Visualizing Network Data for Intrusion Detection

Kulsoom Abdullah, Chris Lee,
Gregory Conti, John A. Copeland

June 16, 2005

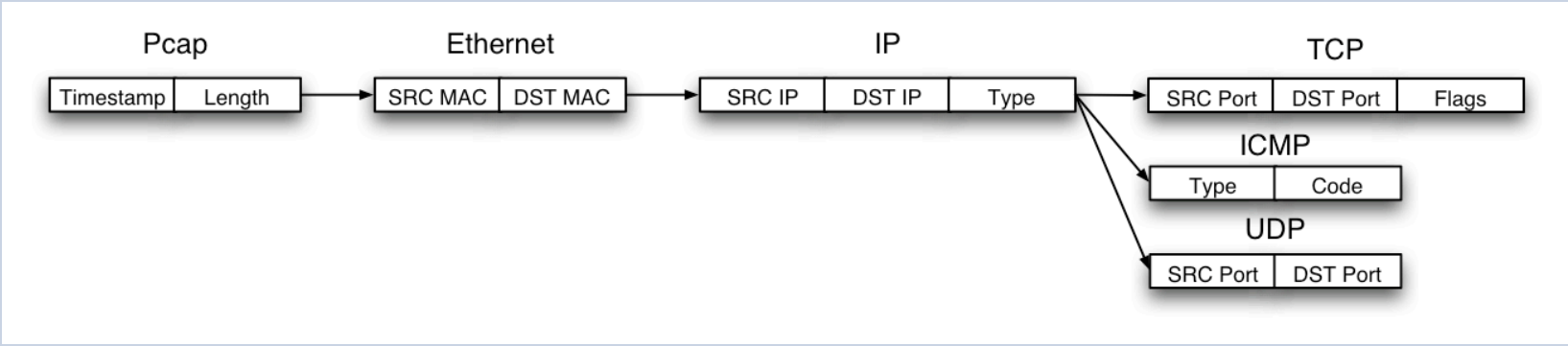
Motivation/Background

- Network traffic capacity is greater than systems can process.
- Network attacks have not decreased, current security tools are insufficient.
- Network attacks can be characterized by ports activity.
- Information visualization helps to provide insights and understands in datasets vs. just text alone.
- We want to provide an overview with details on demand.

Related Work

- FlowScan
  - Stacked area chart used to show NetFlow statistics.
- NVisionIP
  - Shows port flow count per IP address, separated by common and uncommon ports.
- PortVis
  - Different level of details shown in multiple views.
  - Matrix is used to show the entire port range.
  - Uses normalization methods to show variance.
- We use stacked histograms to show individual packet statistics for instantaneous results.
- We show packet count/byte over time where aggregate port activity is grouped and shown initially.
- Cube root normalization is used to show pattern and variance over time.

Packet Capture & Header fields



- Pcap timestamp, port numbers are used currently.
- Advantage is that packets are parsed online, where real time processing can be performed not having to wait for a flow to end.
- The payload is not checked, but with so much traffic data, processing each packet would overburden a monitoring system.

