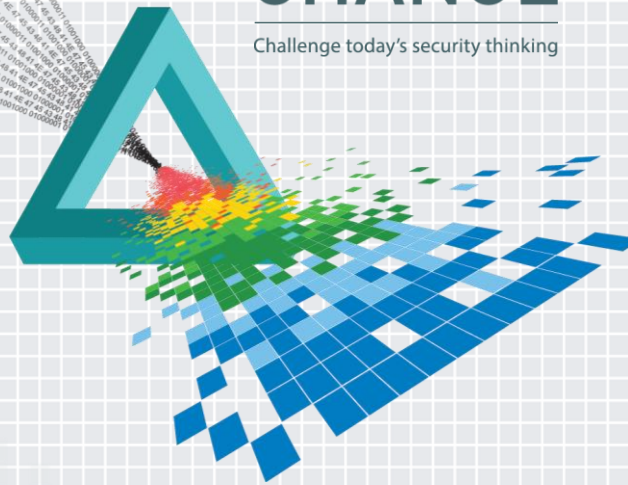
RSA Conference2015
San Francisco | April 20-24 | Moscone Center

CHANGE
Challenge today's security thinking



SESSION ID: CSV-F01

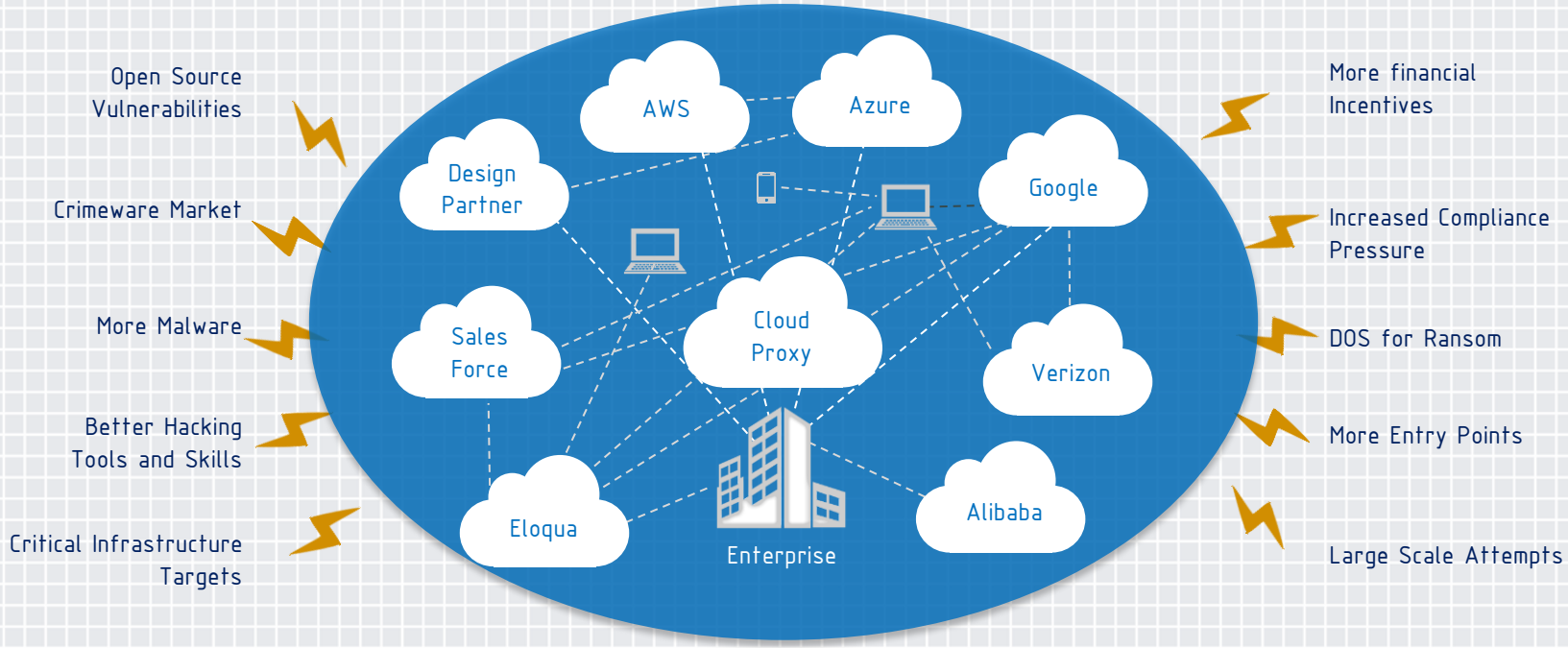
We're Gonna Need a Bigger Boat

Alan Ross
Senior Principal Engineer
