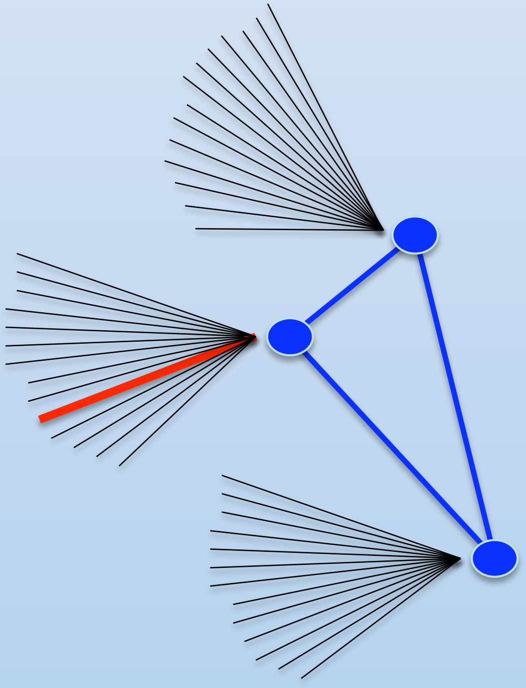
DMnet: Detection Mitigation Network

A Behavioral Analysis System Supporting Trust Measurements



Owen McCusker, Scott Brunza,
Sonalysts, Inc.
{mccusker,brunza}@sonalysts.com

Dr. Carrie Gates,
CA Labs
carrie.gates@ca.com

Joel Glanfield, Diana Paterson,
Dalhousie U.
{joelglanfield, paterson}@cs.dal.ca

An	To an
ADAPTIVE	ADAPTIVE
DISTRIBUTED	DISTRIBUTED
DYNAMIC	DYNAMIC
Approach	Threat

SONALYSTS

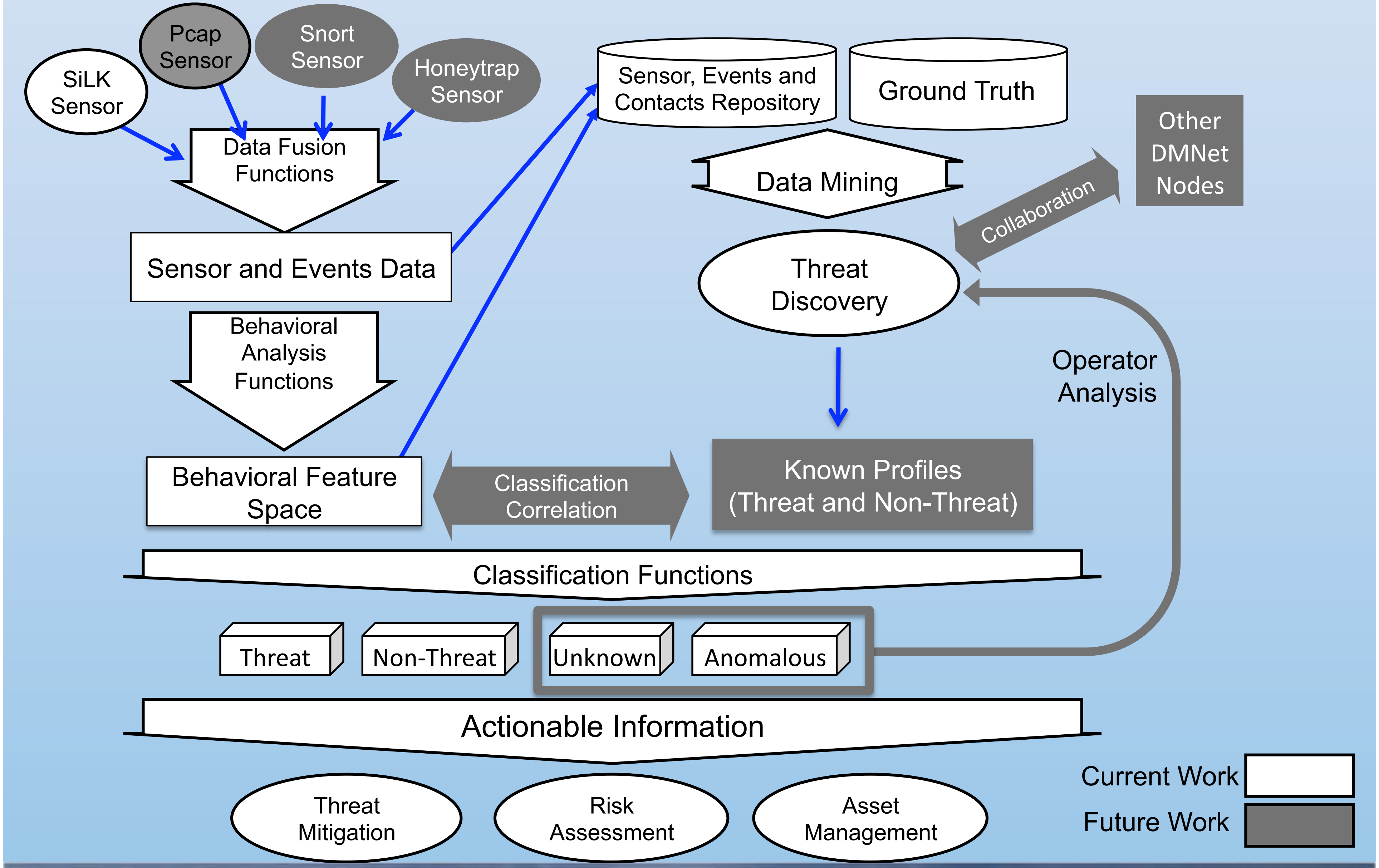
Agenda

- The Need
- Behavior-based Classification
- Trust Derived from Behavior
- Where We are Today

The Need

- The cyber threat is distributed, dynamic, and multi-scale in time
  - IDS technology is focused on “**single source**” solutions, “**single time-scales**”.
  - Threats are buried in the noise of everyday traffic
  - Cyber defense technologies adapt mostly through the use of signatures (exception: Anomaly Detection)
- We need enabling technologies that facilitate the creation of **adaptive** and **open** distributed defense technologies
- Our Contribution:
  - Creation of an **aggregated behavioral feature space**
  - **Separation** of **trust** from **behavior aggregation**
  - Initial use of **ontology** to map behavior to threat
  - Share behaviors between COIs to break through privacy barriers