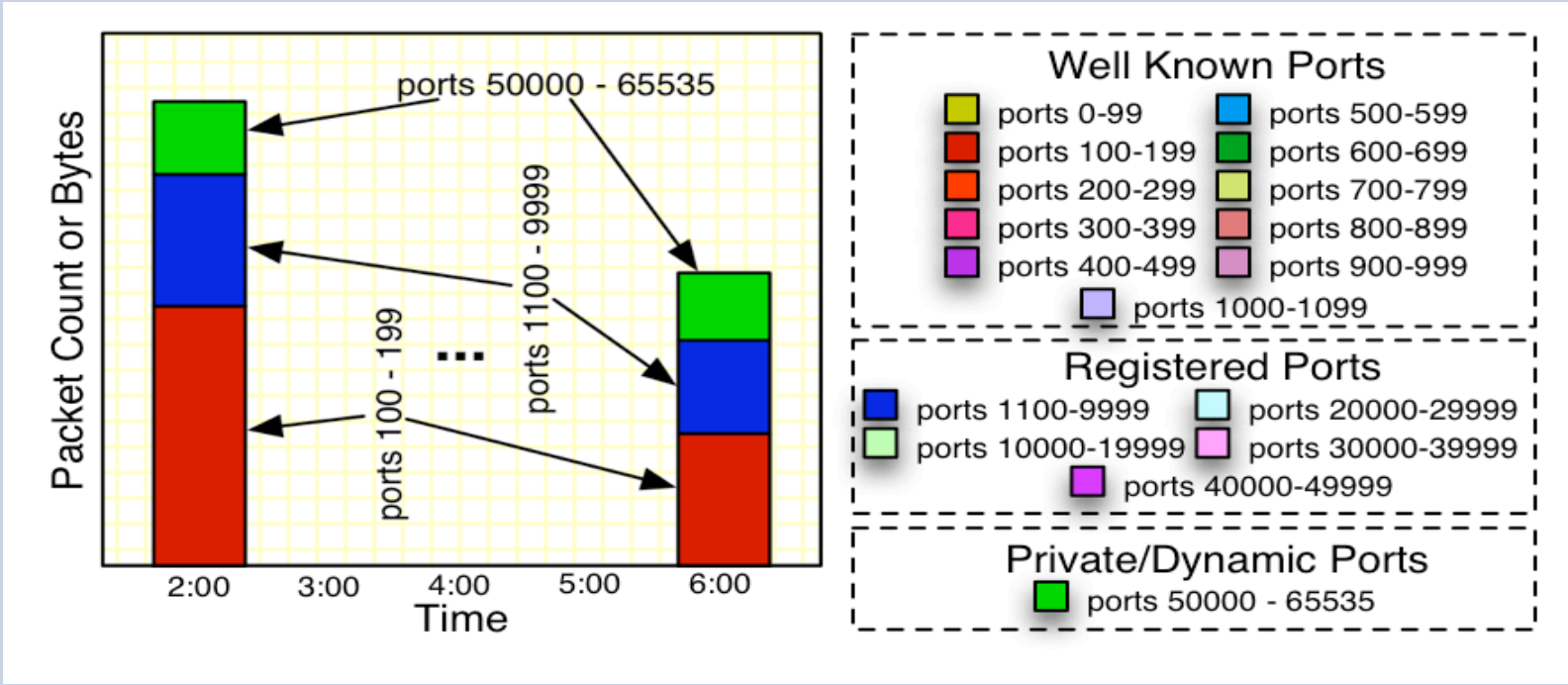
Forensic vs. Real time

- Browsing text logs for real time and forensic analysis is tedious.
- Real time is more challenging since it contains legitimate traffic in addition to malicious.
- Currently, we have used Honeynet traffic. Some techniques can be applied to real time traffic.

Histograms

- Histograms are easy to interpret and good for visualizing large datasets.
- Values can be compared relative to each other, which is useful in visualizing time patterns.
- For 3 variable plotting, we use 2D stacked, rather than 3D for less program complexity and for more accurate value interpretation.

