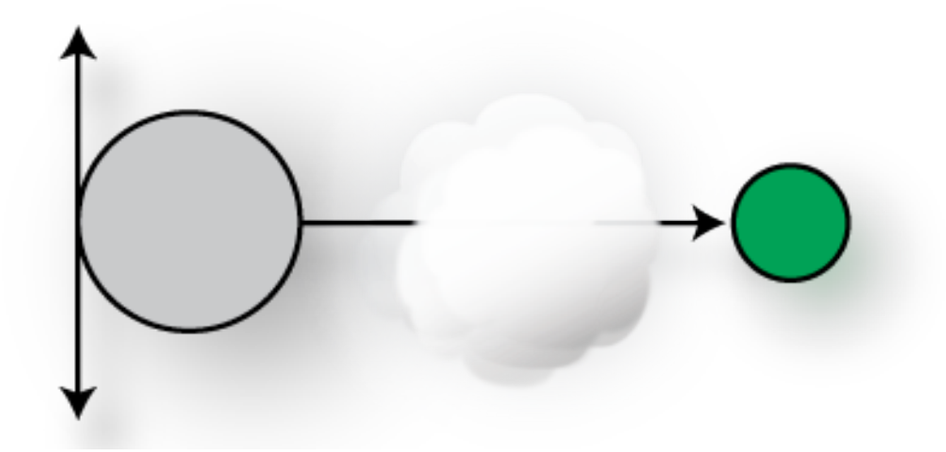


Argus reading from packet files or network and writing directly to disk

QoSient

Data Collection

- Local Generation Remote Collection
  - Most high performance systems use this strategy
    - Provides explicit scalability and performance capabilities
    - Relieves argus from physical device blocking
    - Network interfaces generally faster than local storage devices
  - Introduces network transport issues
    - Reliability, connection vs. connection-less, unicast vs multicast, congestion avoidance, access control and confidentiality

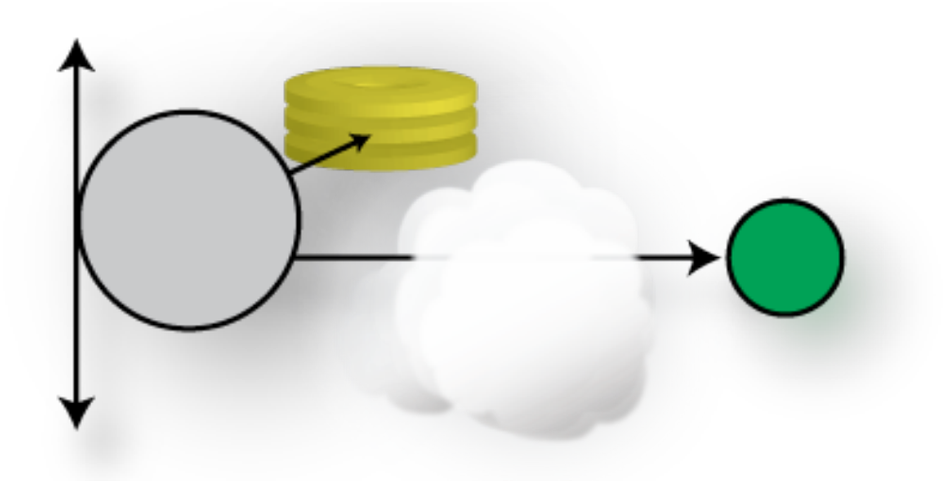


Argus reading from the network and writing directly to network based client

QoSient

Data Collection

- Local Storage and Remote Collection
  - Used when data reliability is most critical
    - Local storage provides explicit data recovery
    - File collection provides additional distribution flexibility
    - Scheduled transport
  - Reduces ultimate sensor performance
    - Argus itself is doing a lot of work
    - Packet processing is really the ultimate limit



Argus reading from the network and writing directly to disk and network based client