Approach: DMnet Architecture



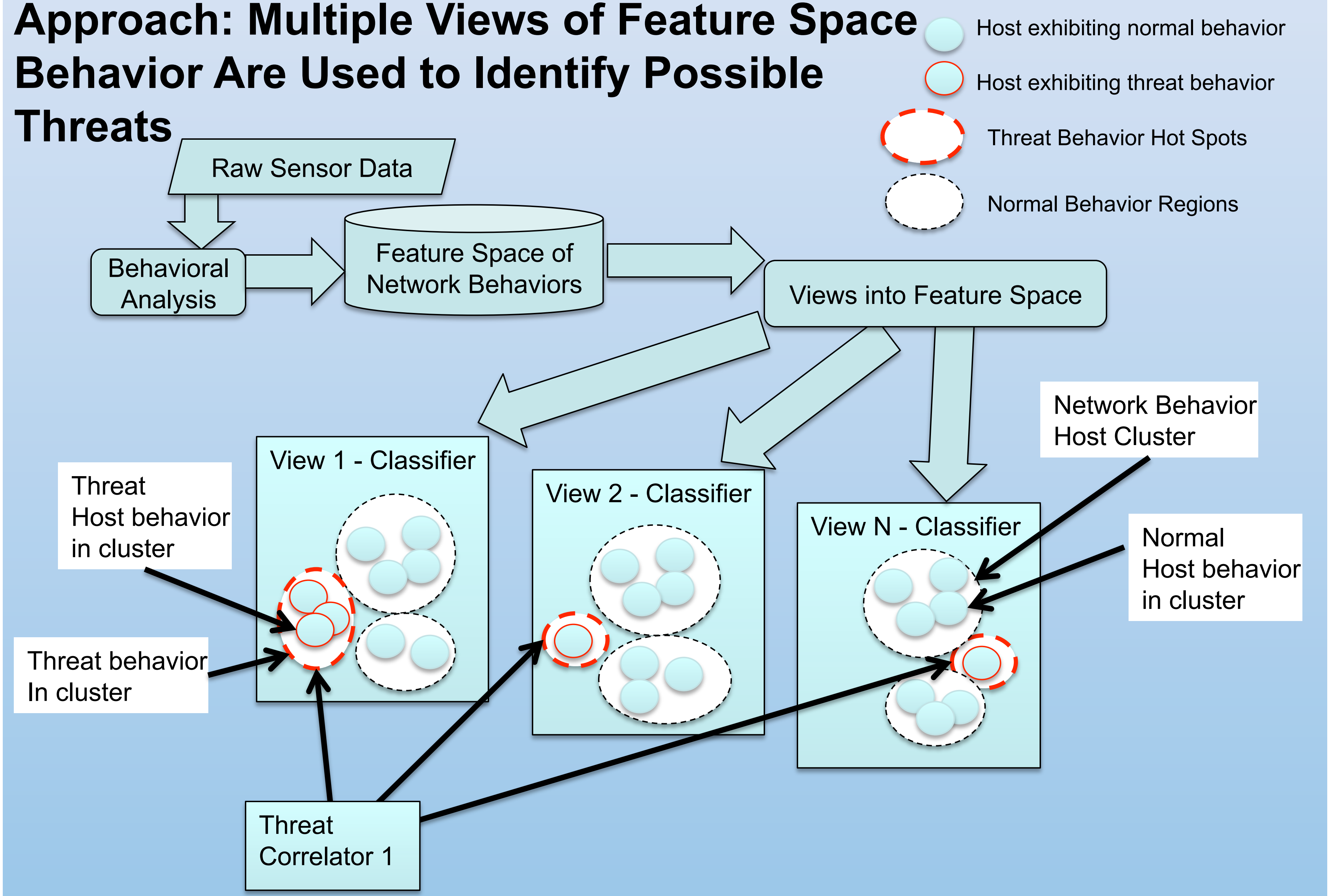
Behavior-based Classification

- Ingest events and data from multiple sensor types
  - Architecture supports different sensor types
  - Currently using SiLK
- Derive features from capture events and data
  - Create a rich feature space used for behavioral analysis
  - Leverage primitive features during analysis
- Identify Behaviors
  - The Goal is to create a **behavioral language** used to describe and identify cyber threats
  - Based on analyzing feature space using different n-tuple sets.

Behavior-based Classification

- Threats are detected by the **identification of multiple behavioral patterns**
  - Has an "analog" to OCR and voice recognition
- Threat behavior can be characterized / detected using **adaptive heuristics**.
  - Success despite primitive state of current rule set.
  - Architecture will support concurrent use of more complex and adaptive heuristics.
- Analysis can be **enhanced** by:
  - **Increased Community of Interest size** (number of correlated network sensors),
- **Automation** necessary to improve / expand analysis.
  - There is a need for a common behavioral **Ontology**
- **Application of "Hyperplaning" for Botnet Detection**
  - University of Connecticut

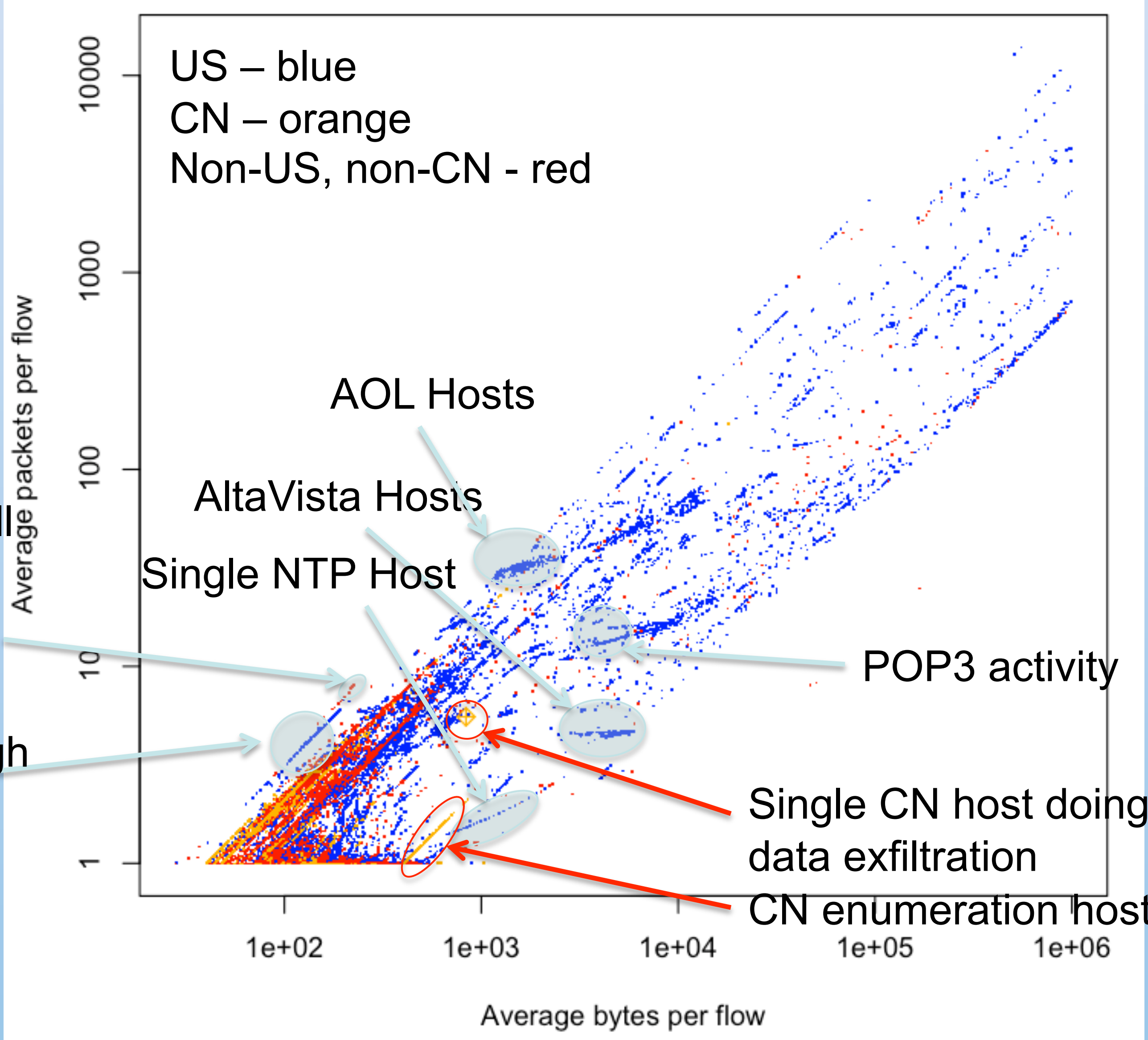
Approach: Multiple Views of Feature Space Behavior Are Used to Identify Possible Threats