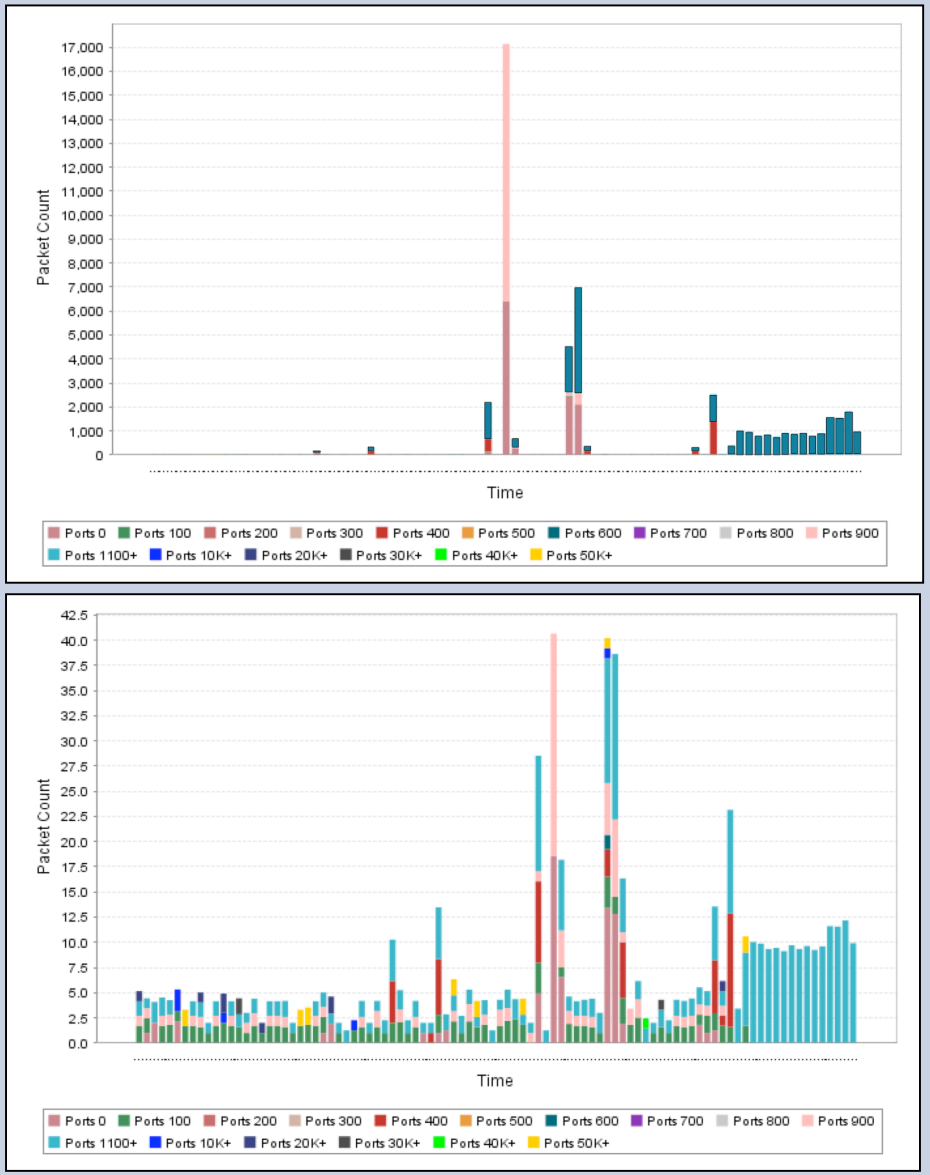
Concept Diagram



- X axis is time.
- Y axis is port count (interval).
- Port number (ordinal) ranges are grouped and mapped by color (more info in port scaling).