Intel Corporation

Grant Babb
Research Scientist
Intel Corporation

IT Analytics: All about the changing Enterprise

Cloud + Evolving Threat Landscape = Complex Security Needs



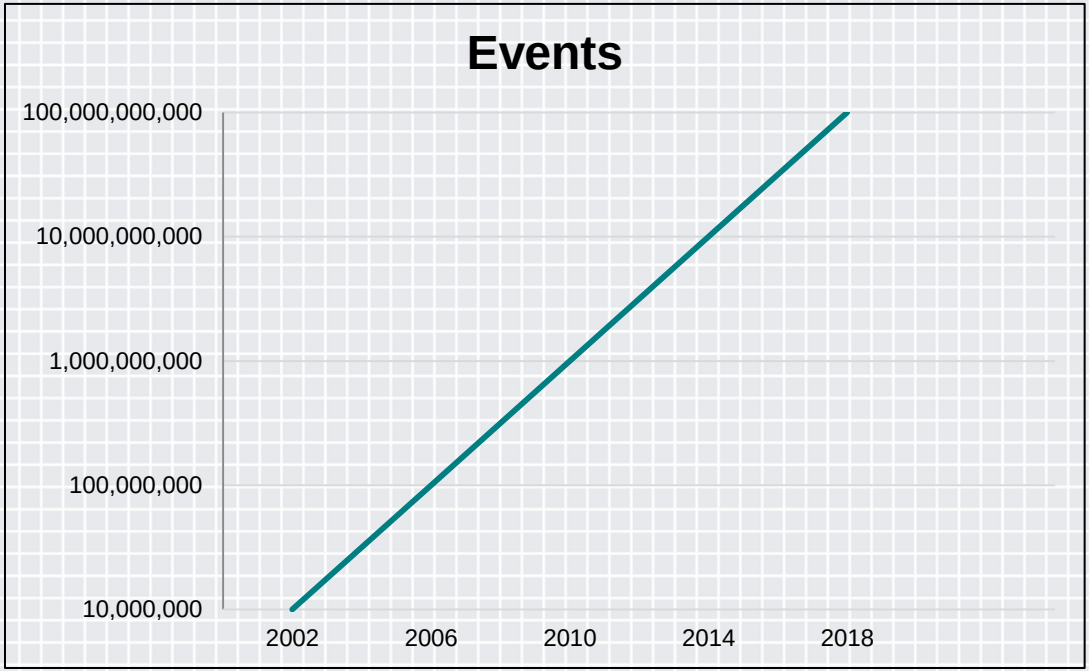
We Need Visibility, Transparency, and Control

What “Scale”?