Host exhibiting normal behavior
Host exhibiting threat behavior
Threat Behavior Hot Spots
Normal Behavior Regions
Raw Sensor Data
Behavioral Analysis
Feature Space of Network Behaviors
Views into Feature Space
View 1 - Classifier
View 2 - Classifier
View N - Classifier
Network Behavior Host Cluster
Normal Host behavior in cluster
Threat Host behavior in cluster
Threat behavior In cluster
Threat Correlator 1

SONALYSTS

Benefit: Weak Signal Cyber Detection (Threat signals stand out of the Noise)

CN Data Exfiltration Case Post-Event Comparison of Selected Host



AOL Hosts

AltaVista Hosts

Single NTP Host

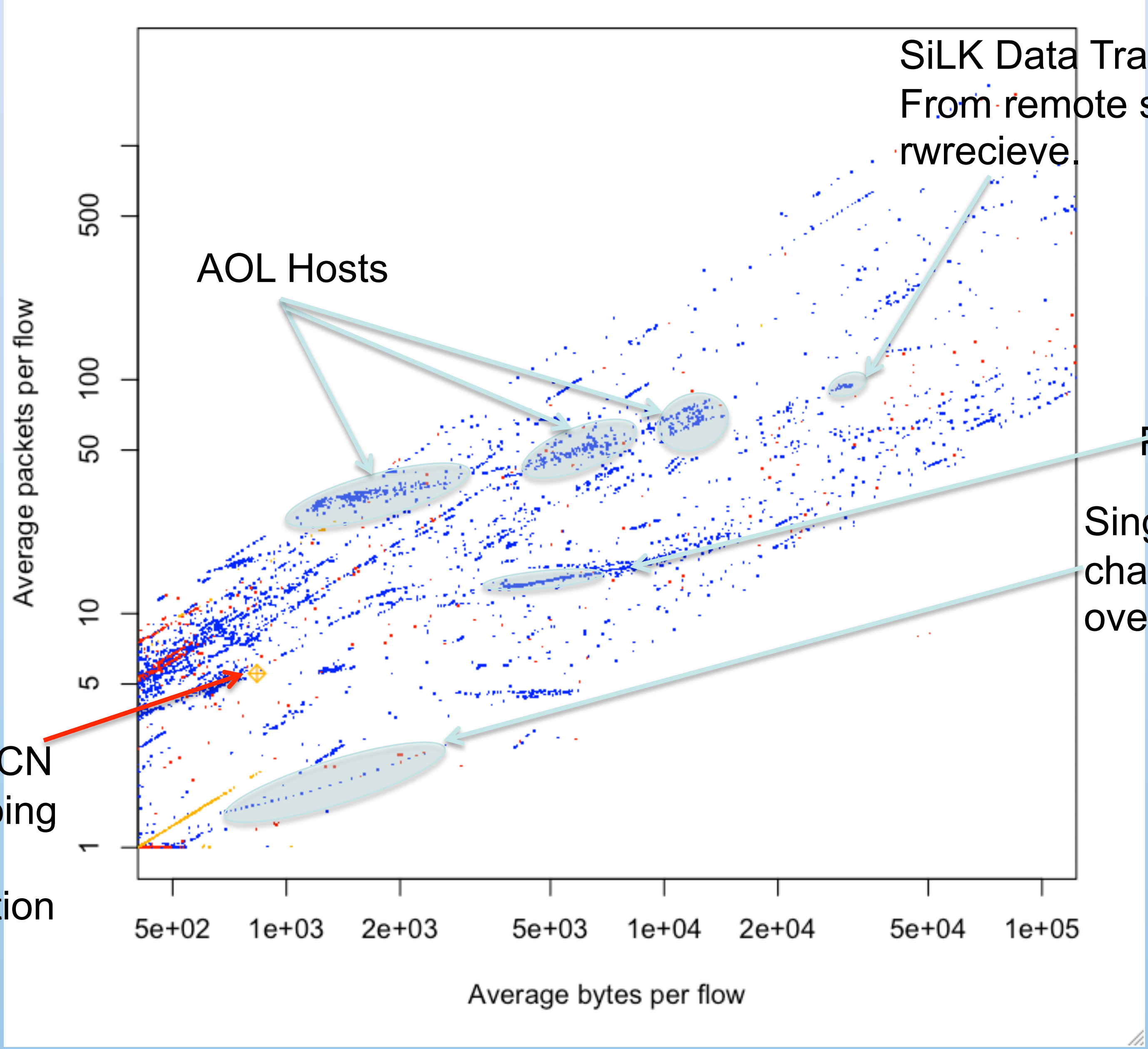
Pings hitting firewall from around globe. Panther Express included.

Hosts NAT'd through firewall

POP3 activity

Single CN host doing data exfiltration

CN enumeration hosts



SiLK Data Transfers From remote site. rwsend, rwrecieve.