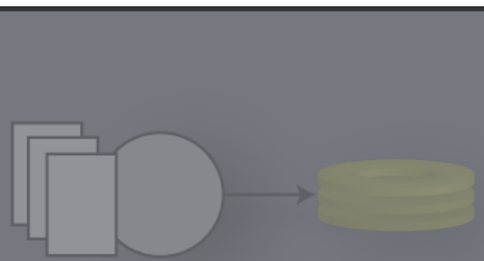
QoSient

Data Collection

Complex Collection Hierarchies

QoSient

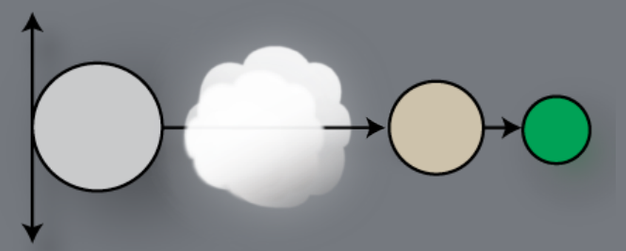
Data Collection



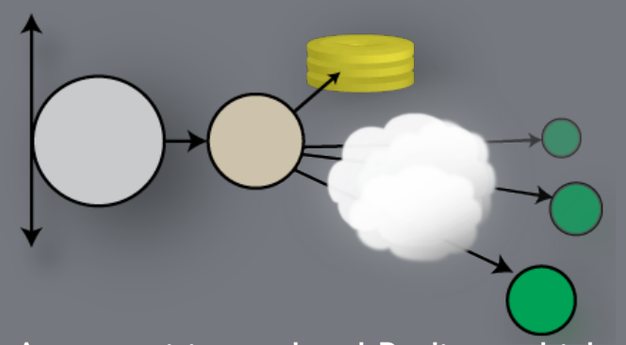
Argus reading from packet files or network and writing directly to disk

Argus reading from the network and writing directly to network based client

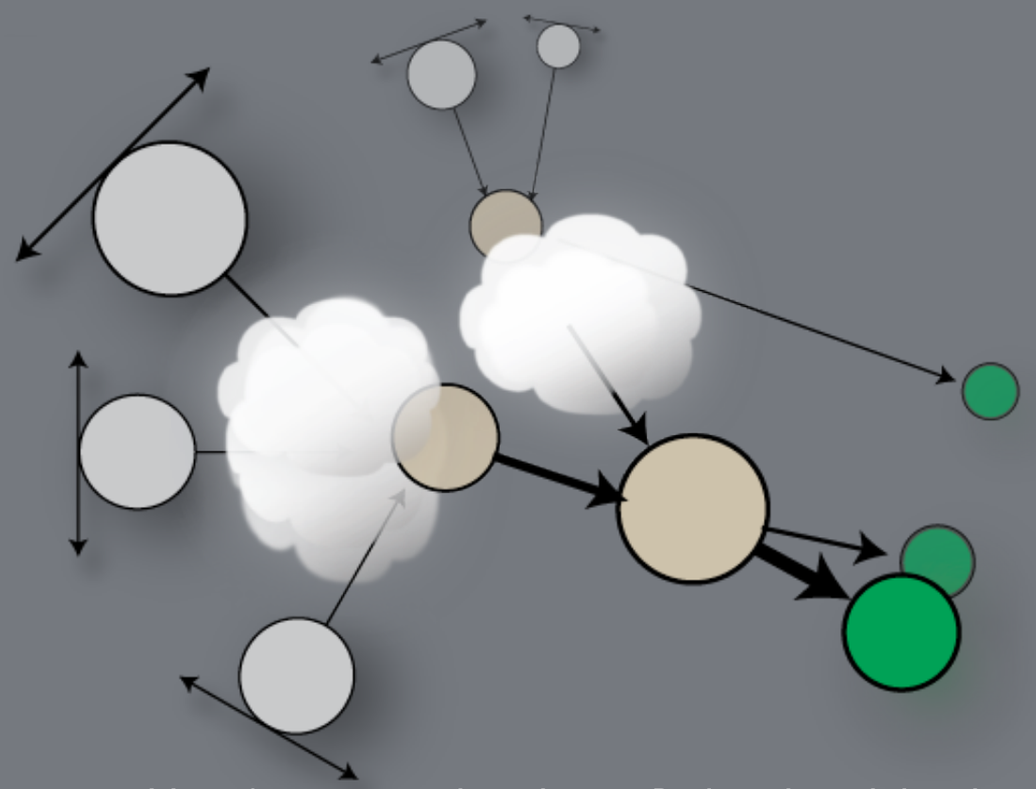
Argus reading from the network and writing directly to disk and network based client



Argus reading from the network and writing directly to a network Radium, writing to a client



Argus writing to local Radium which is writing directly to disk and to network based clients



Many Argi writing directly to a Radium based distribution network, which is providing data to a set of clients.

Data Collection

- Radium
  - Primary argus data distribution technology
  - Radium is a ra* program with an argus output processor.
    - Read from many sources
    - Write to many clients
    - Serve up argus data files
    - Process/transform data
    - Configuration is combo of argus() and ra()
- Supports very complex data flow machine architectures.