Intel IT Today

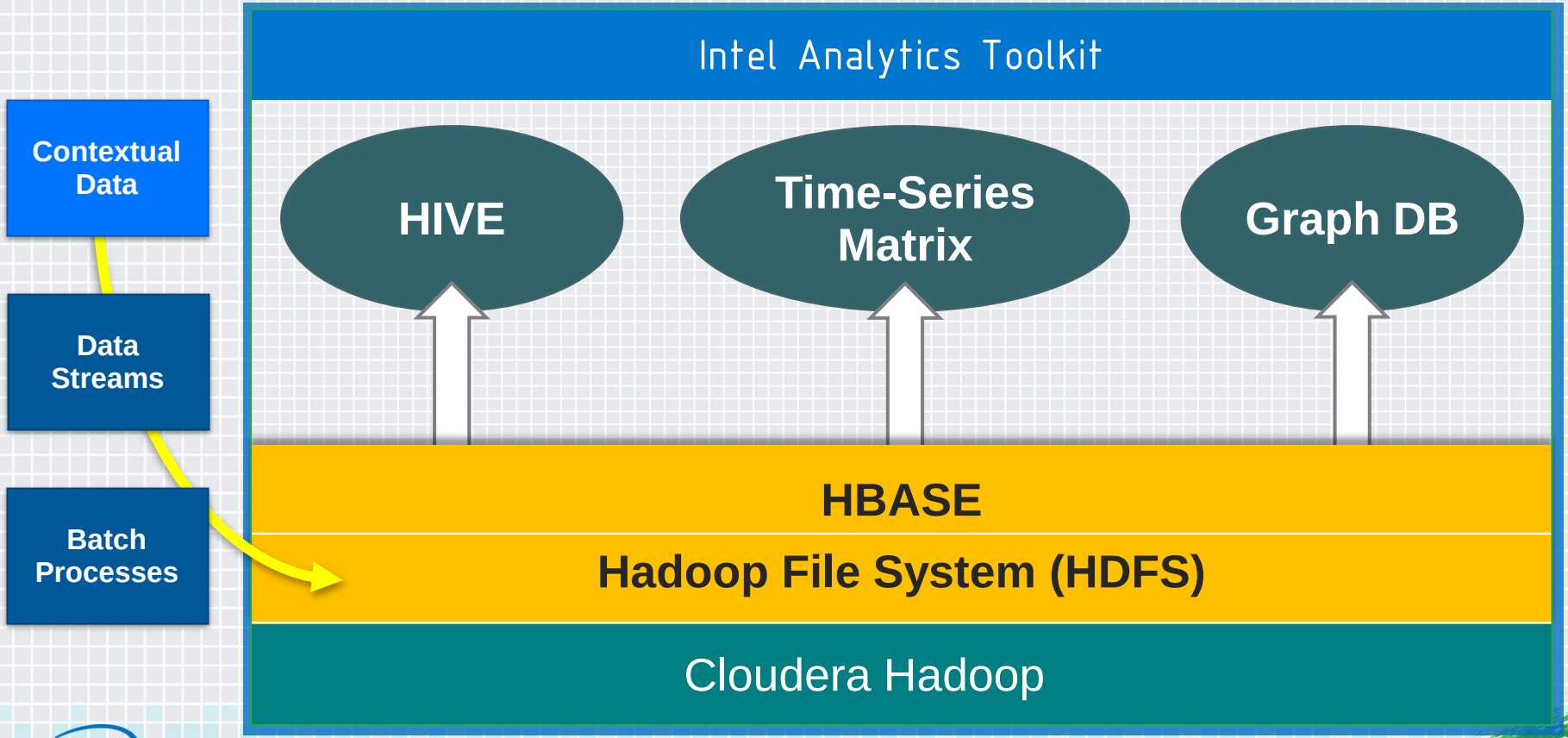
(if fully instrumented):

- 30-40 billion network related events/day
- Adding cloud environments could drive that to 50+ billion
- Not including platform, app, etc.