Graph Scaling

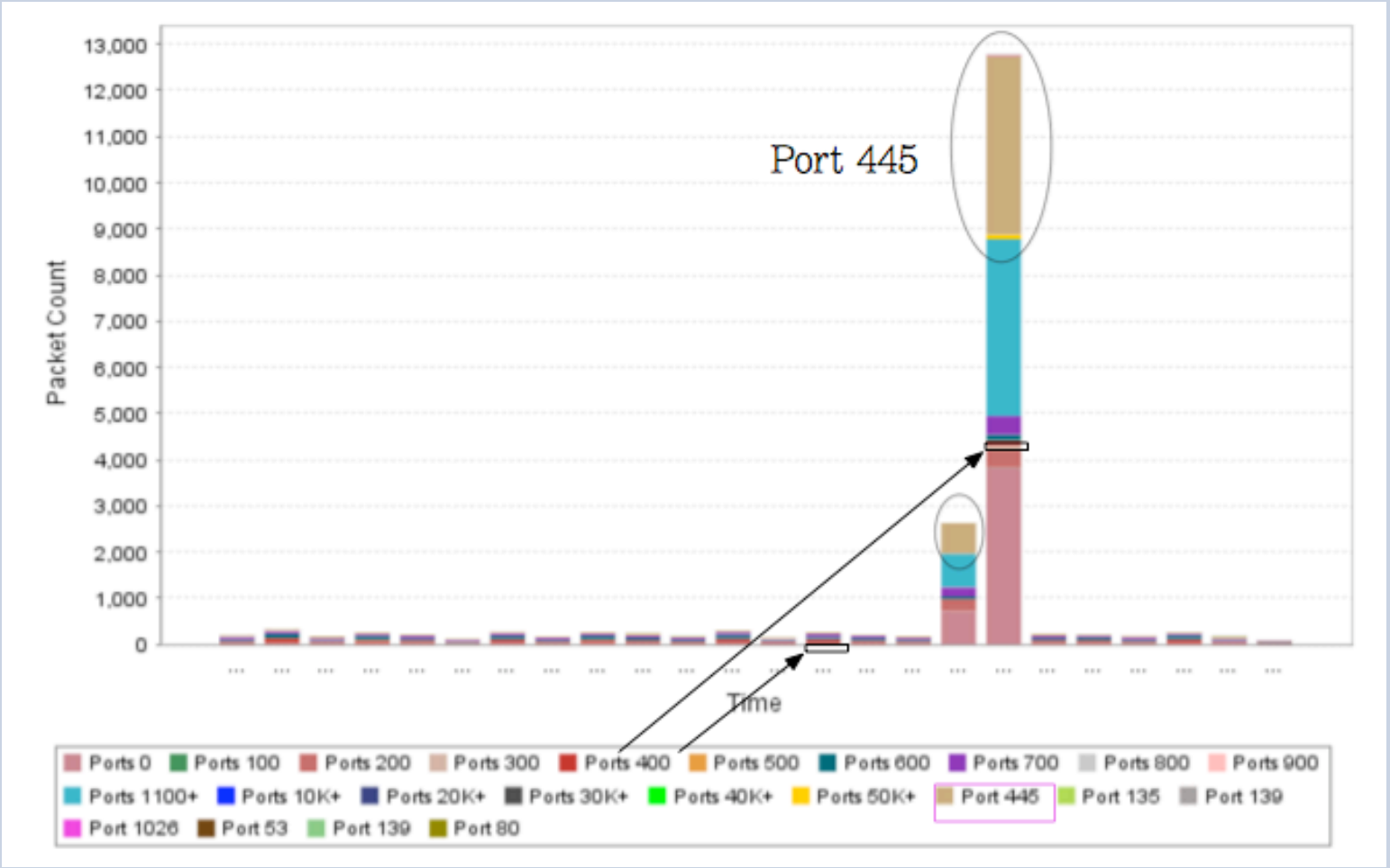
- Goals: Avoid overlap and occlusion.
- Network traffic statistics are highly variable, and high values can skew the scale.
- Cube root scales the range of values including zero with the additional benefit that values less than one, but greater than zero, are still mapped to positive values.
- Infovis methods can be used (filtering, zoom, and mouseovers).



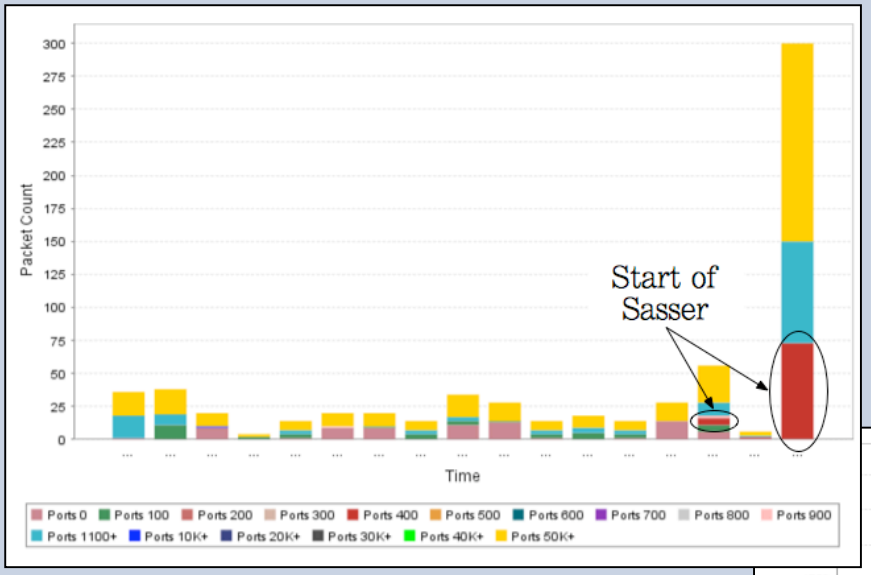
- Botnet traffic, real count (top) and normalized (bottom)

