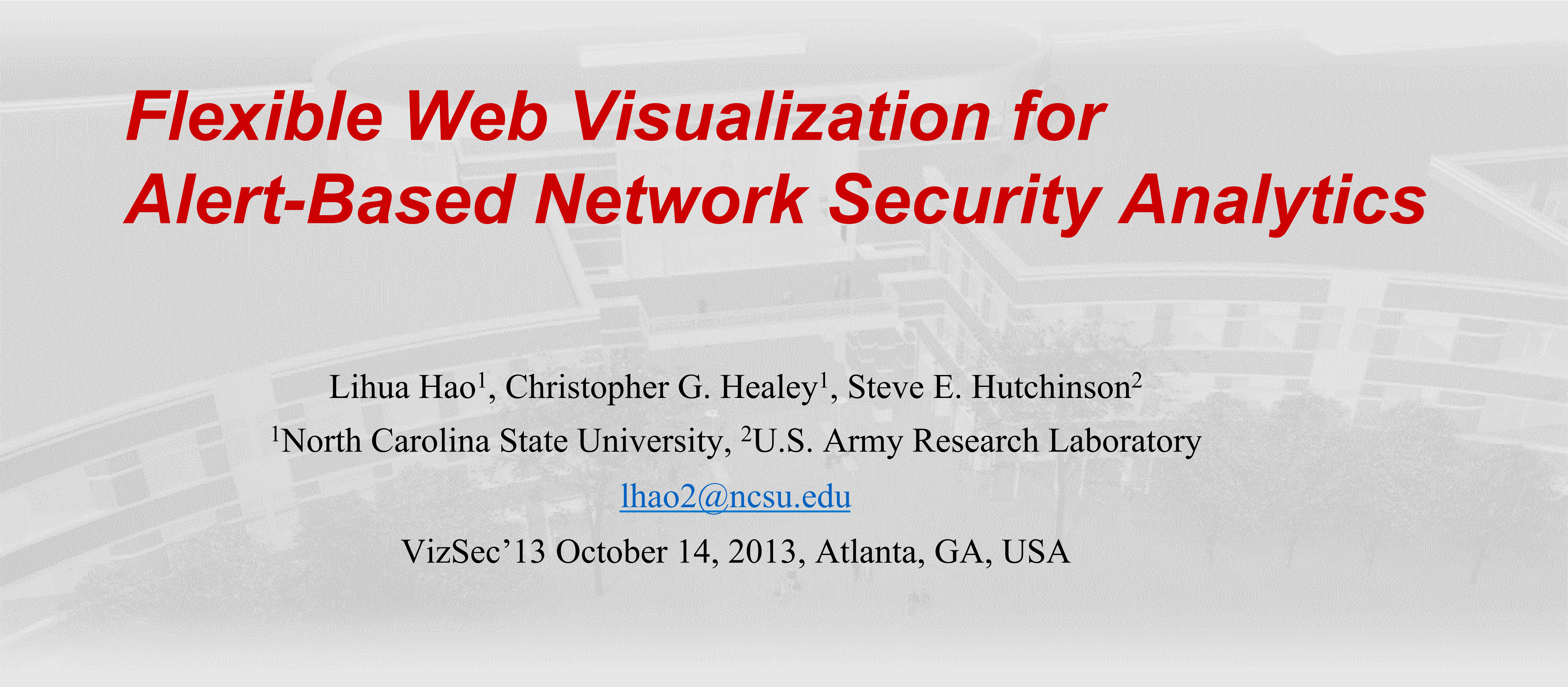


Flexible Web Visualization for Alert-Based Network Security Analytics

Lihua Hao[1], Christopher G. Healey[1], Steve E. Hutchinson[2]

[1]North Carolina State University, [2]U.S. Army Research Laboratory

lhao2@ncsu.edu

VizSec'13 October 14, 2013, Atlanta, GA, USA

Introduction

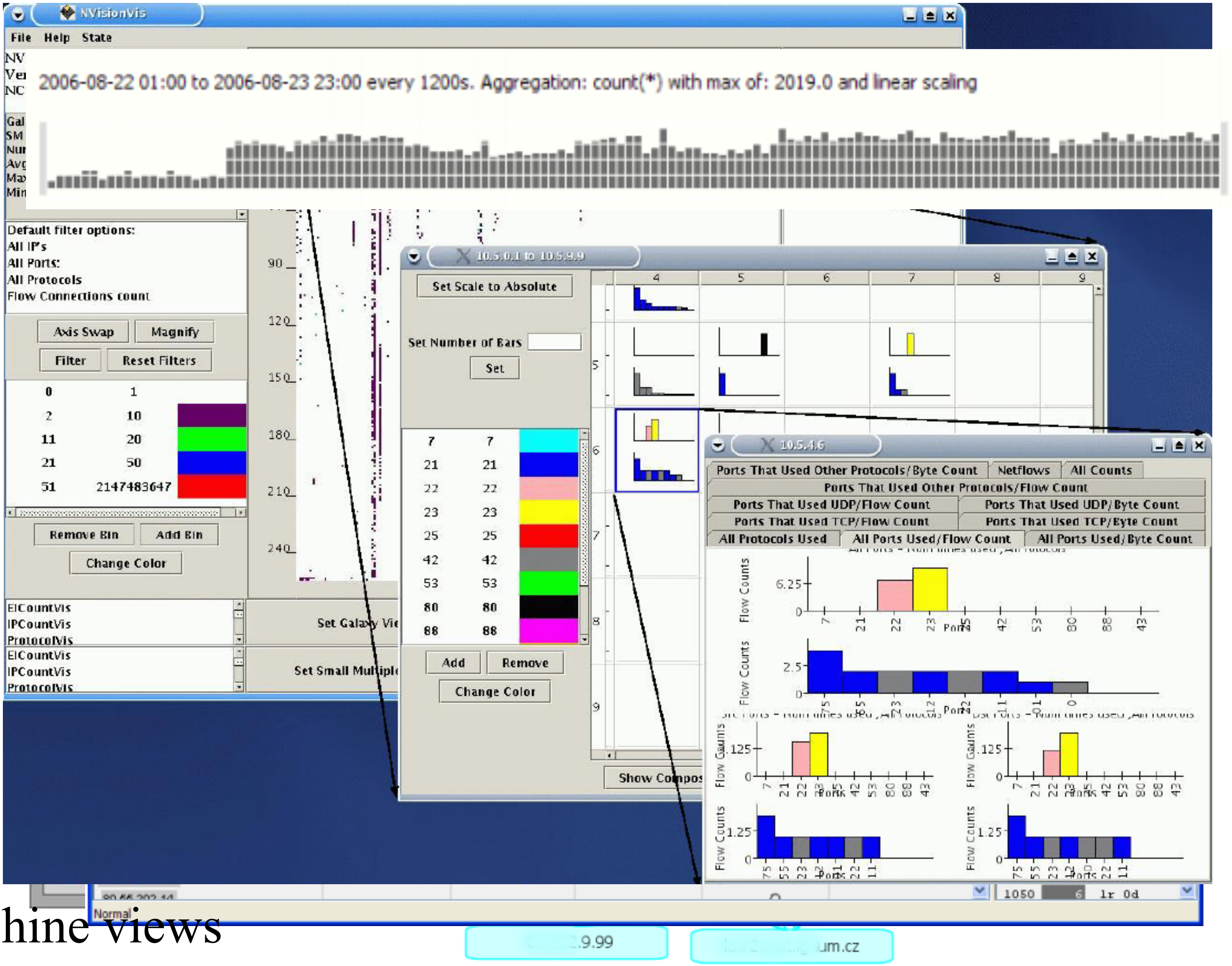
- Building a visualization tool for Army Research Laboratory (ARL) network security analysts
- Driven by analysts
  - Our approach does not focus explicitly on network security ***data***, but rather on network security ***analysts***
  - ***"Don't fit our problem to your tool. Build a tool to fit our problem."***
- We must balance
  1. Meeting needs of the analysts.
  2. Applying knowledge and best practices from visualization.