QoSient

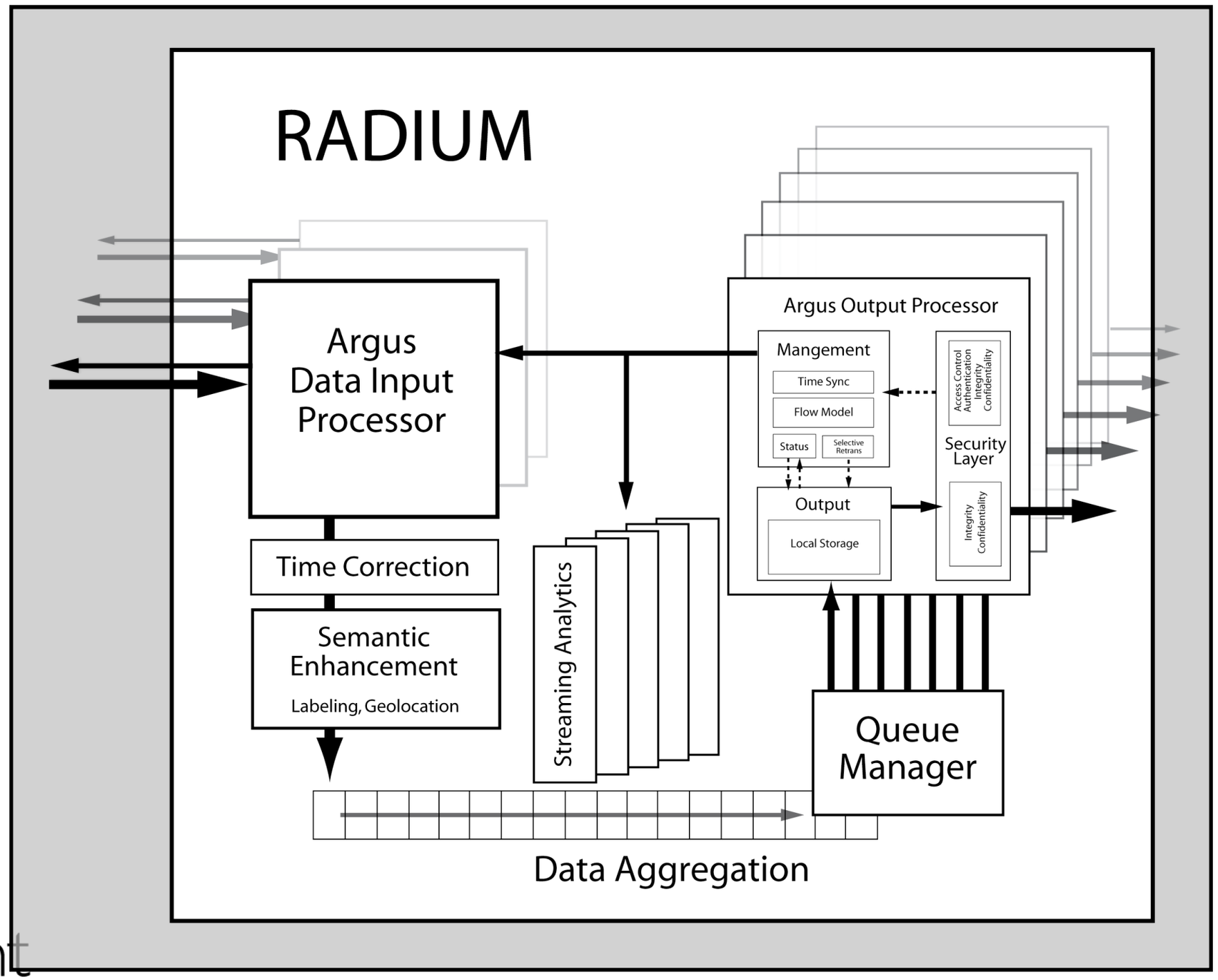
Radium

- Hybrid Argus and Argus client
  - Argus Client
    - Read argus data from all supported files and streams
    - Can read Netflow, Sflow, Jflow and FlowTools data
    - Reads up to 256 argus data sources, generates 1 output
    - Supports most ra* functions by design:
      - Filtering
      - Labeling - full ralabel.1 functionality
      - Flow Correction - time sync correction, direction
      - Aggregation - rabins() behavior
      - Stream Analytics - future work
  - Argus
    - Supports 256 argus data output processors
      - One radium, one output stream x256
      - Independent processors, independent outputs
        - Different transports, filters, sockets, files, etc.....

