Cyber Analytics for US-CERT

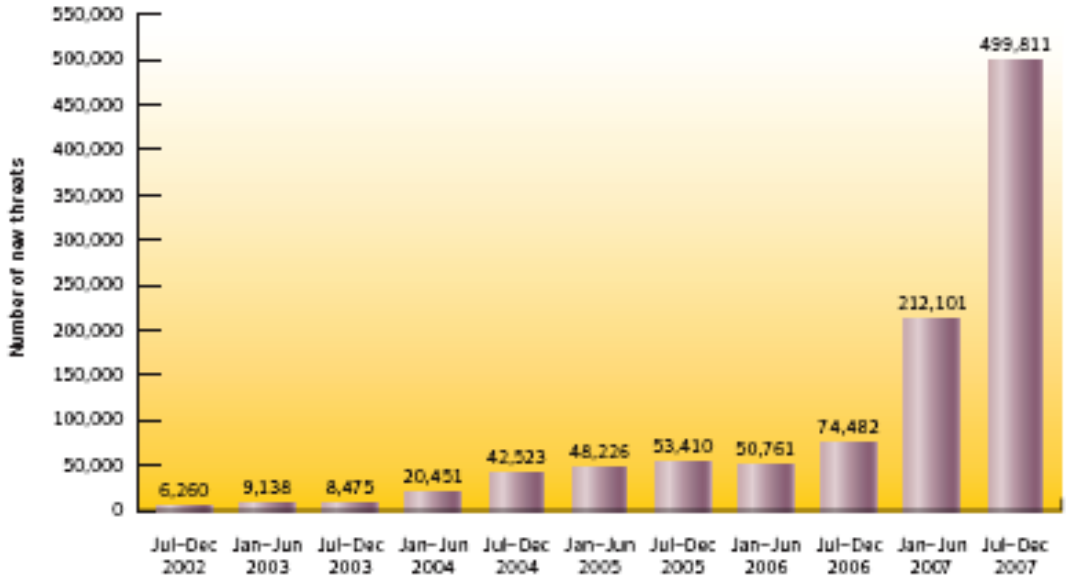
Transitioning network flow visualization from the laboratory to the watch floor

Bill Pike
Pacific Northwest National Laboratory
bill.pike@pnl.gov

John Gerth & Justin Talbot
Stanford University
gerth@graphics.stanford.edu

Pacific Northwest
NATIONAL LABORATORY

Visual analytics for high-volume data



- Challenges to anomaly detection and characterization in computer network communications:
  - Lots of data (billions of transactions/day)
  - Lots of unique actors
    - IPv4: 4.3 million unique IP addresses
    - IPv6: $6.67 * 10^{27}$ IP addresses *per square meter*
  - Lots of noise

Pacific Northwest
NATIONAL LABORATORY

What can visual analytics do for cyber security?