AOL Hosts

POP3 activity

Single NTP server changing behavior over month

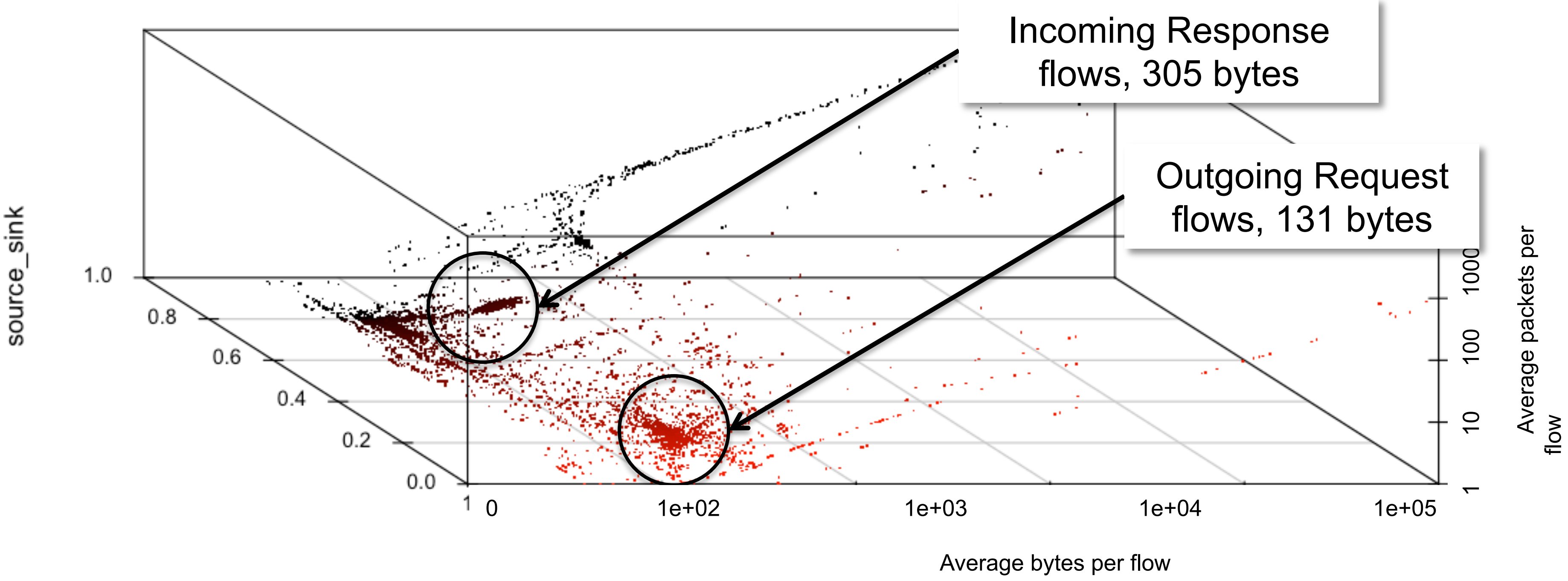
Single CN host doing data exfiltration

Differentiator: Adding in another visual dimension (source/sink) to separate out host behaviors

October (ingested 5 days)
Total Contacts: 251K

Red is source, Black is Sink
DHT Protocol

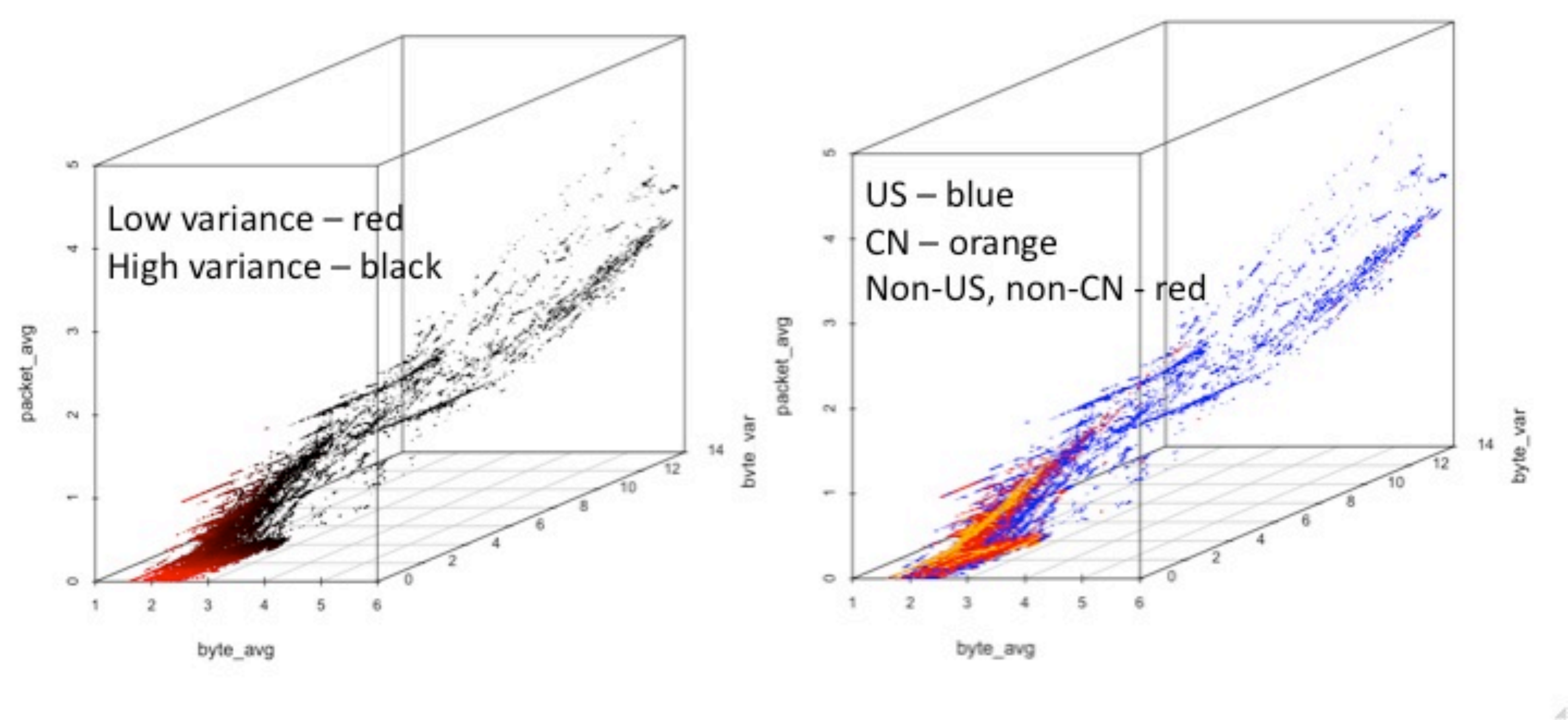
Log(Avg Byte) vs Log(Avg Packet) vs Source-Sink