QoSient

Argus Collection Design

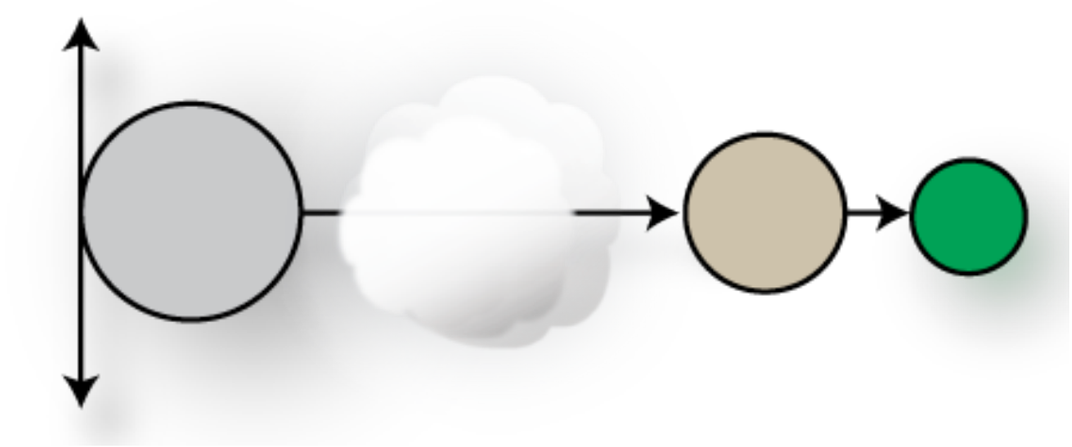
Radium Process



QoSient

Data Collection

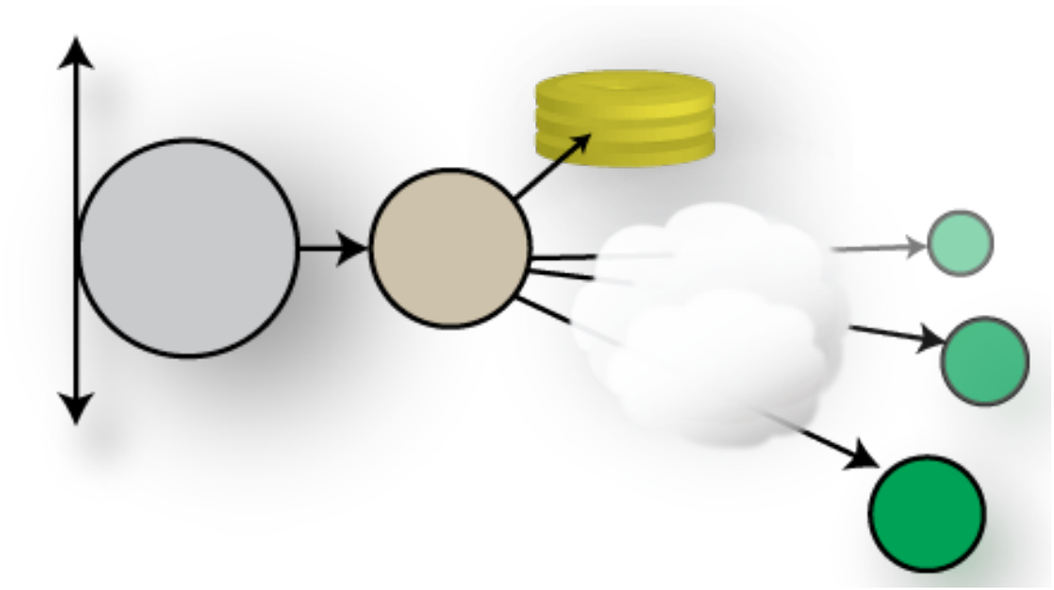
- Local Generation Remote Distribution
  - Most prevalent strategy used in argus-3.0
    - Provides explicit scalability and performance capabilities
    - Provides most stable collection architecture from client perspective
    - Single point of attachment for complete enterprise
  - Least reliable of 'advanced' strategies
    - Radium failure interrupts continuous stream collection, with no opportunity for recovery



Argus reading from the network and writing directly to a network Radium, writing to a client

Data Collection

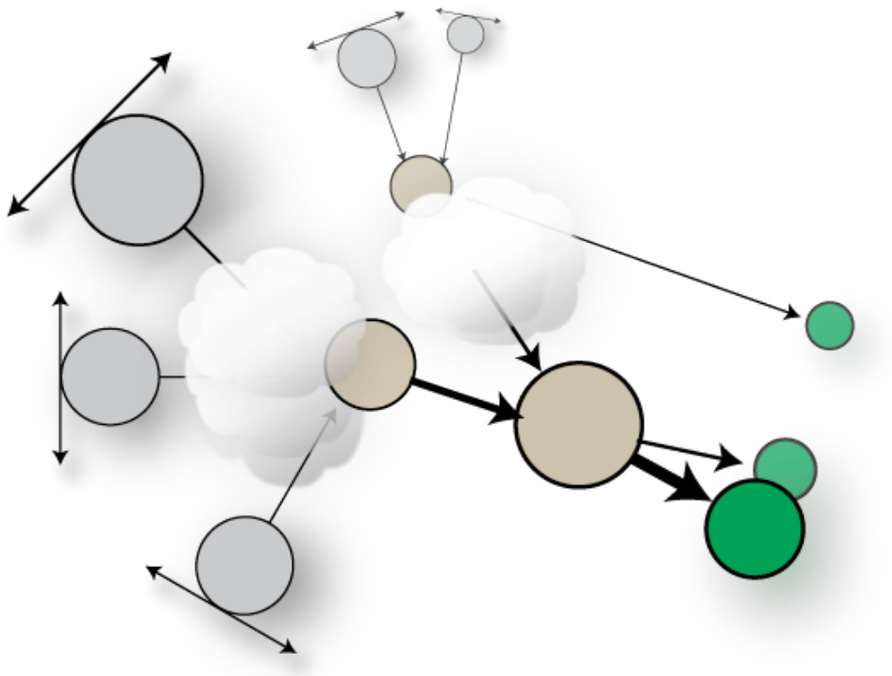
- Local Distribution and Storage
  - Best methodology
    - Provides explicit scalability and performance capabilities
    - Provides most reliable collection architecture
    - Multiple points of attachment, multiple points of control
  - Most expensive strategy at data generation
    - Radium deals with device and remote client requests for data which does come with a processor and memory cost