Design Constraints

1. Mental models
   - "Fit" the mental models the analysts use to investigate problems
2. Working environment
   - Integrate into the analyst's current working environment (web browser for ARL analysts)
3. Configurability
   - Static, pre-defined presentations of the data are typically not sufficient
4. Accessibility
   - The visualizations should be familiar to analysts, to avoid steep learning curves
5. Scalability
   - Support query and retrieval from large data sources
6. Integration
   - Augment the analyst's current problem-solving strategies with useful support

Existing Visualization Techniques

- Node-link graphs
  - Portall, HoNe, LinkRank
- Treemaps
  - NetVis, NFlowVis
- Timelines and Event Plots
  - Aggregate value over events
  - Capture patterns of individual events
- Basic Charts
  - Snorby, NVisionIP
- Zooming, Multivariate
  - NVisionIP: galaxy, small multiple, and machine views
  - VisFlowConnect: global, domain, internal, and host views

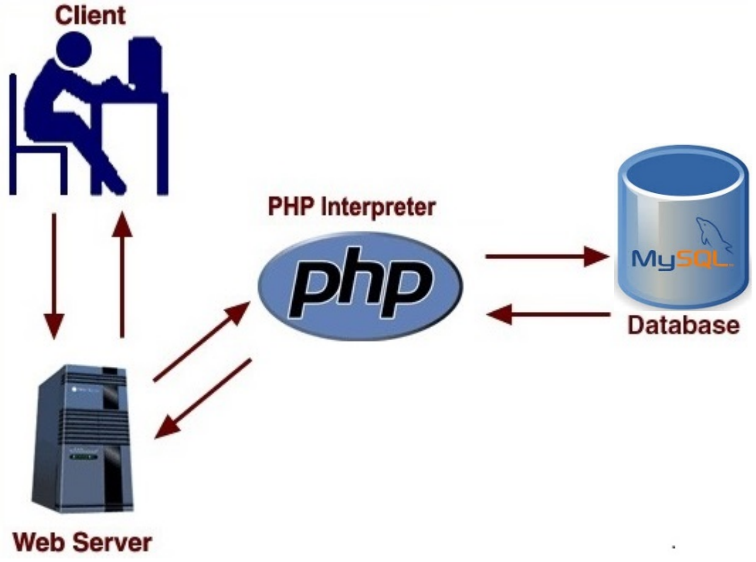


2006-08-22 01:00 to 2006-08-23 23:00 every 1200s. Aggregation: count(*) with max of: 2019.0 and linear scaling