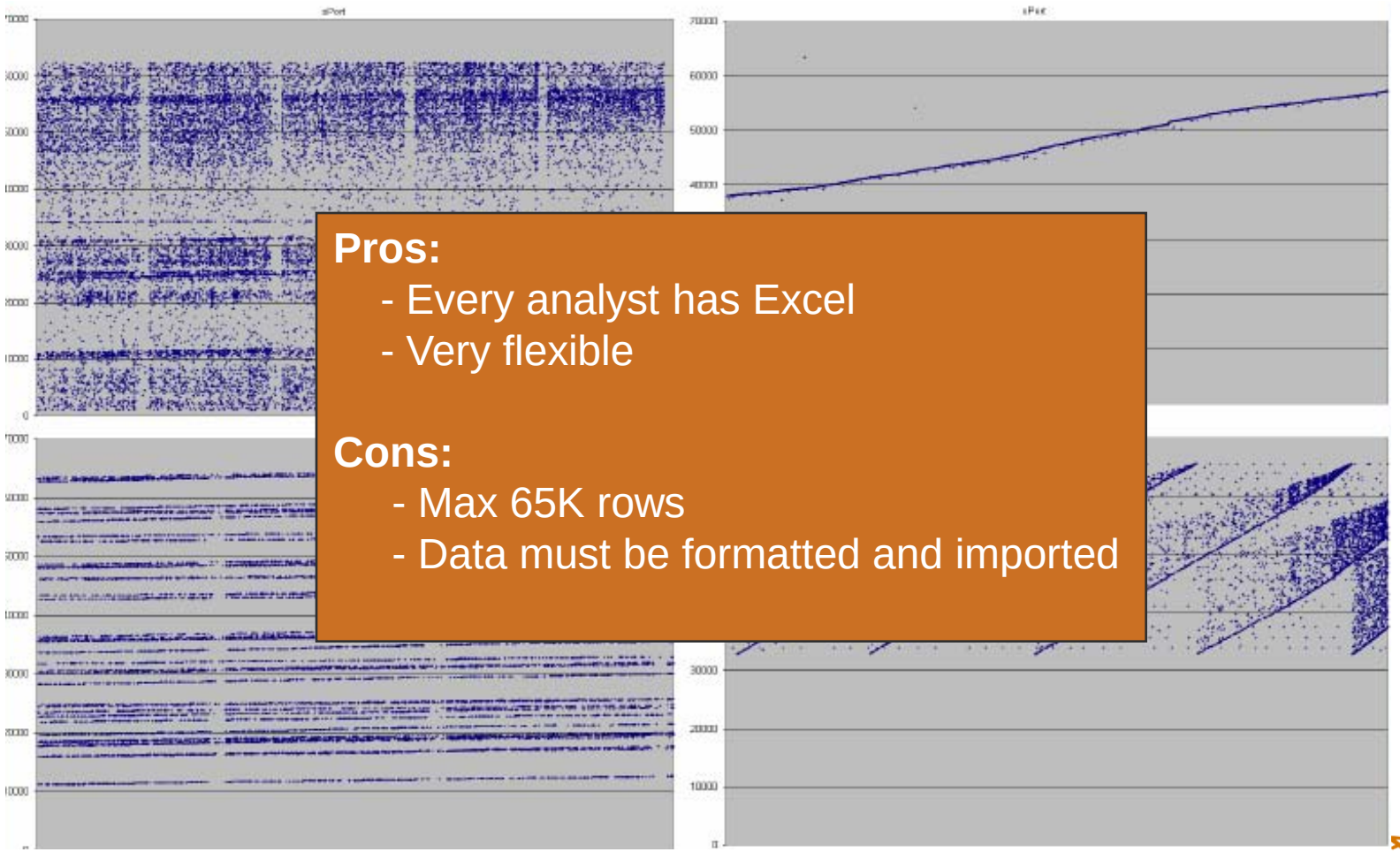
- If we know what we're looking for, we can build a signature to detect it. **But what's in the data that we don't already know to look for**?
- **Approach:** Create a new ability to scale between "50,000ft" situational awareness and "ground level" analysis of individual transactions.
  - **PNNL** | NUANCE network flow overviews
  - **Stanford** | Isis event browsing and correlation
- **Goals:**
  - Deploy a scalable visualization suite at US-CERT to visually discover emerging threats in high-volume streaming data.
  - Link laboratory and academic products into a single suite.

Pacific Northwest
NATIONAL LABORATORY

US-CERT Mission

- Protect critical infrastructure in cyberspace – both public and private sector.
  - Analyze and reduce cyber threats and vulnerabilities.
  - Disseminate cyber threat information.
  - Coordinate incident response activities.
- US-CERT's EINSTEIN program collects summary network traffic information at agency gateways and provides a high level view of federal government network connections.
- US-CERT analysts use EINSTEIN data to correlate cross-agency network events.

One current visualization tool for EINSTEIN flow data



Pros:

- Every analyst has Excel
- Very flexible

Cons:

- Max 65K rows
- Data must be formatted and imported

Pacific Northwest
NATIONAL LABORATORY

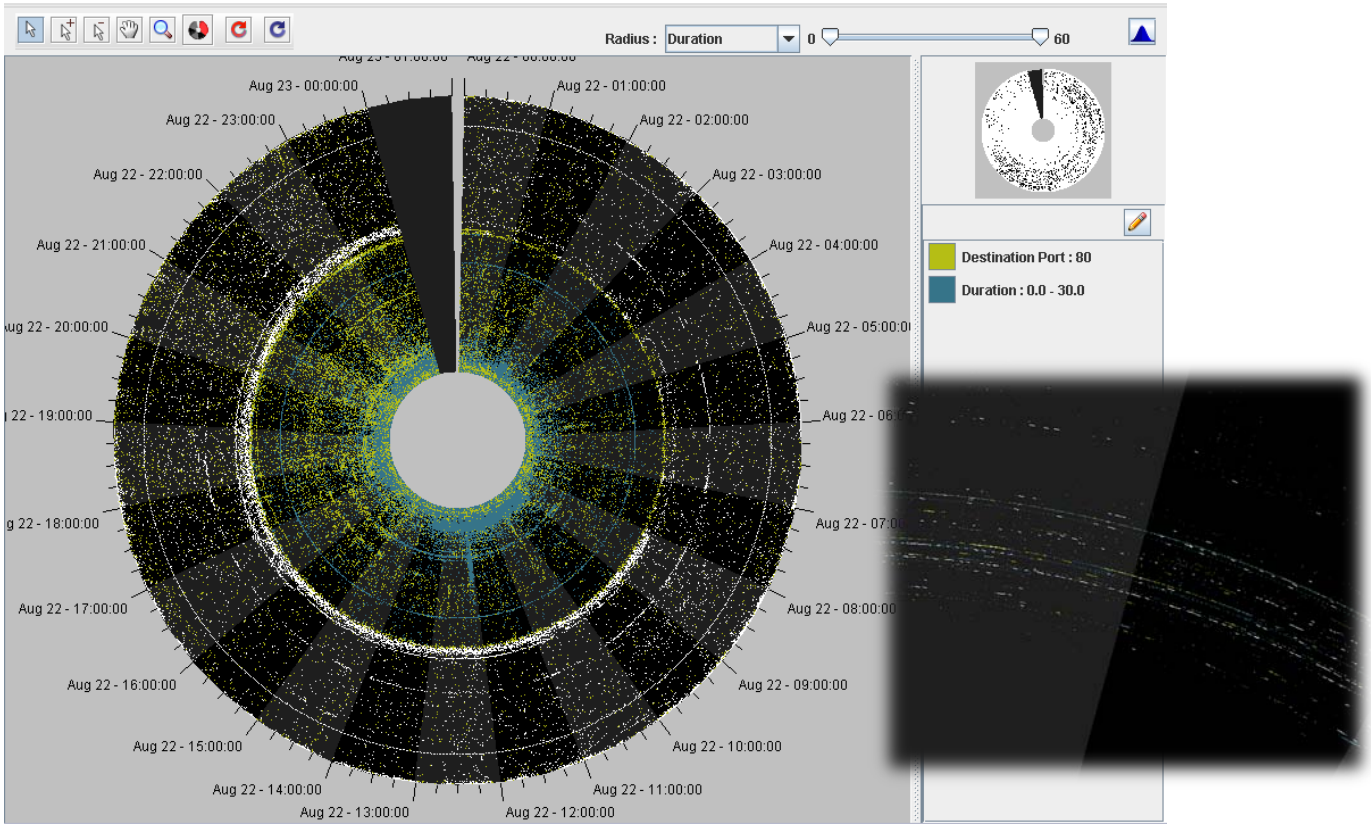
Our approach
Scalable exploration of network flows

- Collect analyst requirements
- Customize existing tools to support a new level of situational awareness and exploratory analysis
  - **PNNL:** NUANCE Traffic Circle
    Generate high level overviews of large data sets.
  - **Stanford:** Isis
    Construct event narratives and preserve investigation history
- Support analytic workflow
  - Start with NUANCE overviews; when interesting events discovered, send extracts to Isis for detailed analysis.

Pacific Northwest
NATIONAL LABORATORY

NUANCE Traffic Circle
Scalable exploration of network flows

- Interactive visualization of patterns in high volume netflow data.