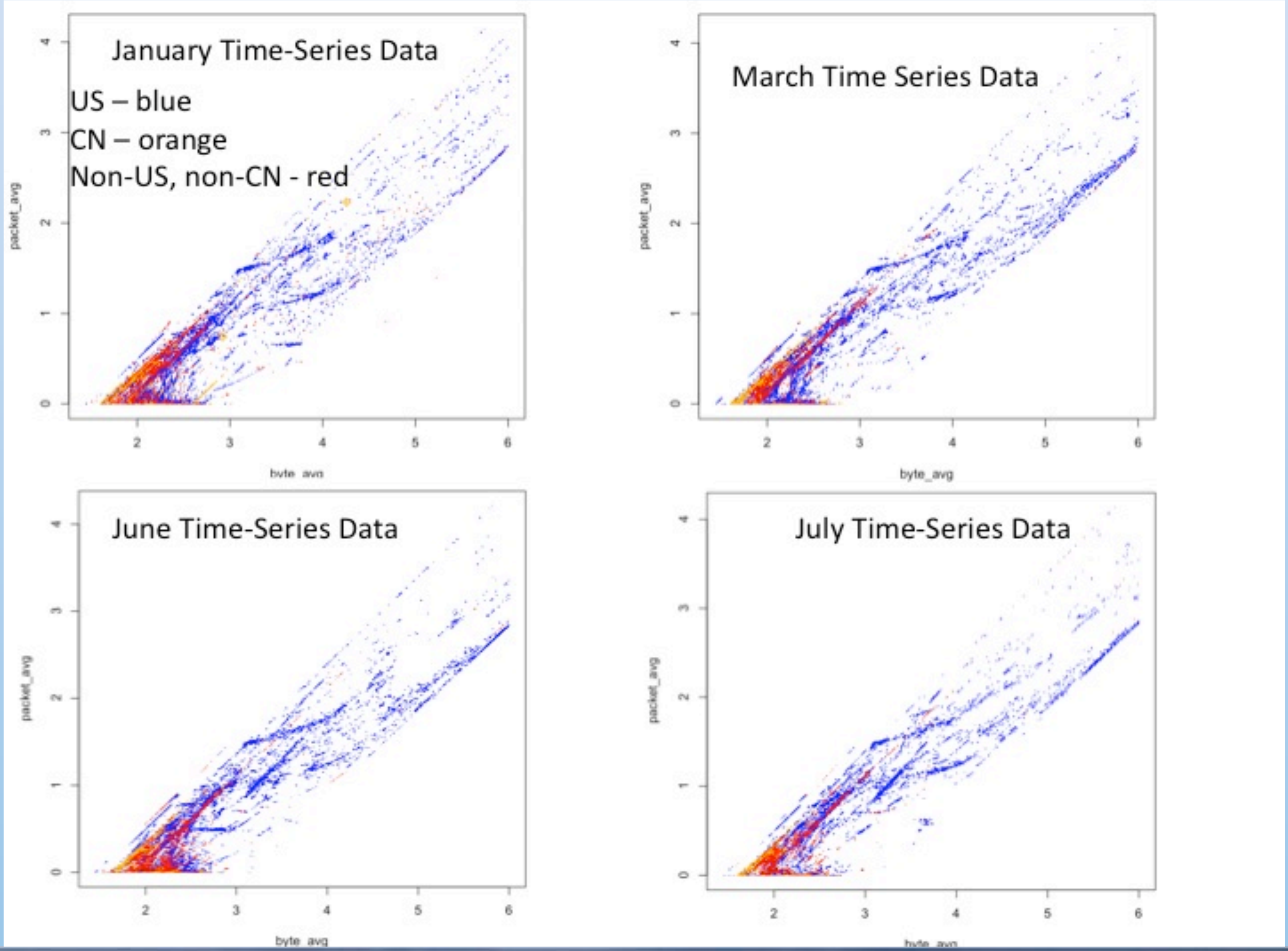
Incoming Response flows, 305 bytes

Outgoing Request flows, 131 bytes

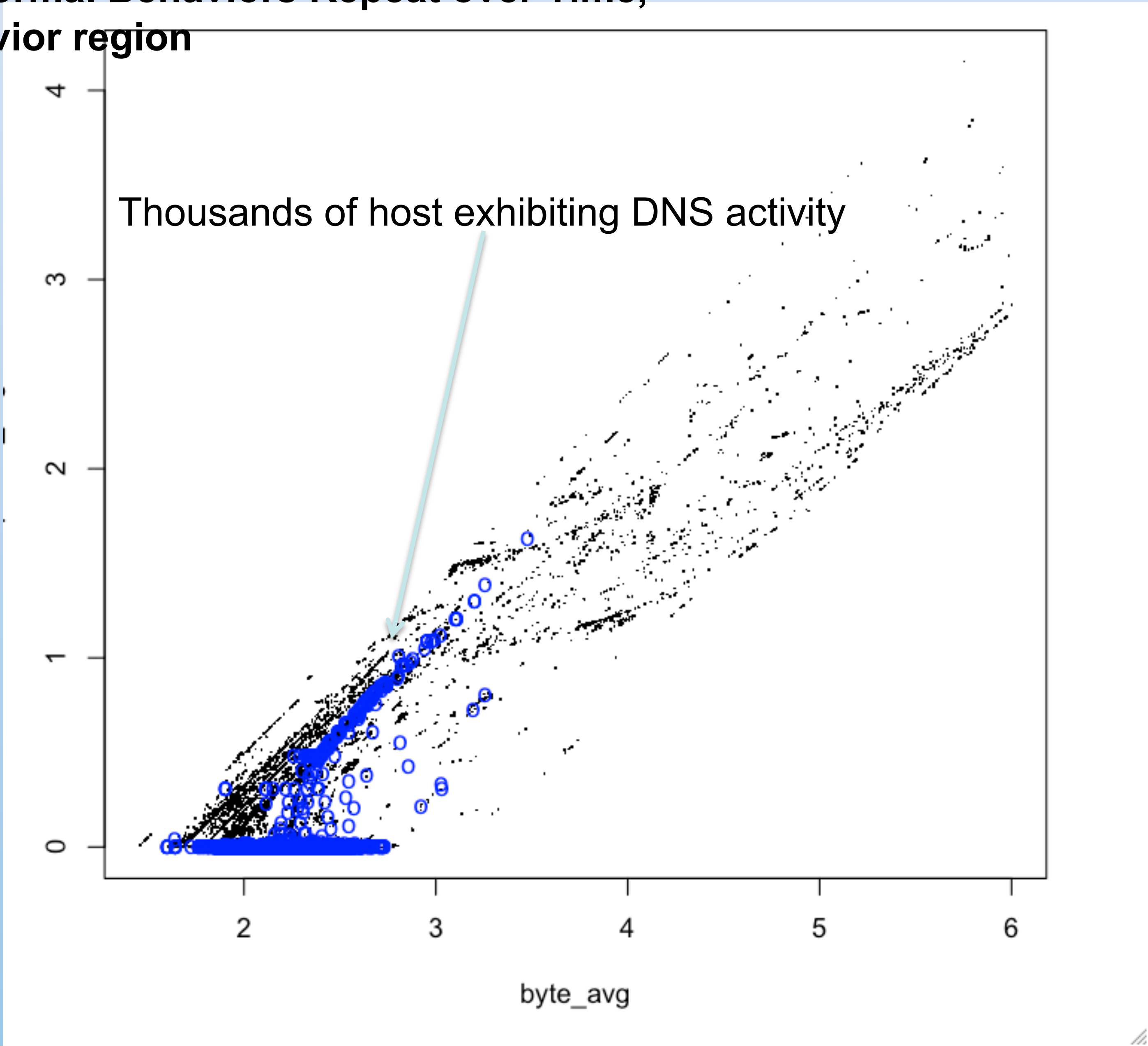
Differentiator: Adding in another visual dimension (byte variance) to separate out host behaviors