Argus writing to local Radium which is writing directly to disk and to network based clients

Data Collection

- Complex data flow machine architectures
  - Architecture of choice for scalability
    - Provides explicit scalability and performance capabilities
    - Provides most parallelism
    - Multiple points of attachment, multiple points of control
  - Can get a little complex
    - Merging of multiple flows, multiple times, introduces complex data duplication issues, and allows for complex, incompatible data schemas to co-exist



QoSient

Many Argi writing directly to a Radium based distribution network, which is providing data to a set of clients.

Radium

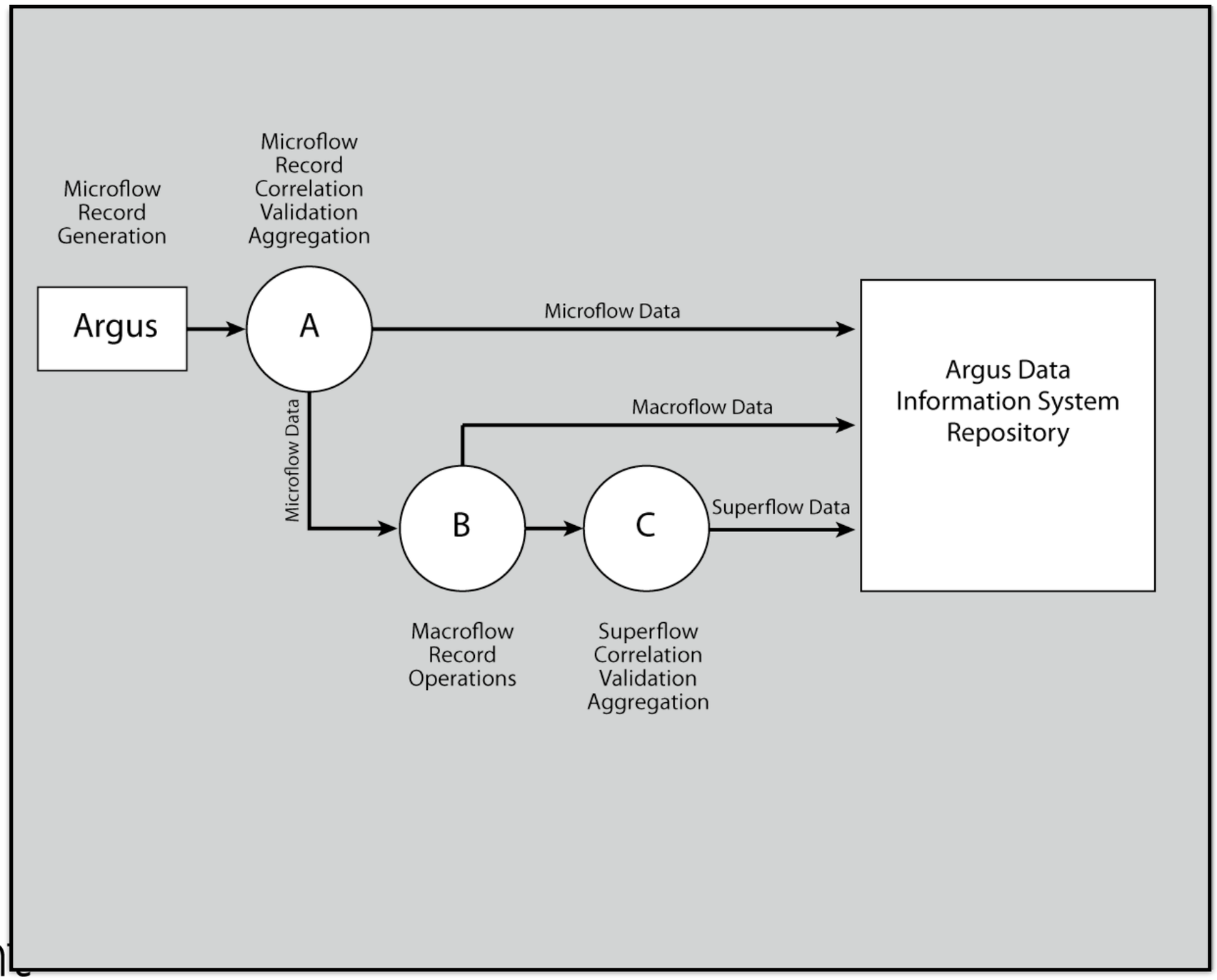
- Real-time operation
  - Radium, is designed as a non-blocking data distribution node
  - Implemented as multi-threaded input and output processor(s)
    - Input processed and placed in single process queue
      - Read up to 256 argus data sources
      - Generates 1 output data stream
    - Queue manager continuously distributes records to the collection of output processors
    - The more cores, the less queuing, locking and scheduling