Default filter options:
All IP's
All Ports:
All Protocols
Flow Connections count

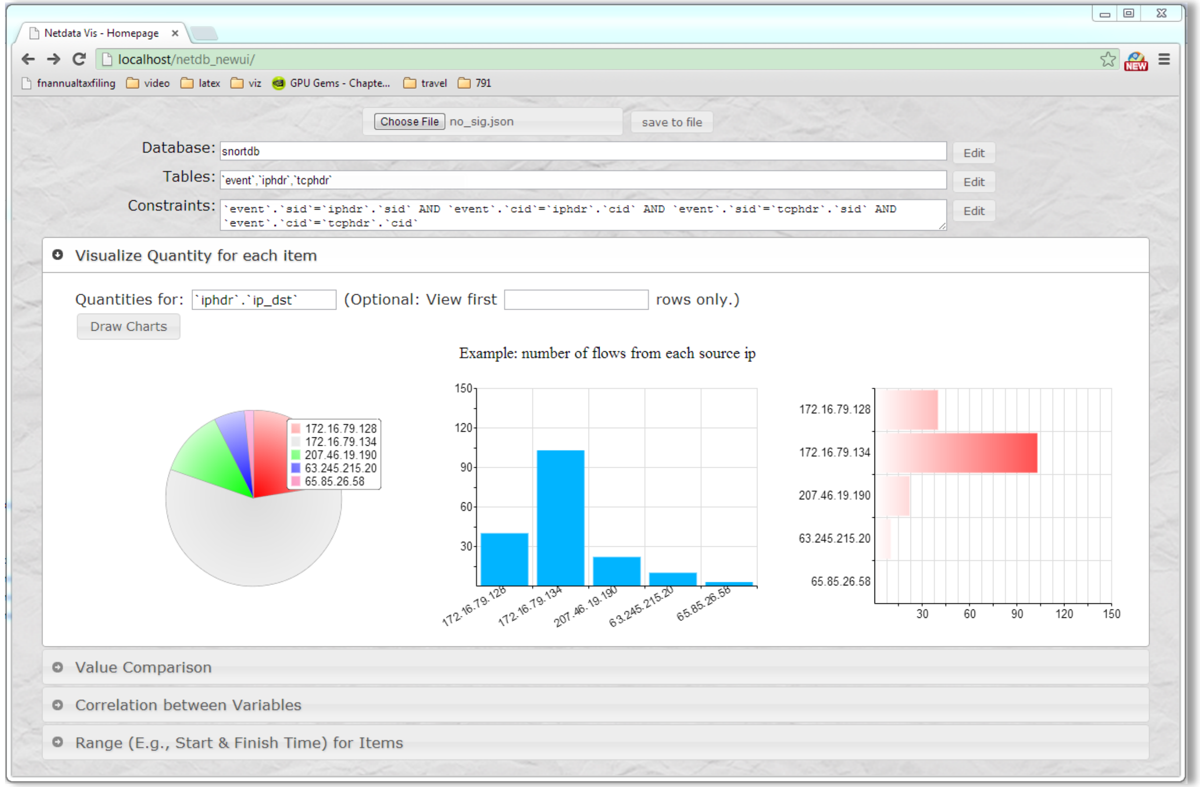
Data Management

- MySQL & PHP running on a remote server
  - Provide reasonable *scalability*, efficient data filtering and projection
- No pre-defined table formats
  - Analysts choose columns to visualize, define table correlations and data filtering
  - Provide flexibility and *configurability*
- Cache results of current query in memory
  - Generate queries to retrieve the new data on demand
- Full SQL is available on demand to the analyst
  - System suggests visualization with automatically generated SQL queries
  - Analysts can manually *configure* system suggestions



Web-Based Visualization

- ARL analysts work in a browser
  - “Fit” analysts’ *working environment*
- HTML5 canvas element
  - No external plug-ins required
  - Run in any modern web browser
- Use 2D charts
  - Common in other security visualization systems
  - Effective for presenting values, trends, patterns and relationships our analysts want to explore
  - Provides *accessibility*



Netdata Vis - Homepage

localhost/netdb_newui/

fnannualtaxfiling | video | latex | viz | GPU Gems - Chapte... | travel | 791

Choose File no_sig.json | save to file

Database: snortdb

Tables: `event`,`iphdr`,`tcphdr`

Constraints: `event`.`sid`=`iphdr`.`sid` AND `event`.`cid`=`iphdr`.`cid` AND `event`.`sid`=`tcphdr`.`sid` AND `event`.`cid`=`tcphdr`.`cid`

Visualize Quantity for each item

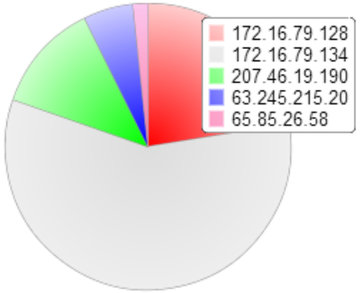
Quantities for: `iphdr`.`ip_dst` (Optional: View first rows only.)

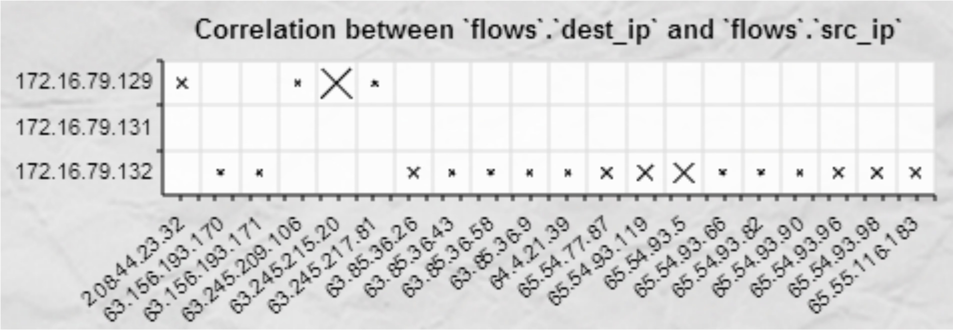
Draw Charts

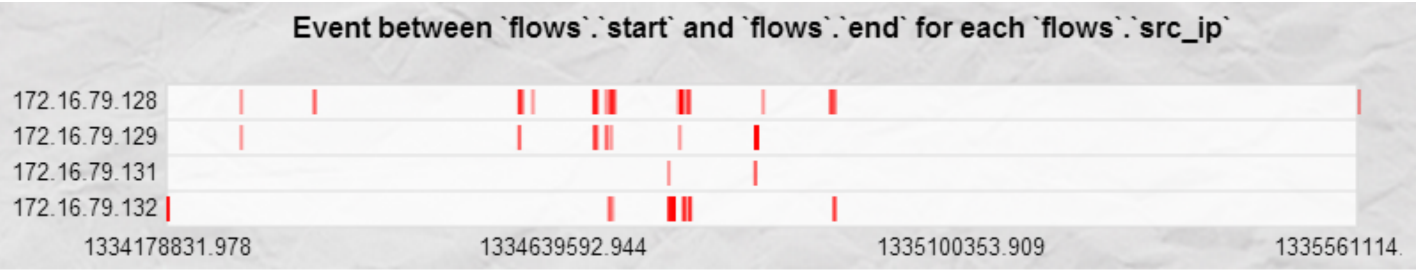
Example: number of flows from each source ip

Analyst-Driven Charts

- RGraph for basic chart visualizations
  - Open source library for visualization with 2D charts
  - Choose charts commonly used in network data visualization
- Assisted chart selection based on data and task (*accessibility*)
  - Pie/bar: proportion and frequency comparison
  - Bar: value comparison over a secondary attribute
  - Scatterplots: correlation between two attributes
  - Gantt: range value comparison
- Suggested chart properties
  - Backgrounds, grids, glyph size, color and type
- Free to change the initial choices