Port Scaling

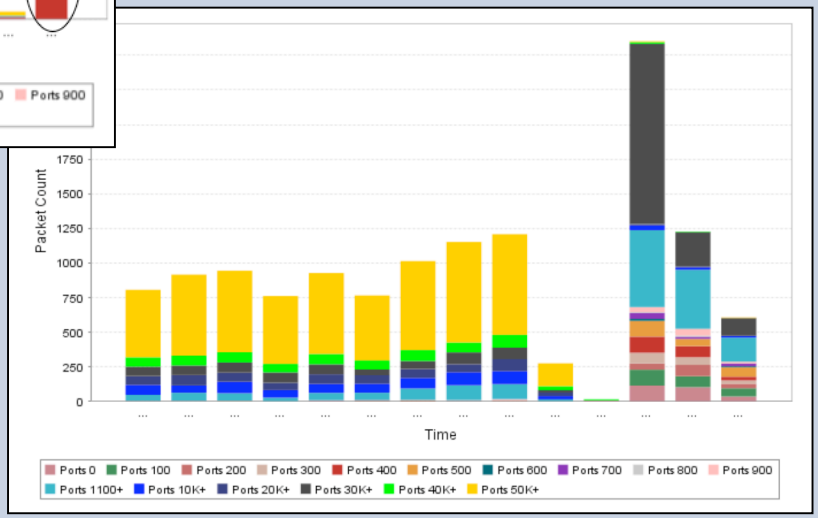
- 65,536 possible port numbers - can not allocate each number to one pixel.
- Ports have been grouped into ranges so we can fit the range on the graph.
- Well-known and commonly assigned ports - 0-1023, 100 in a group.
  - Most traffic here, also most attacks start with these ports.
- Registered ports 1024-49151, 10,000 in a group.
  - Can be used by an application or assigned for a connection attempt to a server.
  - Less traffic here than in well common ports.
- Private or dynamic ports 49152 - 65535
  - No service is typically assigned here.
  - These can still be used by malicious applications.



Ports 445, 135, 1026, 53, and 80 are separated. Common service ports would be separated in regular traffic networks. Here port 445 is circled, separating it from ports 400-499 (red).



Incoming sasser every 5 minutes



30 min. network scan, plotted every 2 min.