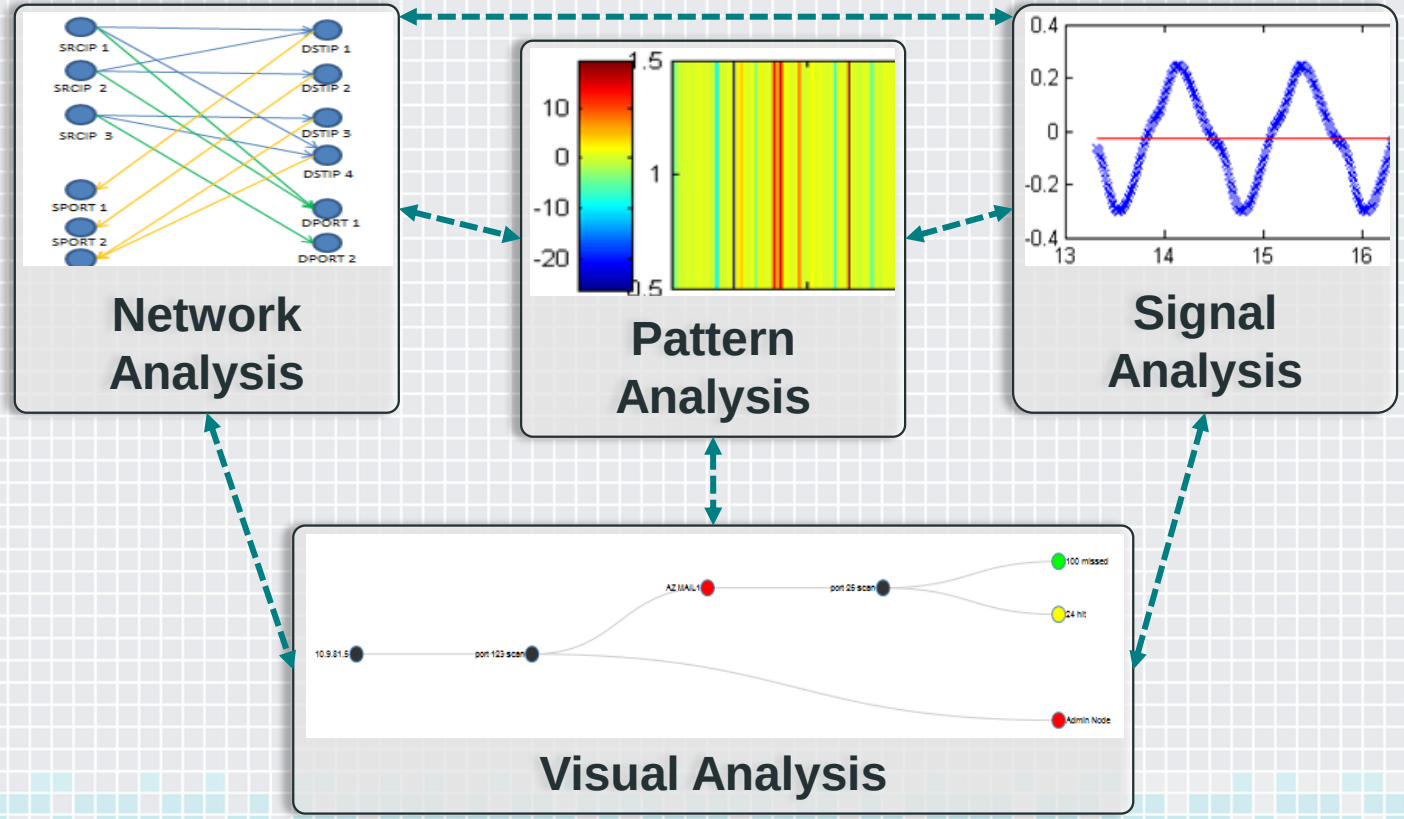


We need a platform that can scale to Hundreds of Billions of events/day!

“How do we process the data?”