  - Aggregation and analytics introduce delay
    - rabins() function demands buffer holding times
      - Aggregation over a fixed period of time.
    - Stream Analytics - process within locked time "bin"
    - Queue manager must wait for analytics to complete

QoSient

Radium

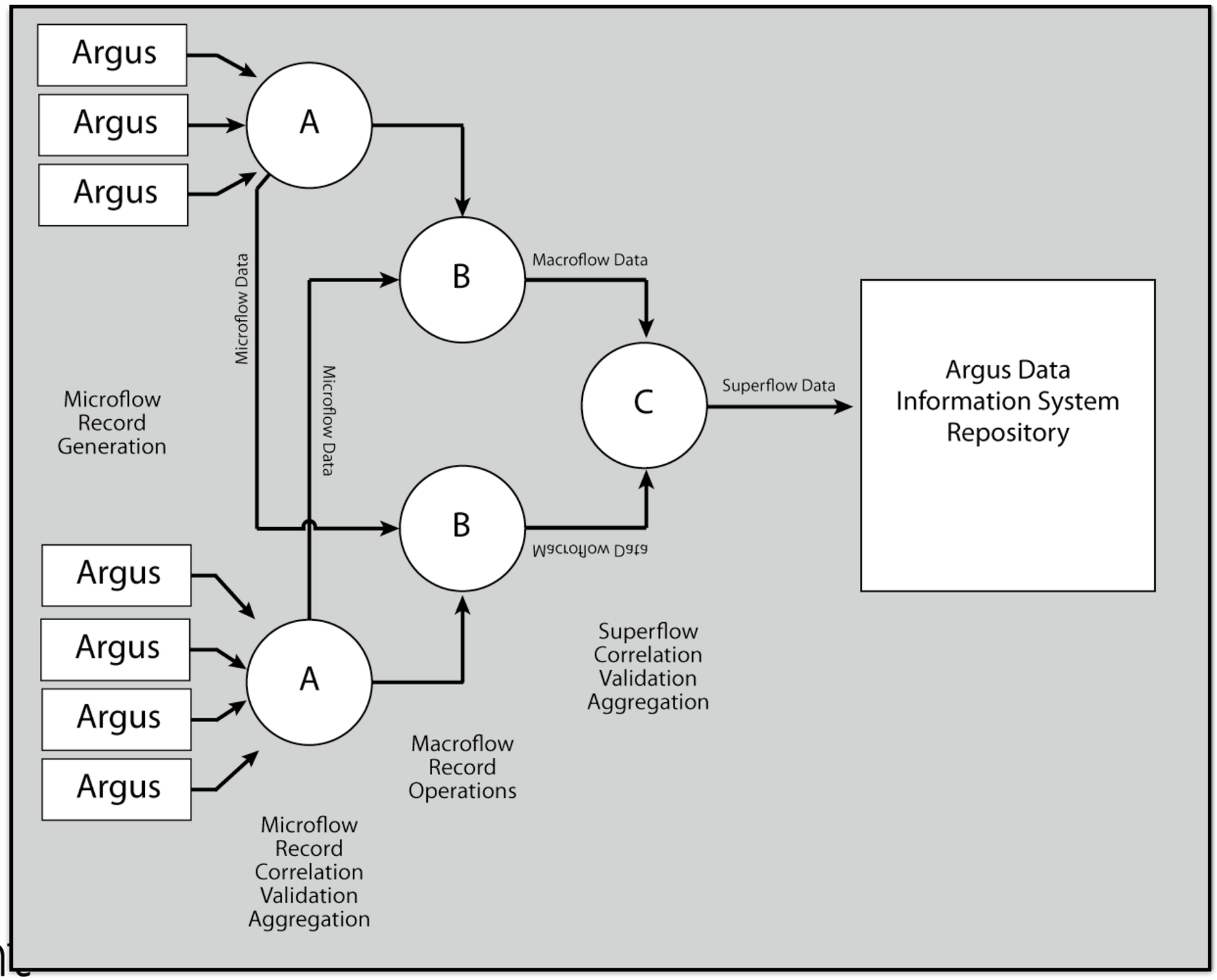
Data Flow Machine Architectures



QoSient

Radium

Data Flow Machine Architectures



QoSient

Argus Repositories

Complex Collection

QoSient

Argus Repositories

- Argus Repository Establishment
  - Formal Ingest/Disposition
- Repository Function
  - Primitive Data Repository
    - General Archive
    - Access Control
    - Retention Policies
    - Modification Policy (Compression)
  - Derived Data Repositories