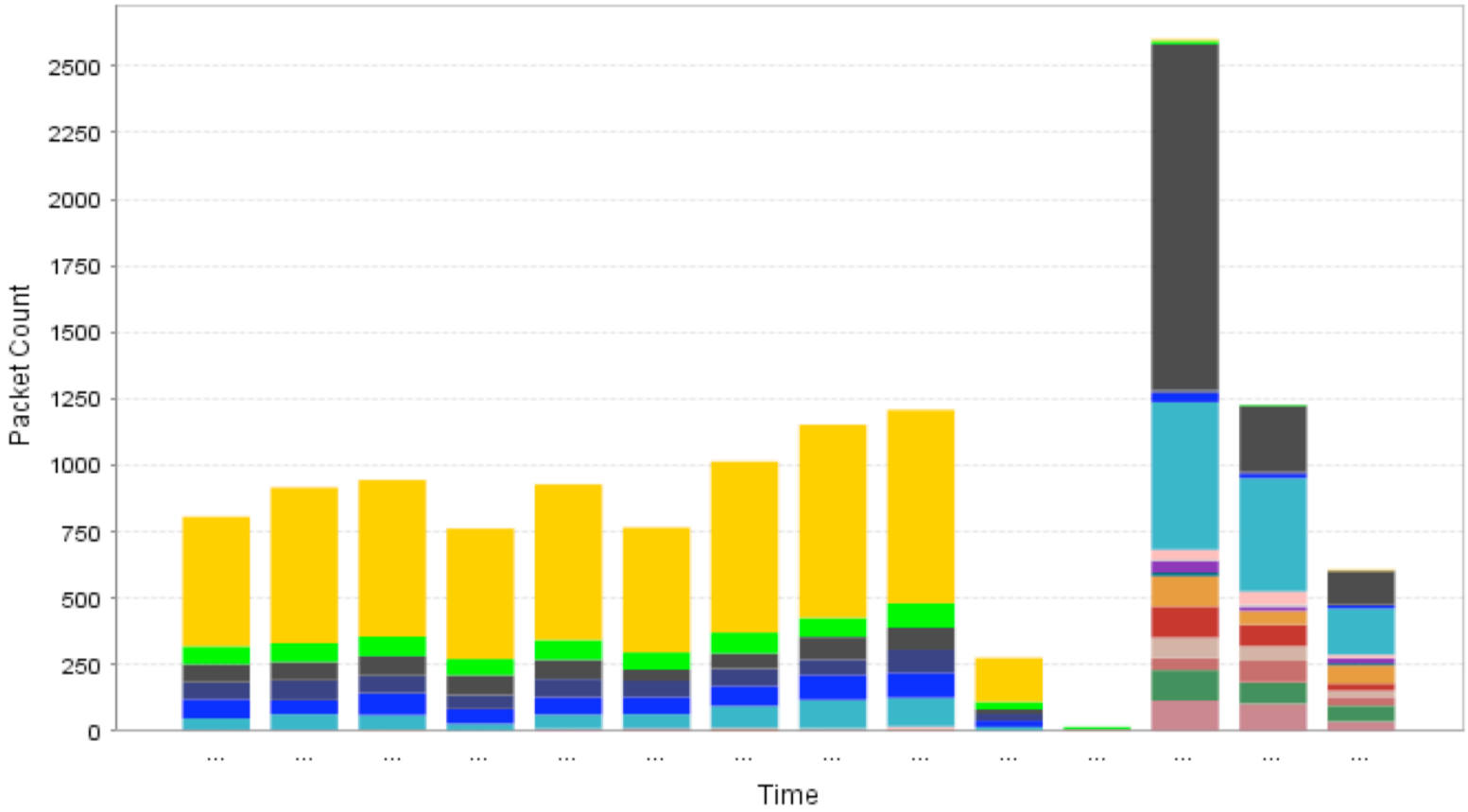
- A small time sample good for quick activities (fast network scans, DoS, fast propagating worms).
- A large time sample is better for viewing slow network scans, overall trends in a network over a long period of time.