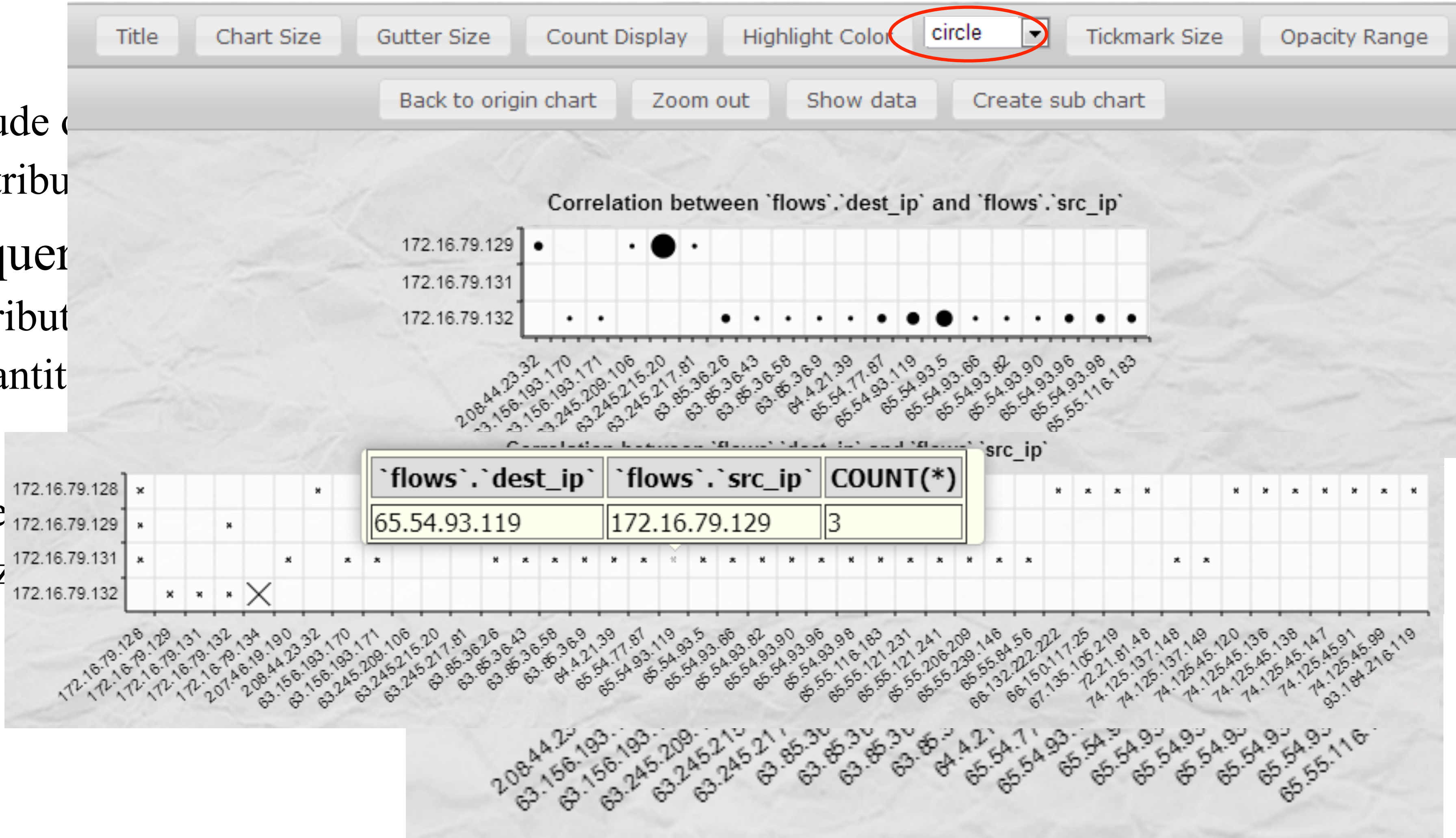






Interaction

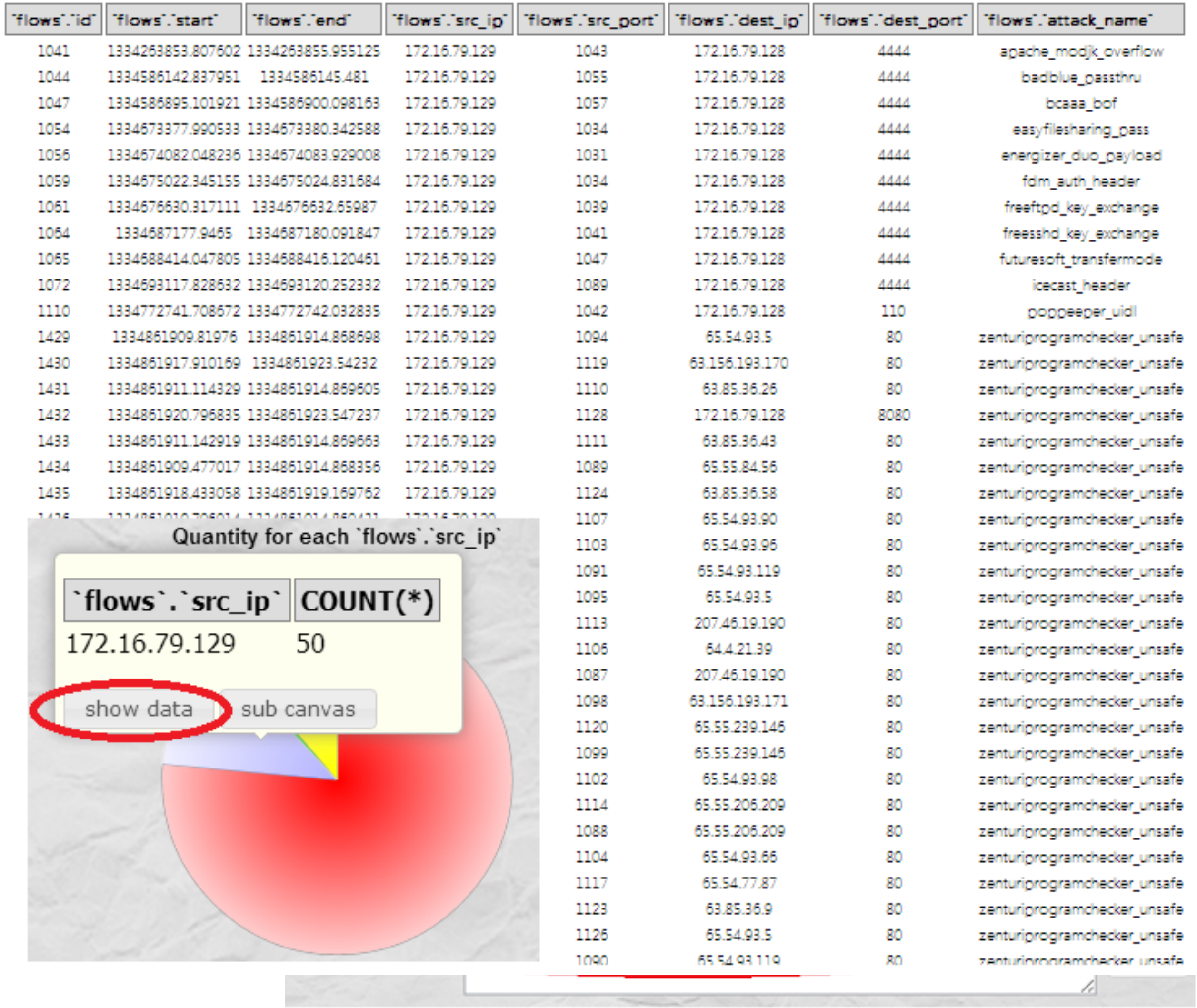
- Intelligent zoom
  - Redraw chart to include d
  - Rescale the visual attribu
- Tooltips for value quer
  - Display the exact attribut
  - Provide access to quantit
- Toolbars
  - Customize glyph size
  - Change chart title, siz