QoSient

Argus Repositories

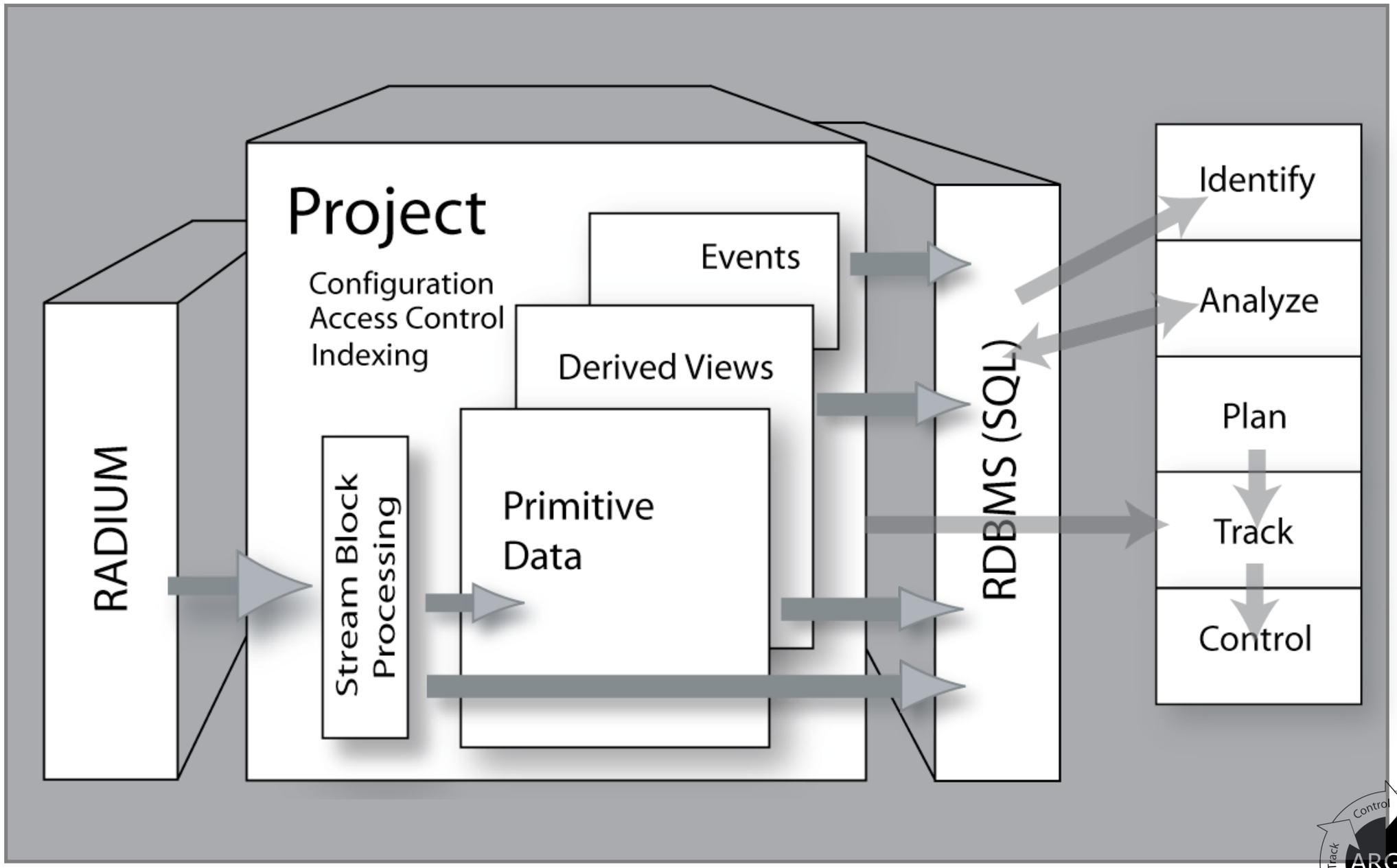
- Native File System
  - Simplicity
  - Performance
  - Compatibility

- Relational Database System (RDBMS)
  - Extensive Data Handling Capabilities
  - Complex Management Strategies
  - Performance Issues