IP Address Scaling

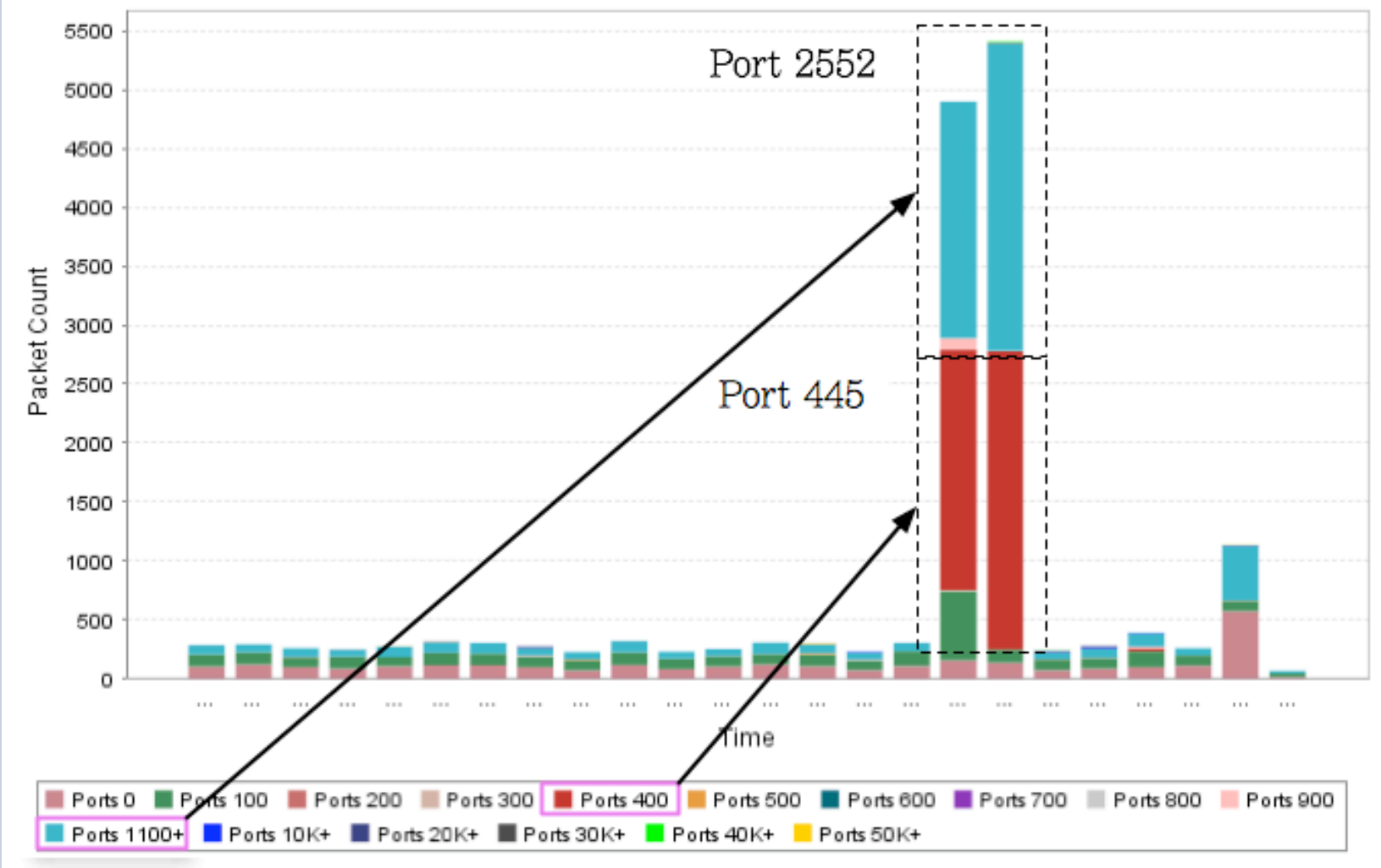
- 4 billion IPs total.
- Matrix method has been used in SnortView, NVisionIP.
- Filtering on hosts in VizFlowConnect.
- This is still a work in progress for future implementation.
  - Possible ideas are pivoting the axis, and highlighting.