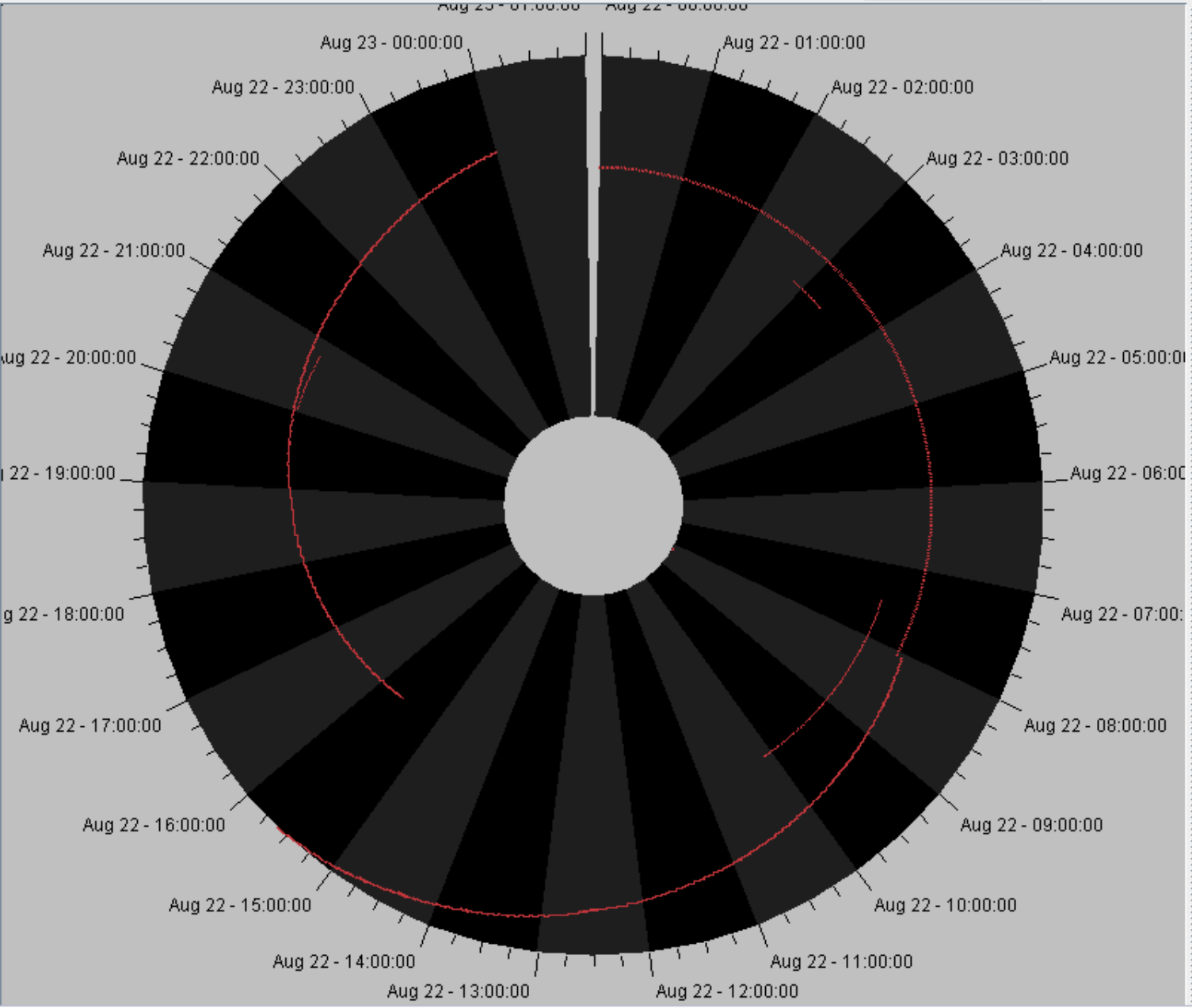


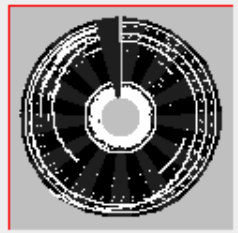
Pacific Northwest
NATIONAL LABORATORY

Nuance - Traffic Circle

File Edit View

Radius : Source Port 0 65535



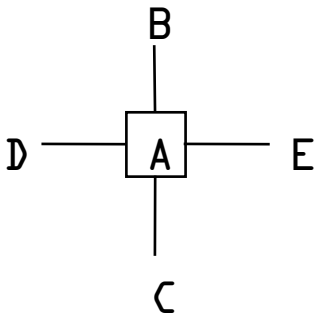


Destination Port : 6660 - 6669

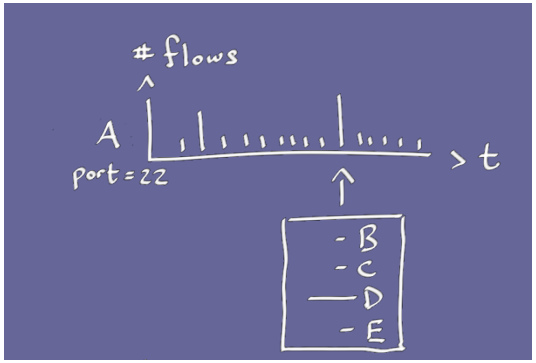
Destination Port : 6670 - , - 665

Isis
Using progressive multiples to explore flows

- **Progressive multiples…**
  - Make exploration history visible
  - Support backtracking
  - Allow rows to be reordered, revealing structure and event sequencing
  - Compare events of different nodes using time-oriented displays