QoSient

Argus Processing Design

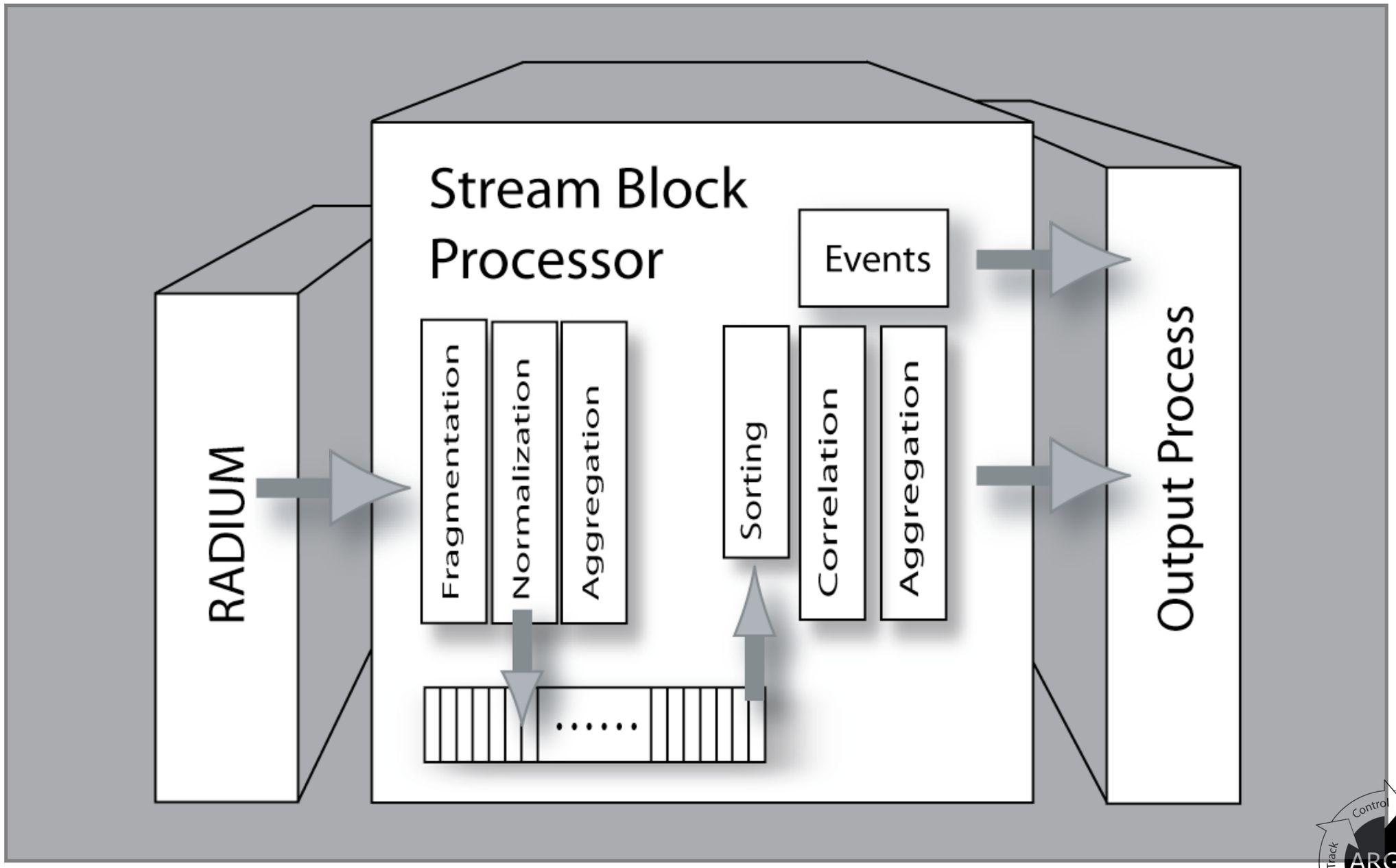
Radium Stream Block Processor



QoSient

Argus Processing Design

Stream Block Processor



QoSient

Argus Repositories

Data Ingest Support

- Stream Block Processing
  - rasplit
  - rastream
  - rabins
  - rasqlinsert

QoSient

Argus Repositories
Best Common Practices

- File system archives
  - Primitive and derived data file systems
  - RDBMS managed complex indexing
  - rastream
    - /sourceId/year/month/day file structure
    - 5 minute files
      - 288 entries per day
      - Matches native file system performance for searching
      - Analogous to Google's Big Table filesystem
- RDBMS based archives
  - Short term data held in RDBMS
  - Rolled into file based system after N days.
  - Binary data inserted into database
  - Primitive data schema includes 'autoid'
  - Table names provide explicit partitioning

QoSient

Argus Repositories
Real Time Processing Strategies

- rasqlinsert based data insertion / management
  - Complete argus data analytic engine
    - Complex aggregation support
    - Semantic enhancement
    - Time and data correction

- Continuous flow status maintained in table
  - Configurable update refresh intervals
  - Idle timeout options provides windowed SA

- RDBMS handles concurrency: updates and access
- RDBMS enabled trigger support

QoSient

Argus Repositories

Real Time Processing Strategies

- rasql based client data access
  - RDBMS handles multiple access
  - Maintains cache management and concurrency
  - Local and remote access through federation

QoSient

Argus Client Programs

- Basic Operations
- Aggregation
- Data Splitting
- Graphing
- User Data Processing
- Semantic Enhancement
- Anonymization

QoSient

Argus Client Programs

- Basic Operations
  - Printing, Filtering, Sorting, Splitting, Aggregation
  - Collection, Archiving, Anonymization
- Graphing/Visualization/GUI
- Data Enhancement
  - Labeling
  - Geolocation
- Database Support
- User Data Processing
- Analytics

QoSient