Threat models

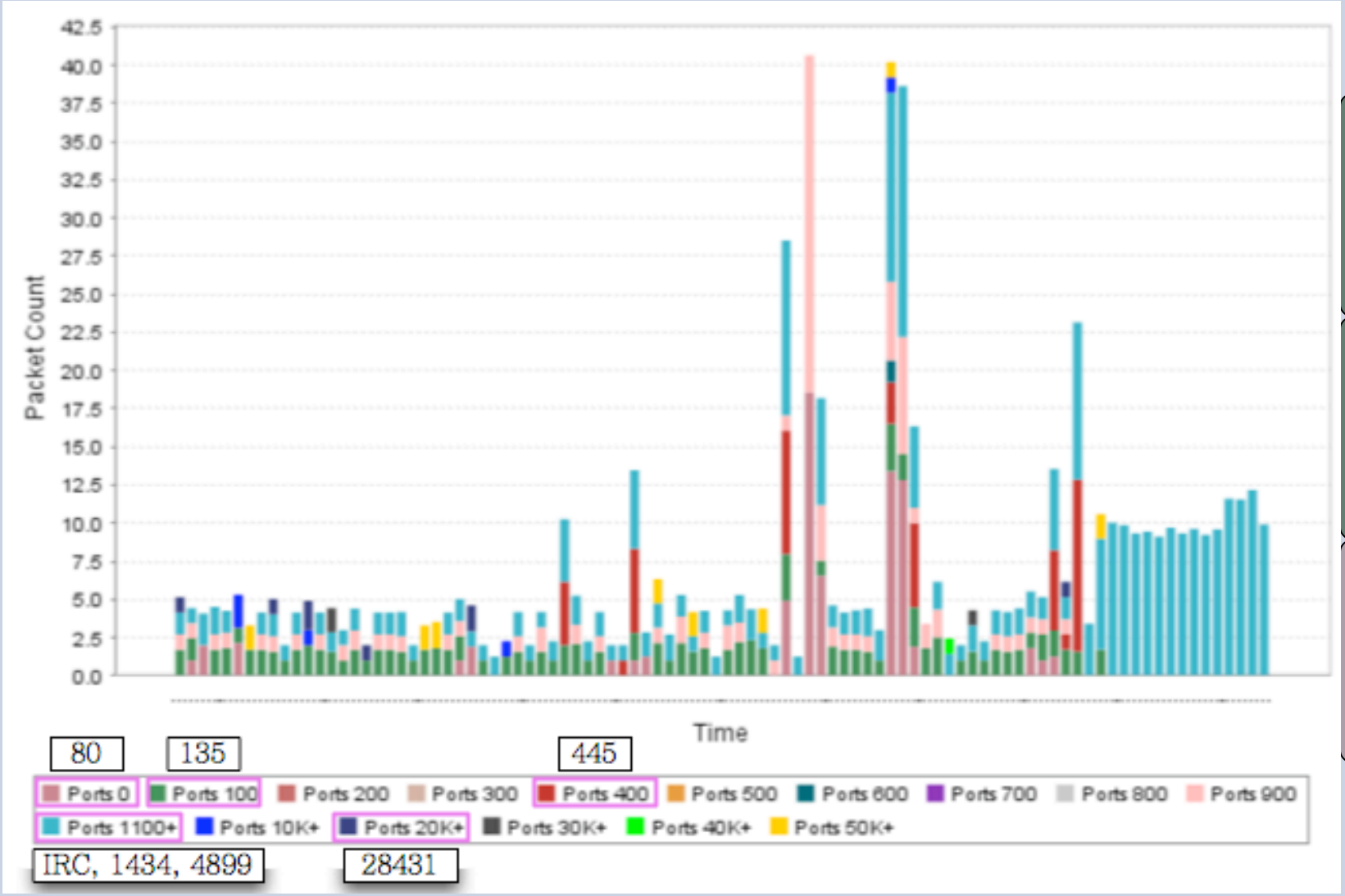
- These types of attacks were selected because they occur the most.
- The botnet capture can be representative of backdoor/trojan behavior.


Packet Count
2500
2250
2000
1750
1500
1250
1000
750
500
250
0
Time
Ports 0
Ports 100
Ports 200
Ports 300
Ports 400
Ports 500
Ports 600
Ports 700
Ports 800
Ports 900
Ports 1100+
Ports 10K+
Ports 20K+
Ports 30K+
Ports 40K+
Ports 50K+

Port Scan


Port 2552
Port 445
Packet Count
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Time
Ports 0
Ports 100
Ports 200
Ports 300
Ports 400
Ports 500
Ports 600
Ports 700
Ports 800
Ports 900
Ports 1100+
Ports 10K+
Ports 20K+
Ports 30K+
Ports 40K+
Ports 50K+

Sasser



Botnet with normalization

Conclusion

- Good for detecting malicious activities that affect ports.
- Gives an overview of all port usage on a network.
- Non-port based activity can not be detected.
  - Gaining root access.
- **Future Work**
  - Incorporate other header fields (e.g. ICMP, IP) for non port based attacks.
  - Implement more info vis methods, HCI.
  - Possibly incorporate the tool with multiple views of a network & other data (alarms, netflow).

Questions
Feedback

Contact:

{kulsoom, chris, copeland}@ece.gatech.edu

conti@cc.gatech.edu