`flows`.`dest_ip`	`flows`.`src_ip`	COUNT(*)
65.54.93.119	172.16.79.129	3

Correlated Views

- A sequence of visualizations to track a
  - Correlate multiple data sources
  - Explore data at multiple levels of details
- Correlated charts
  - Select sub-regions of a chart as input for a f
  - Generate constraints to extract data of intere
  - Add additional constraints, tables, or attribu
- Raw data spreadsheets for value exam
  - Text-based examination: a conventional app
  - “Fit” the analyst’s *working environment, me*



`flows`.`id`	`flows`.`start`	`flows`.`end`	`flows`.`src_ip`	`flows`.`src_port`	`flows`.`dest_ip`	`flows`.`dest_port`	`flows`.`attack_name`
1041	1334263853.807602	1334263855.955125	172.16.79.129	1043	172.16.79.128	4444	apache_modjk_overflow
1044	1334586142.837951	1334586145.481	172.16.79.129	1055	172.16.79.128	4444	badblue_passthru
1047	1334586895.101921	1334586900.098163	172.16.79.129	1057	172.16.79.128	4444	bcaaa_bof
1054	1334673377.990533	1334673380.342588	172.16.79.129	1034	172.16.79.128	4444	easyfilesharing_pass
1056	1334674082.048236	1334674083.929008	172.16.79.129	1031	172.16.79.128	4444	energizer_duo_payload
1059	1334675022.345155	1334675024.831684	172.16.79.129	1034	172.16.79.128	4444	fdm_auth_header
1061	1334676630.317111	1334676632.65987	172.16.79.129	1039	172.16.79.128	4444	freeftpd_key_exchange
1064	1334687177.9465	1334687180.091847	172.16.79.129	1041	172.16.79.128	4444	freesshd_key_exchange
1065	1334688414.047805	1334688416.120461	172.16.79.129	1047	172.16.79.128	4444	futuresoft_transfermode
1072	1334693117.828632	1334693120.252332	172.16.79.129	1089	172.16.79.128	4444	icecast_header
1110	1334772741.708672	1334772742.032835	172.16.79.129	1042	172.16.79.128	110	poppeeper_uidl
1429	1334861909.81976	1334861914.868698	172.16.79.129	1094	65.54.93.5	80	zenturiprogramchecker_unsafe
1430	1334861917.910169	1334861923.54232	172.16.79.129	1119	63.156.193.170	80	zenturiprogramchecker_unsafe
1431	1334861911.114329	1334861914.869605	172.16.79.129	1110	63.85.36.26	80	zenturiprogramchecker_unsafe
1432	1334861920.796835	1334861923.547237	172.16.79.129	1128	172.16.79.128	8080	zenturiprogramchecker_unsafe
1433	1334861911.142919	1334861914.869663	172.16.79.129	1111	63.85.36.43	80	zenturiprogramchecker_unsafe
1434	1334861909.477017	1334861914.868356	172.16.79.129	1089	65.55.84.56	80	zenturiprogramchecker_unsafe
1435	1334861918.433058	1334861919.169762	172.16.79.129	1124	63.85.36.58	80	zenturiprogramchecker_unsafe
[illegible]	[illegible]	[illegible]	[illegible]	1107	65.54.93.90	80	zenturiprogramchecker_unsafe
				1103	65.54.93.96	80	zenturiprogramchecker_unsafe
				1091	65.54.93.119	80	zenturiprogramchecker_unsafe
				1095	65.54.93.5	80	zenturiprogramchecker_unsafe
				1113	207.46.19.190	80	zenturiprogramchecker_unsafe
				1106	64.4.21.39	80	zenturiprogramchecker_unsafe
				1087	207.46.19.190	80	zenturiprogramchecker_unsafe
				1098	63.156.193.171	80	zenturiprogramchecker_unsafe
				1120	65.55.239.146	80	zenturiprogramchecker_unsafe
				1099	65.55.239.146	80	zenturiprogramchecker_unsafe
				1102	65.54.93.98	80	zenturiprogramchecker_unsafe
				1114	65.55.206.209	80	zenturiprogramchecker_unsafe
				1088	65.55.206.209	80	zenturiprogramchecker_unsafe
				1104	65.54.93.66	80	zenturiprogramchecker_unsafe
				1117	65.54.77.87	80	zenturiprogramchecker_unsafe
				1123	63.85.36.9	80	zenturiprogramchecker_unsafe
				1126	65.54.93.5	80	zenturiprogramchecker_unsafe
				1090	65.54.93.119	80	zenturiprogramchecker_unsafe