Benefit: Normal Behaviors Repeat over Time, DNS behavior region

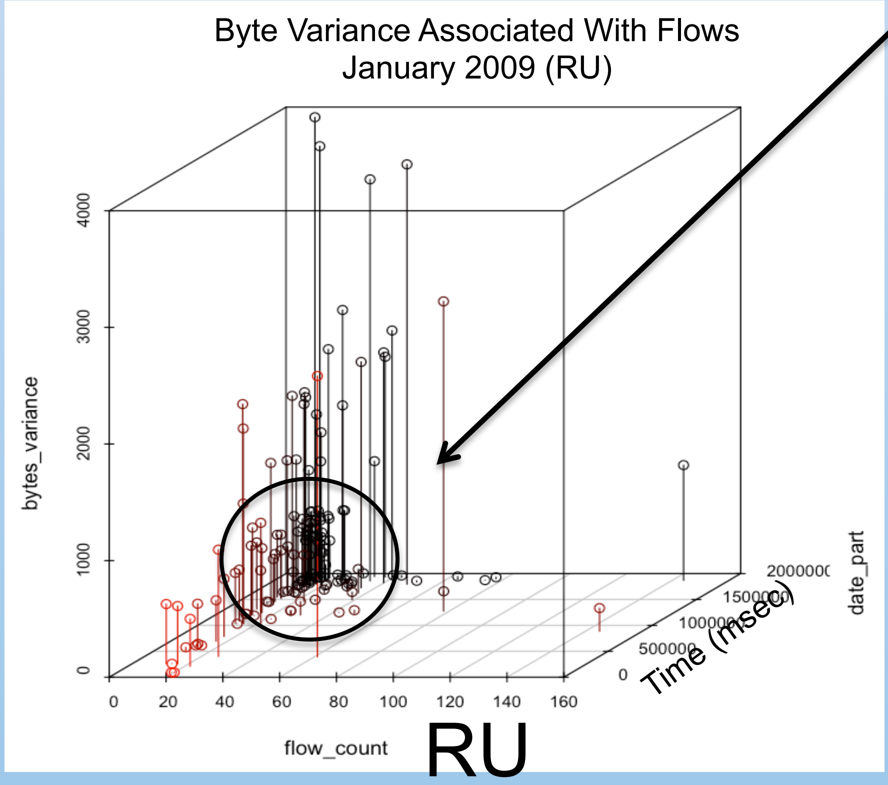


Benefit: Normal Behaviors Repeat over Time, DNS behavior region

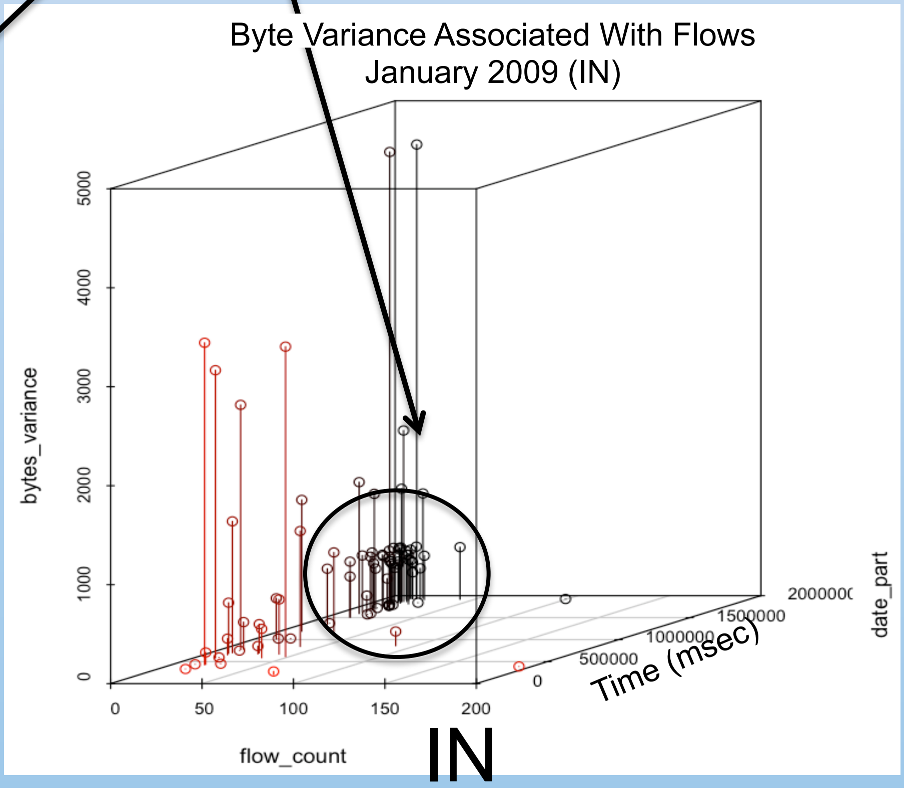


Benefit: Threat Anticipation Regional Byte Variance Over Time

Unexplained / suspect behavior originating from two different regions at the same time.