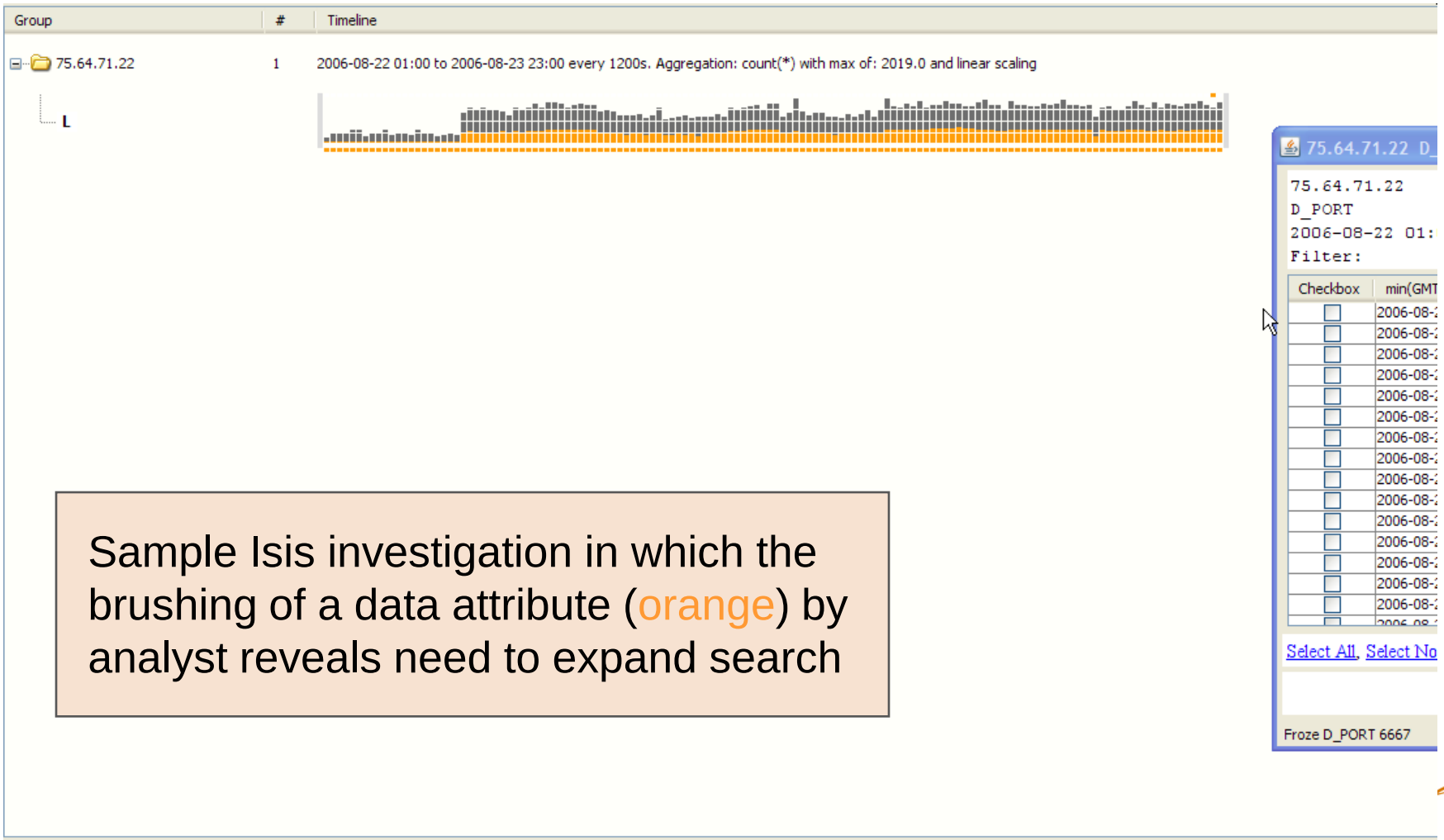


Traffic involving node A as node-link diagram



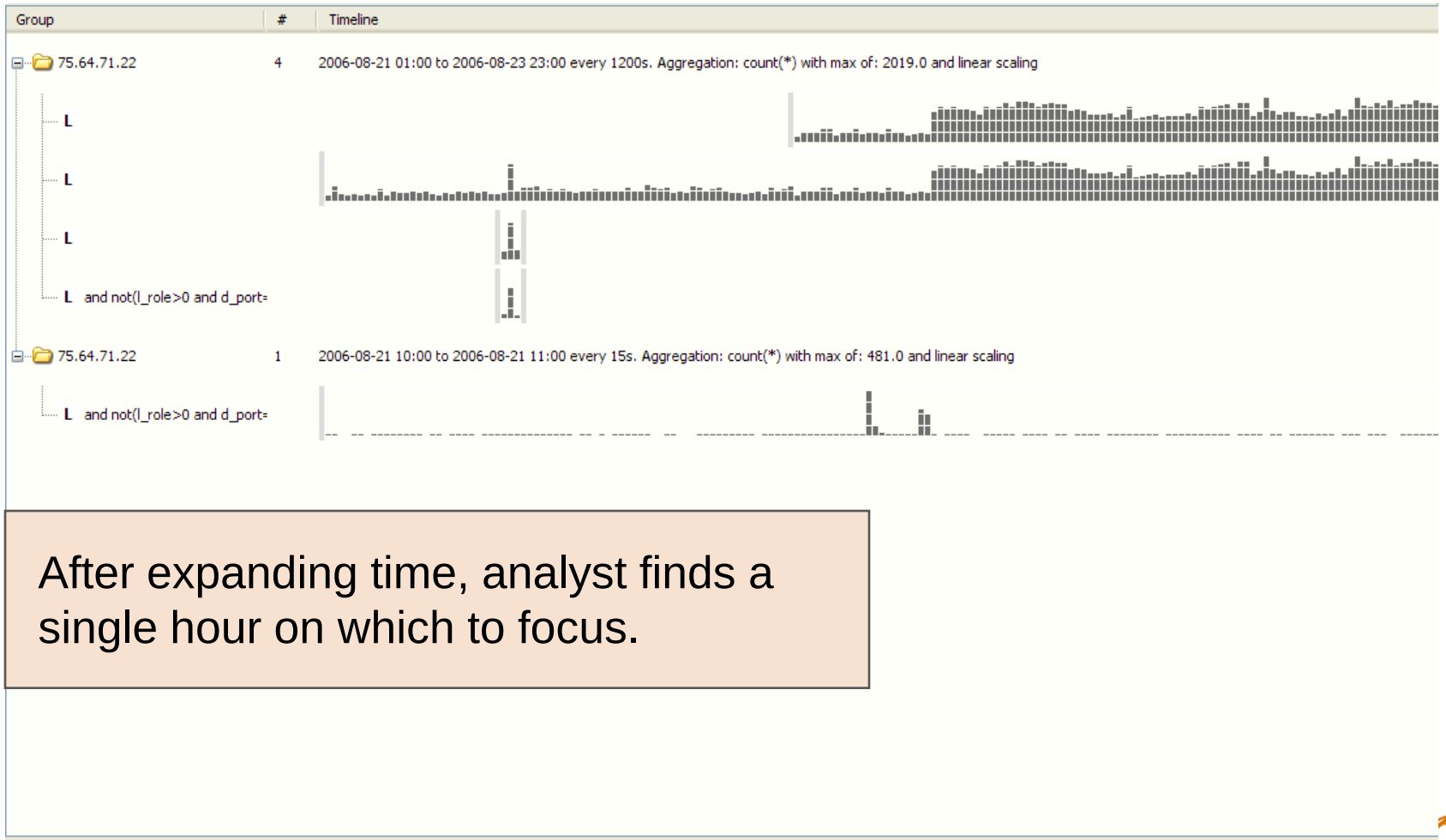
Traffic involving A as a timeline

Isis
Using progressive multiples to explore flows



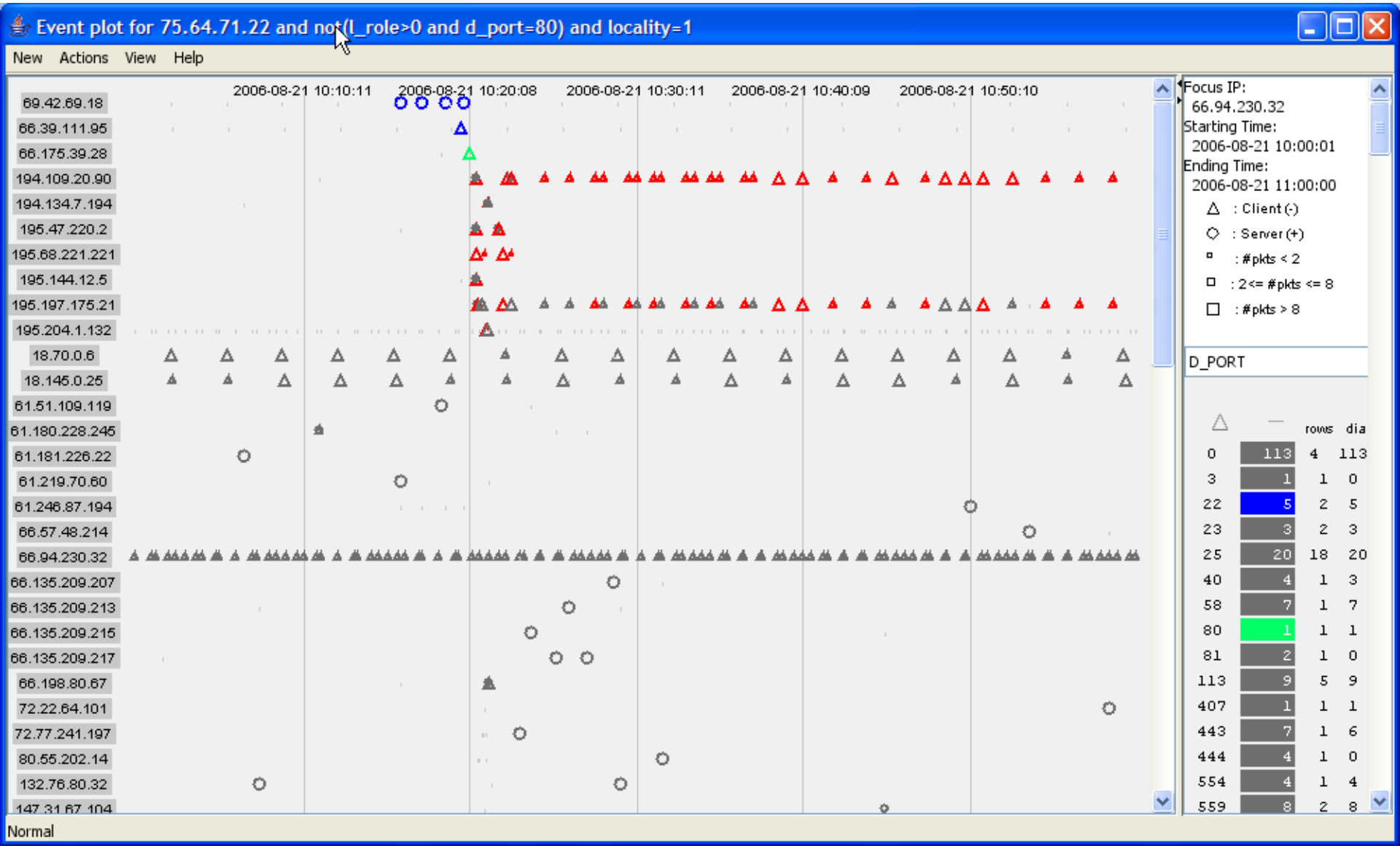
Sample Isis investigation in which the brushing of a data attribute (orange) by analyst reveals need to expand search

Isis
Using progressive multiples to explore flows



After expanding time, analyst finds a single hour on which to focus.

Event plots
Constructing a narrative



Evolution of tools for US-CERT

- Engage analysts in design reviews, requirements analyses.
- Adapt tools to EINSTEIN flow data by making them *schema agnostic*.
- Allow single tool sessions to visualize multiple tables in order to match existing data workflows.
- Allow timelines and event plots to use any attribute for an axis, not just network addresses.
- Simplify query panel inputs to improve productivity.
- Support analyst-defined calculations and control over panel contents.

Project Outcomes

- For researchers:
  - Understanding real-world workflows
  - Resources to turn tools into re-deployable production systems
- For practitioners:
  - Visual analytics becomes part of daily workflow
  - New ways of viewing data lead to better situational awareness
  - Quicker response time between alert and resolution
- Next steps:
  - Link visualization to modeling; there's only so much data you can visualize!
  - Understand the characteristic behaviors of machines on a network
  - Transition from reactive to proactive security posture

Pacific Northwest
NATIONAL LABORATORY