Quantity for each `flows`.`src_ip`

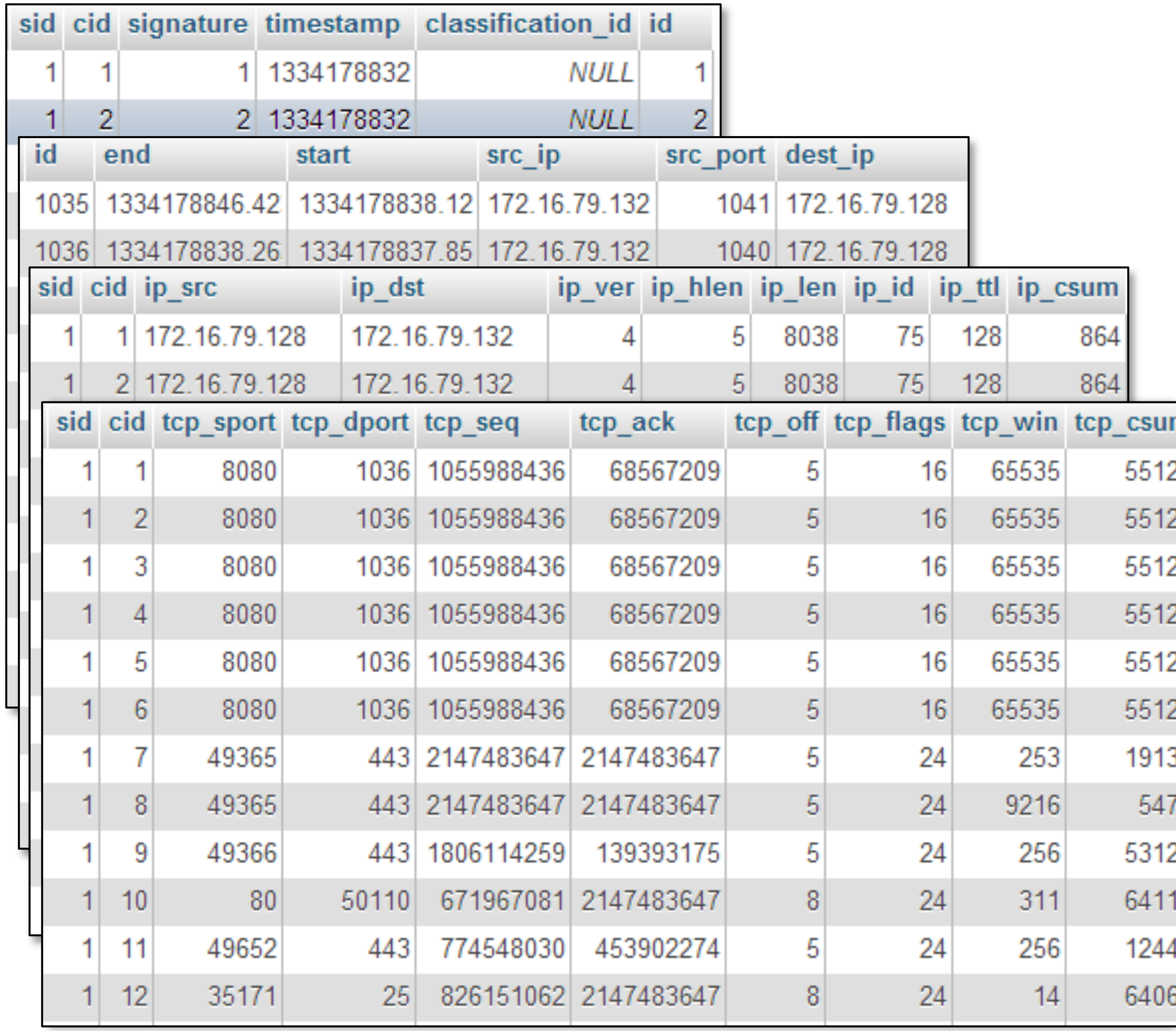
`flows`.`src_ip`	COUNT(*)
172.16.79.129	50

Trap Data

- Need real world data to test the system
- For security reasons, not possible to use data from ARL for testing
- The trap server
  - Data collected by network security researchers at NCSU
  - Real world network traffic in Computer Science building
  - Transmitted to a Snort sensor to perform: (1) intrusion detection and (2) extraction of network packets
  - Stores two types of data: (1) NetFlow data and (2) Snort alerts
- An example file for 24 hours of data
  - 17.4GB of packet headers
  - 938K unique source IPs, 168K unique destination IPs
  - 1.6M flows with 615K alerts

Example Tables

- Tables queried in the visualization
  - **event**: alert signature id and timestamp
  - **flows**: network flow sources and destination IP, port, start and end time
  - **iphdr**: source and destination IP and other information of packet headers
  - **tcphdr**: TCP related information such as source and destination port
- One of our research colleagues acted as the “analyst” in our scenario



sid	cid	signature	timestamp	classification_id	id
1	1	1	1334178832	NULL	1
1	2	2	1334178832	NULL	2

id	end	start	src_ip	src_port	dest_ip
1035	1334178846.42	1334178838.12	172.16.79.132	1041	172.16.79.128
1036	1334178838.26	1334178837.85	172.16.79.132	1040	172.16.79.128

sid	cid	ip_src	ip_dst	ip_ver	ip_hlen	ip_len	ip_id	ip_ttl	ip_csum
1	1	172.16.79.128	172.16.79.132	4	5	8038	75	128	864
1	2	172.16.79.128	172.16.79.132	4	5	8038	75	128	864