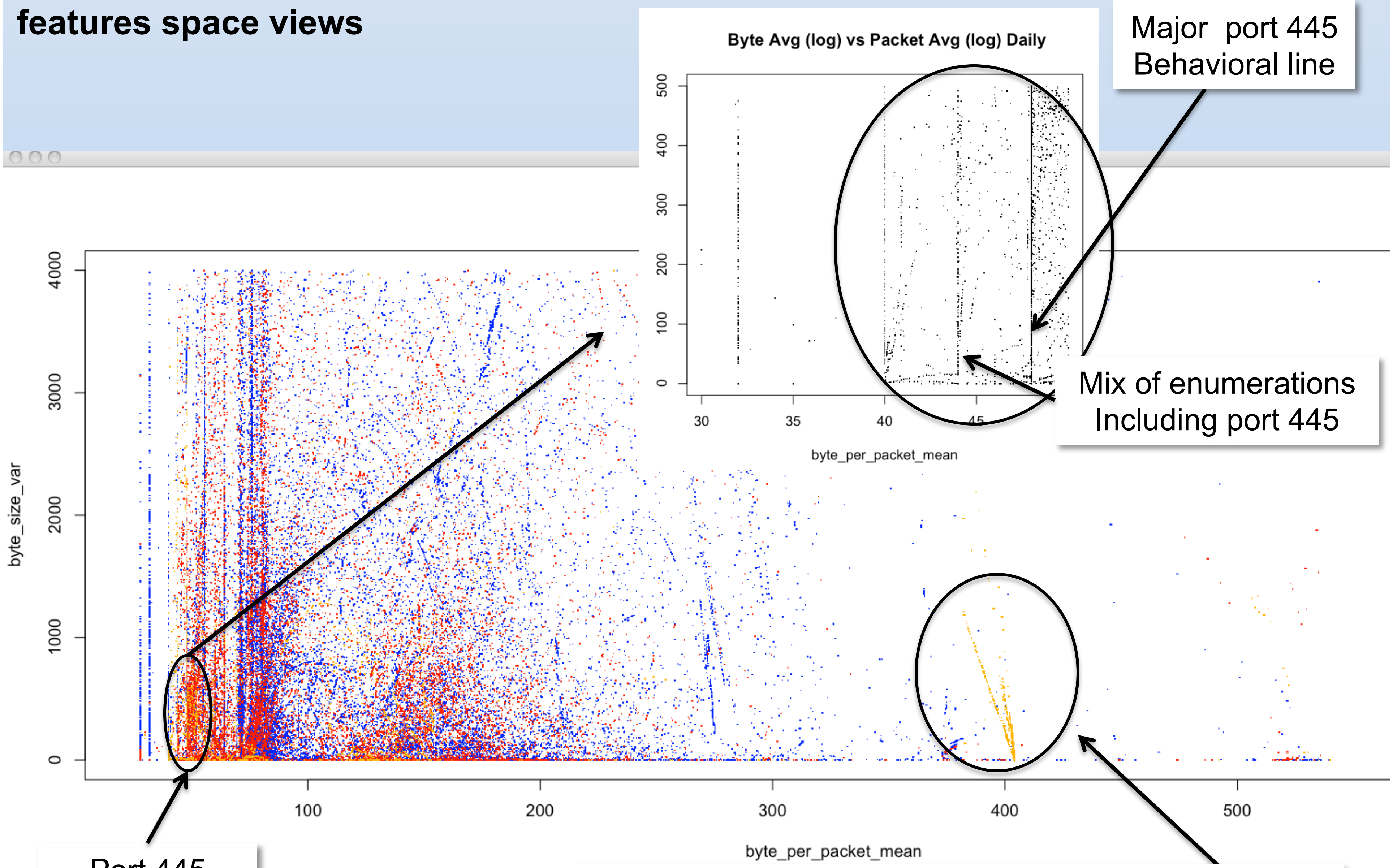


Byte Variance Associated With Flows
January 2009 (RU)



Byte Variance Associated With Flows
January 2009 (IN)

Benefit: Consistent Behaviors within features space views



Major port 445 Behavioral line

Mix of enumerations Including port 445

Port 445 enumerations

SQL Slammer (1434) - Byte per packet mean 404
These hosts send 403 to 404 rations
Number of hosts is 1040, **Mostly from CN**

Trust Derived from Behavior

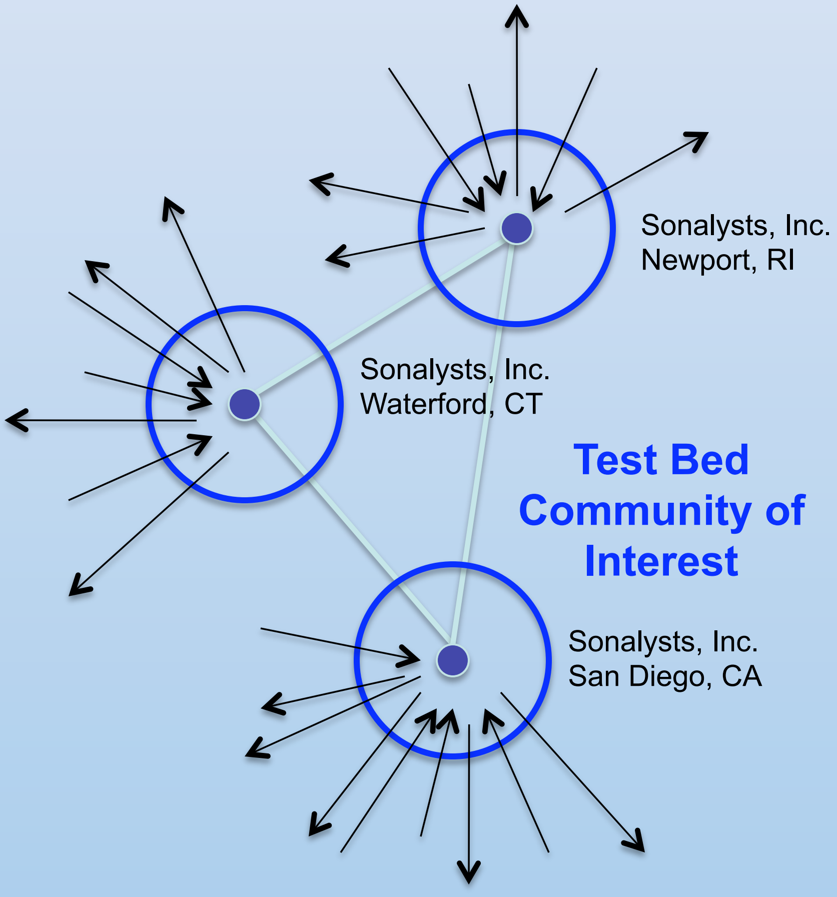
- Trust is **subjective** driven by the security policies of the institution
  - Host network behaviors are **objective**
- Trust is too difficult to share without a common understanding of risk
- The overall trust of a host is a weighted sum of all trust behaviors
  - Each **measured behavior** is given a value of **trust**
- The change in behaviors can be used as a measure in trust
  - Use of multiple protocols, compared to single protocol
  - High variance in byte usage
  - High variance in entropy of payload

Where are we Today?

- ITT – **3rd Party evaluation**
  - Web-base interface, CLI tool
  - Identified threats using a simplified set of heuristics
  - Solid system, more work needed false positives/negatives
- Researching the application of **Biological Immune System (BIS)** concepts to system
  - **Self, Non-self concepts** combined with **Computational Trust**
- Created a **commercial** service for network analysis
  - **Behavioral Analysis leading to Situational Awareness**
- HPC based architecture
  - Created a small cluster of nodes using **OpenMPI** to test scaling our system
- High Bandwidth
  - Just starting integrating and **Endace DAG** for network capture

Approach: Test Bed

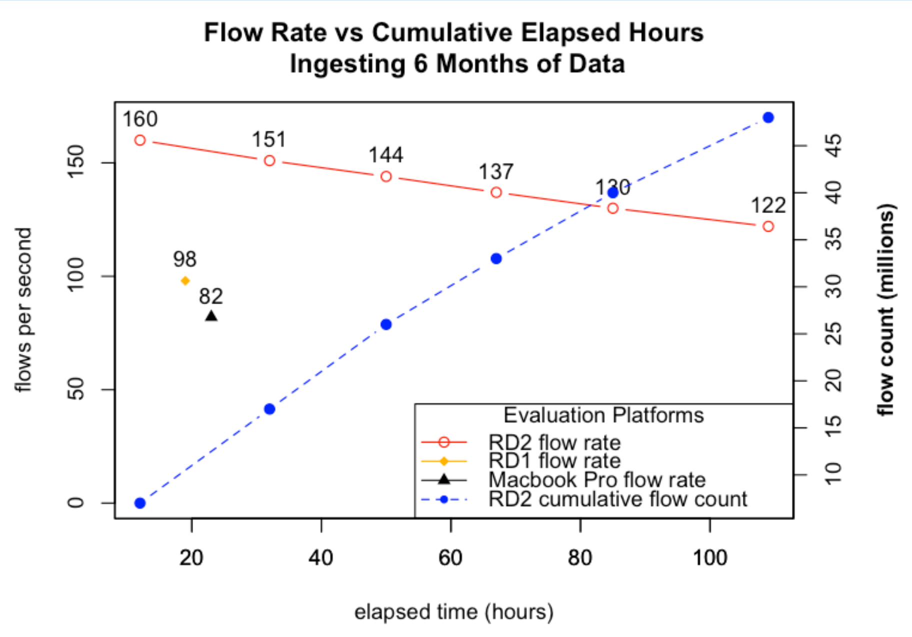
- Using:
  - Networked sensors leveraged across a trusted community of interest,
  - **Host-centric behavioral aggregation**
  - Multiple data fusion and mining methodologies, and
  - Concurrent classification and correlation algorithms.
- To:
  - Connect the dots over long time periods (e.g. months).
  - Detect and characterize threat behavior in near real time.
  - Perform weak signal cyber detection.
- Sharing **just** behaviors **minimizes** impact of **user data privacy**



Today:

- Monitor incoming & outgoing traffic outside the perimeter.
- ~400,000 host contacts / month.
- Characterize host behavior and look for changes over time that suggest threat behavior.
- 72 basic characteristics extracted and synthesized.
- Correlate with country of origin.

DMnet – RD2 Prototype



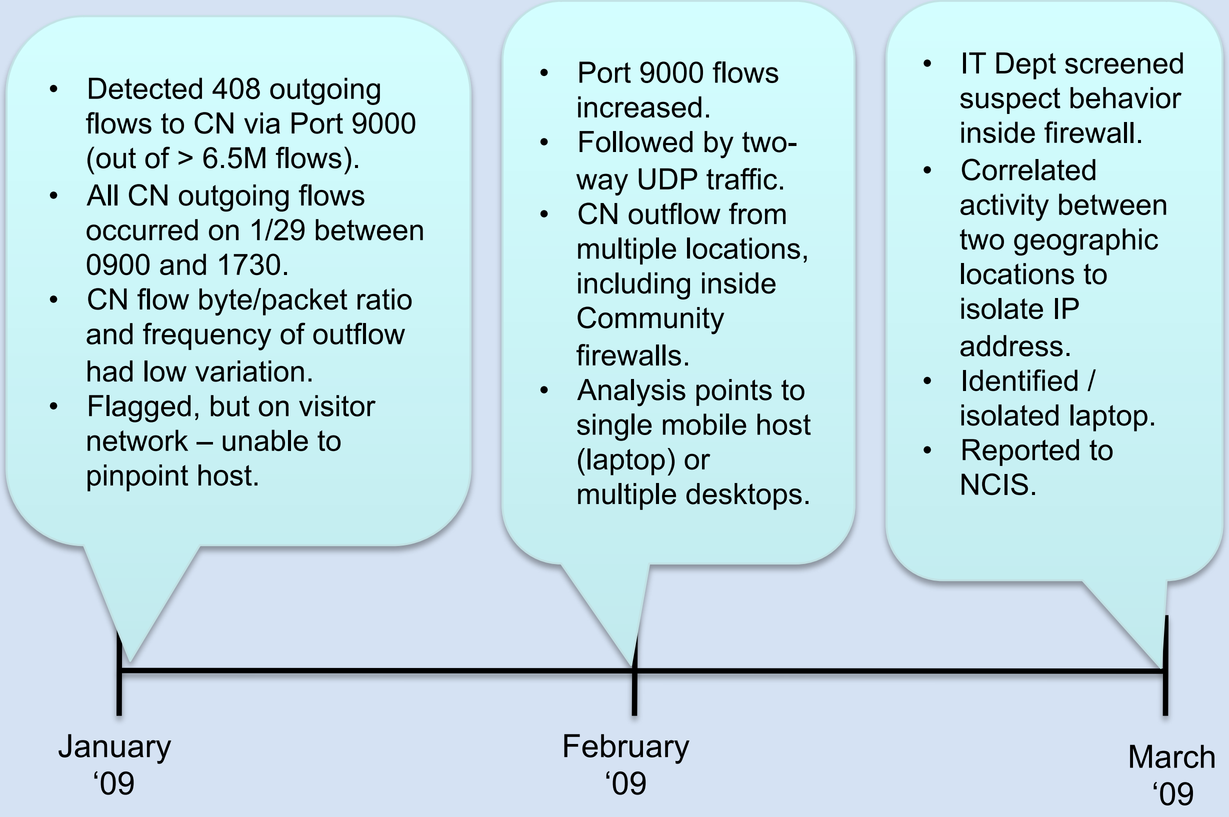
Prototype Evaluation

- Processed 6 months of data on site in 4 days
  - 30 times real time for Sonalysts
  - Improved performance from last year
- TODO:
  - Integrate behavioral learning
  - Integrate classifiers and correlators

Benefit: Successful Detection of Data Exfiltration to a Sophisticated "CN Host" via Company Laptop

Monthly, we conduct semi-automated regional analysis of outgoing network flow.

- Analysis script looks at all incoming and outgoing data to community of interest.
- Heuristics set used to parse data into manageable subsets.
  - Based on location, port usage, direction, port pair bandwidth utilization, IP address bandwidth usage, client-server behavior, protocol analysis.
- Manually review subsets looking for anomalies
  - Trend analysis of past reports.



- Detected 408 outgoing flows to CN via Port 9000 (out of > 6.5M flows).
- All CN outgoing flows occurred on 1/29 between 0900 and 1730.
- CN flow byte/packet ratio and frequency of outflow had low variation.
- Flagged, but on visitor network – unable to pinpoint host.

- Port 9000 flows increased.
- Followed by two-way UDP traffic.
- CN outflow from multiple locations, including inside Community firewalls.
- Analysis points to single mobile host (laptop) or multiple desktops.

- IT Dept screened suspect behavior inside firewall.
- Correlated activity between two geographic locations to isolate IP address.
- Identified / isolated laptop.
- Reported to NCIS.

CN exfiltration was not and likely could not have been detected by existing firewall technology.

Thank you!

Owen McCusker, Scott Brunza,
Sonalysts, Inc.
{mccusker,brunza}@sonalysts.com

Dr. Carrie Gates,
CA Labs
carrie.gates@ca.com

Joel Glanfield, Diana Paterson,
Dalhousie U.
{joelglanfield, paterson}@cs.dal.ca