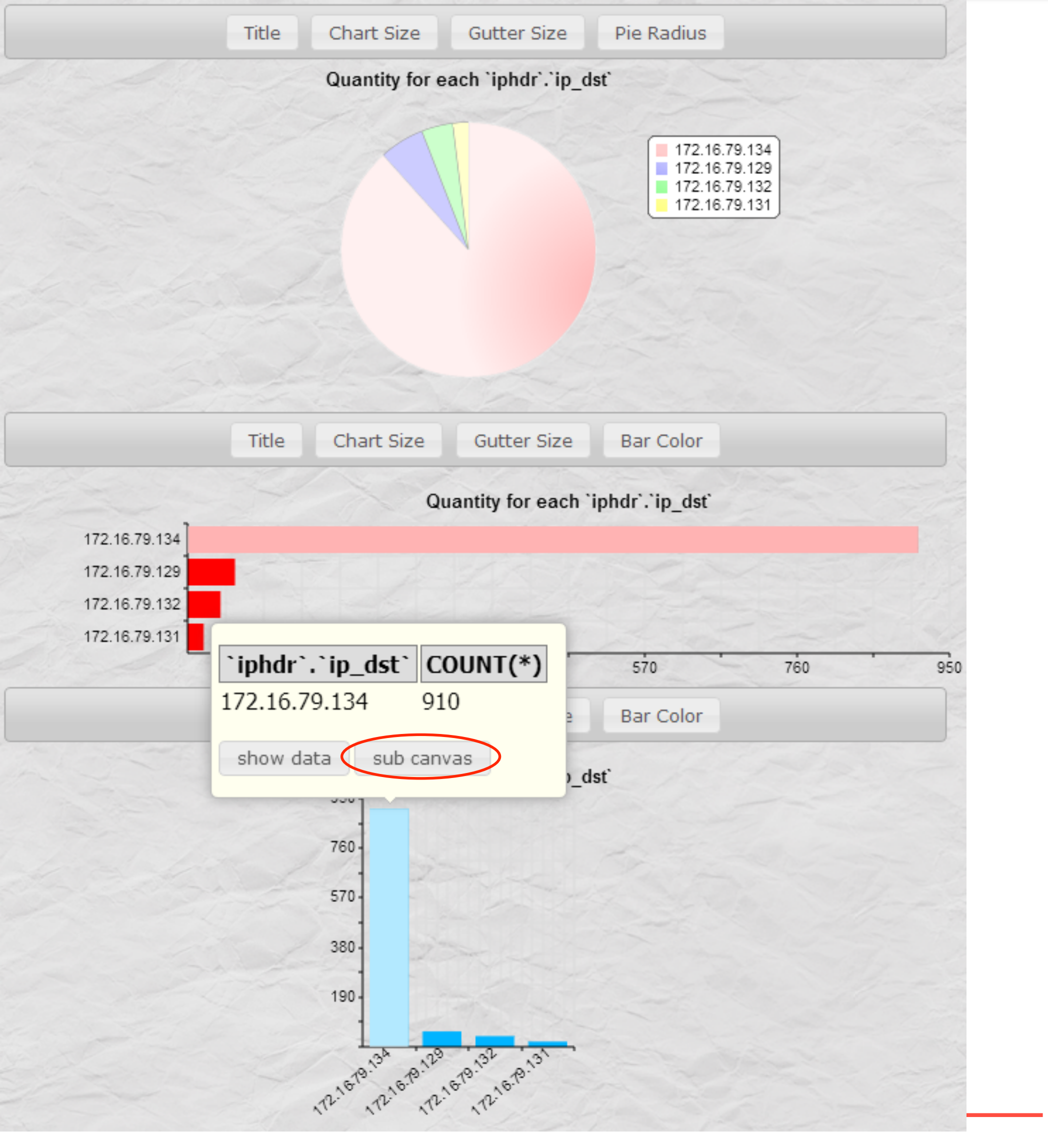
sid	cid	tcp_sport	tcp_dport	tcp_seq	tcp_ack	tcp_off	tcp_flags	tcp_win	tcp_csu
1	1	8080	1036	1055988436	68567209	5	16	65535	5512
1	2	8080	1036	1055988436	68567209	5	16	65535	5512
1	3	8080	1036	1055988436	68567209	5	16	65535	5512
1	4	8080	1036	1055988436	68567209	5	16	65535	5512
1	5	8080	1036	1055988436	68567209	5	16	65535	5512
1	6	8080	1036	1055988436	68567209	5	16	65535	5512
1	7	49365	443	2147483647	2147483647	5	24	253	1913
1	8	49365	443	2147483647	2147483647	5	24	9216	547
1	9	49366	443	1806114259	139393175	5	24	256	5312
1	10	80	50110	671967081	2147483647	8	24	311	6411
1	11	49652	443	774548030	453902274	5	24	256	1244
1	12	35171	25	826151062	2147483647	8	24	14	6406

Aggregate Alerts on Destination IPs

- Visualize number of alerts for each destination IP
- Pie chart, proportion of alerts by destination IP
- Bar chart, absolute numbers of alerts by destination IP

- The majority of the alerts are sent to destination IP 172.16.79.134

- “Sub Canvas” in the tooltip to create correlated chart for target destination IP



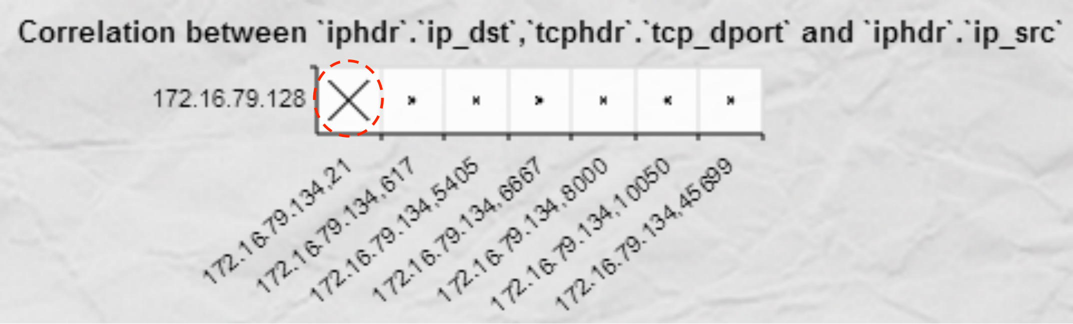
`iphdr`.`ip_dst`	COUNT(*)
172.16.79.134	910

Focus on High-Alert Destination IP

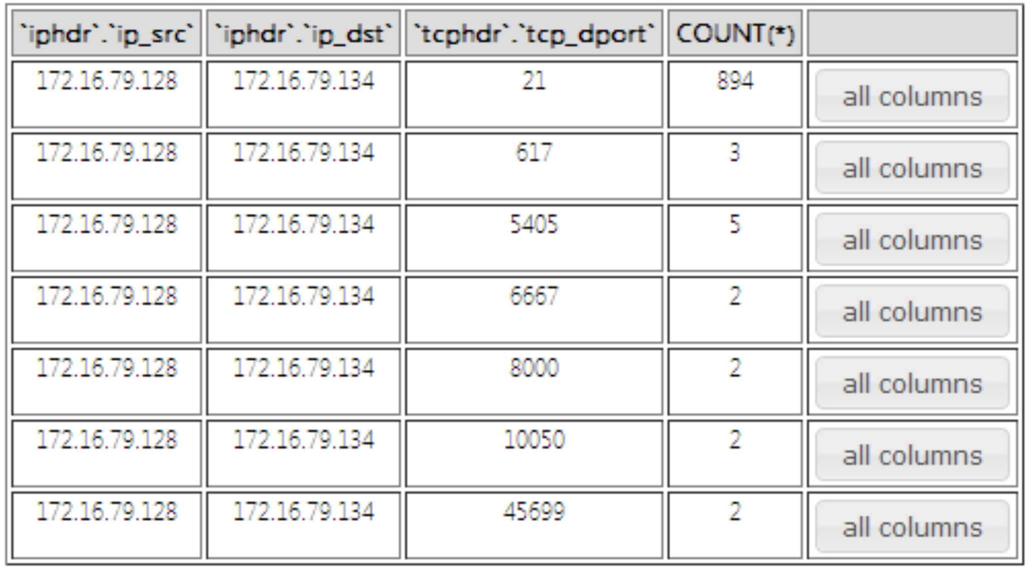
- Focus on the destination IP with the maximum number of alerts (i.e., 172.16.79.134)
- Scatterplot of an analyst-chosen source IP versus the target destination IP and port

- Sizes of scatterplot glyphs indicate number of alerts from the source to the destination/port
- Analyst requests a text table detailing the exact IPs, ports, and alert counts

- Most alerts are sent to port 21 (894 alerts), so follow-on analysis will focus on this port



Correlation between `iphdr`.`ip_dst`,`tcphdr`.`tcp_dport` and `iphdr`.`ip_src`

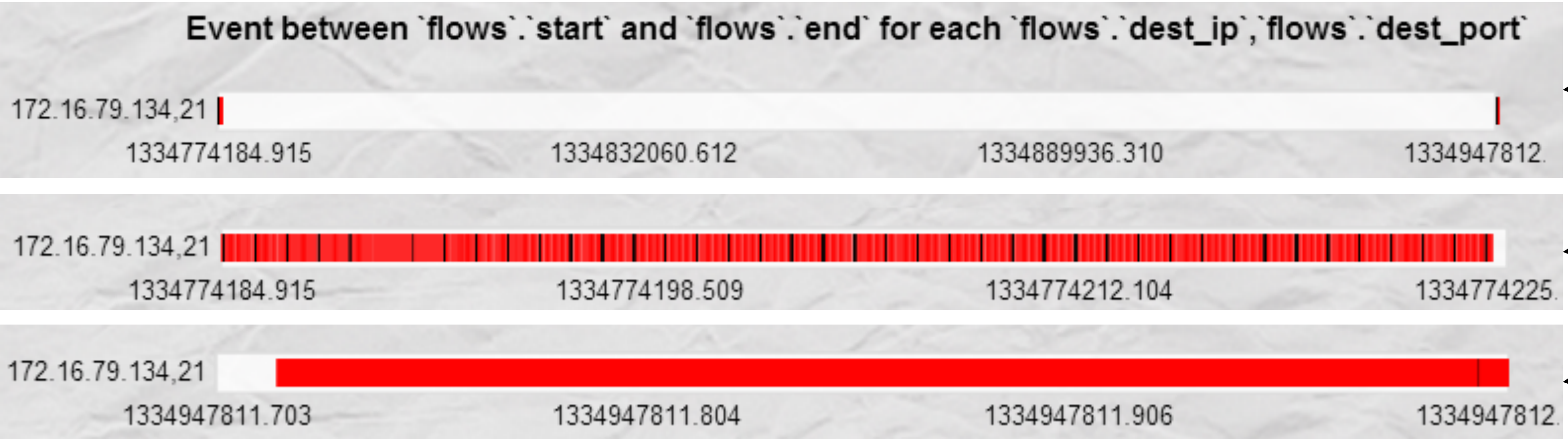


`iphdr`.`ip_src`	`iphdr`.`ip_dst`	`tcphdr`.`tcp_dport`	COUNT(*)	
172.16.79.128	172.16.79.134	21	894	all columns
172.16.79.128	172.16.79.134	617	3	all columns
172.16.79.128	172.16.79.134	5405	5	all columns
172.16.79.128	172.16.79.134	6667	2	all columns
172.16.79.128	172.16.79.134	8000	2	all columns
172.16.79.128	172.16.79.134	10050	2	all columns
172.16.79.128	172.16.79.134	45699	2	all columns

NetFlows for Target Destination IP and port

- Visualize netflow traffic related to the target destination IP on port 21
- Zoom to examine details in left and right flow clusters

- Right flow contains only one alert, does not look suspicious
- Most alerts happened in left flow, may contain attack
- Analyst decides to perform further analysis of traffic associated with left flow
  - E.g., include more tables and attributes to perform deeper analysis



Flows are distributed over two time ranges

Majority of alerts occur in left flow. Look suspicious

Right flow has single alert

Summarization of the Example

- Major steps supported by our visualization tool:
  - High level aggregation to highlight destination IPs with numerous alerts
  - Scatterplots to examine relationship between source IP and suspicious destination IP's ports
  - Correlated netflow visualization to examine timeline of alerts
  - Further analysis will focus on traffic related with the left flow

- Analysts focus on the data they are interested in at a given point in an investigation
- Easy to request follow-on visualizations and modify them to pursue new hypotheses and investigate new findings as they are uncovered

Future work

- Analysis Sandbox
  - Individual analyses can be performed, stored, reviewed and compared
  - Improve an analyst's "working memory" capacity
- Analysis Preferences
  - Track an analyst's actions to better anticipate their strategies for specific types of tasks
  - Use preference elicitation algorithms to track an analyst's interest within a visualization session
- Real-world Validation
  - Not allowed to speak directly with the analysts
  - Coordinate with IT staffs who support the analysts

Contact Information

Christopher G. Healey
Department of Computer Science
North Carolina State University

healey@ncsu.edu

Special Thanks

Peng Ning, CS Department, NC State University
Doug Reeves, CS Department, NC State University
Cliff Wang, Army Research Office