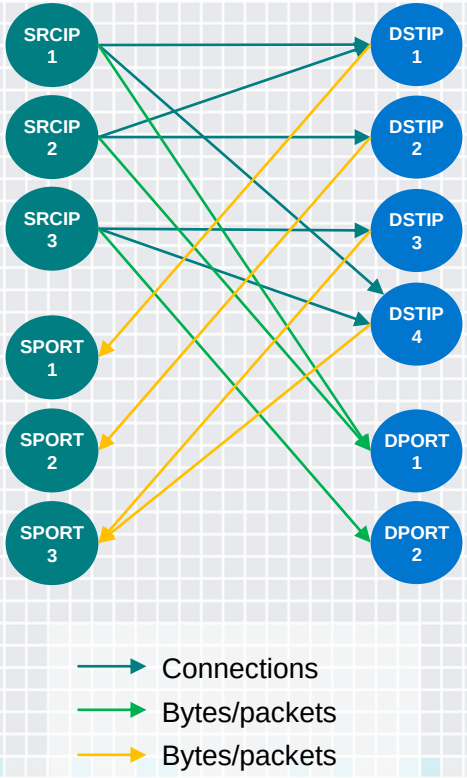


Analytic Approach



Graph Analysis: Latent Dirichlet Allocation

- Strives to put a population into sub-groups based on their similarity
- IP addresses are nodes, flow details are edges
- Use to cluster on known *(profiling)* or unknown *(automated behavior)*

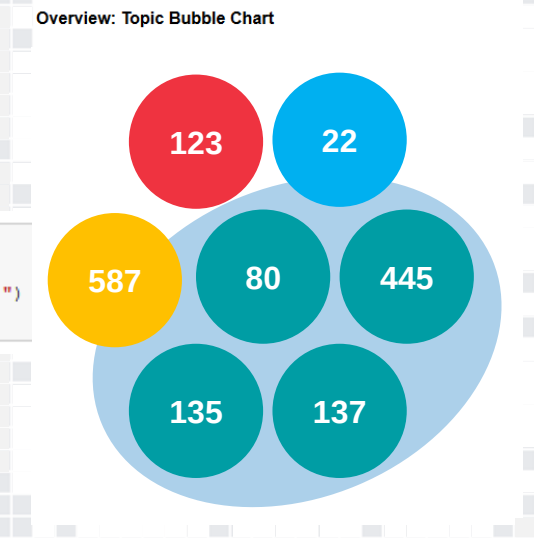


LDA results

Question:

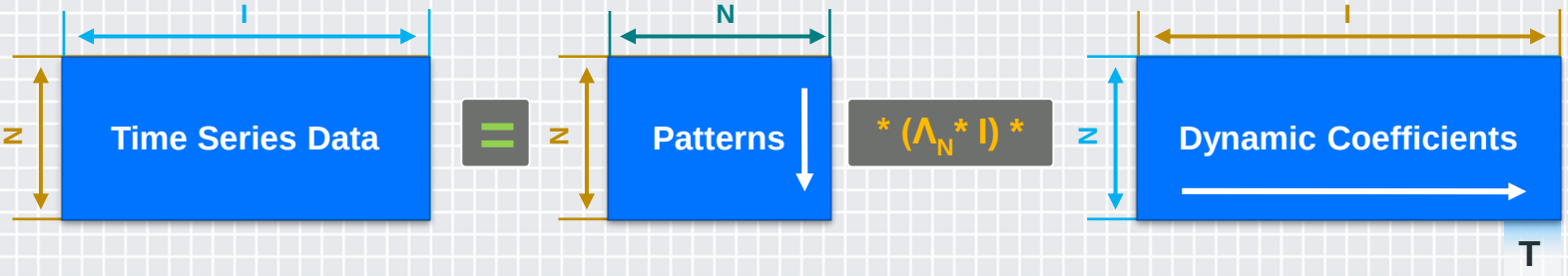
What are the strongest matches for groups based on automated communication to well-known ports ?

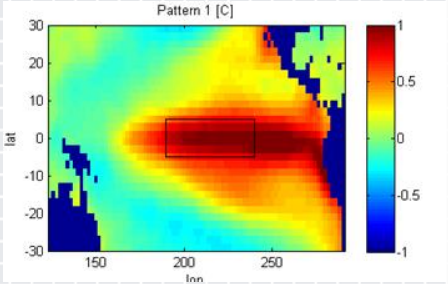
```
In [ ]: gname = 'netflow_topic'
        g = get_graph(gname)
        graph_result1 = g.query.gremlin("g.V.has('dport').has('lda_result',T.gte,0.9f).has('dport',T.lte,1024)")
        print 'results retrieved'
```

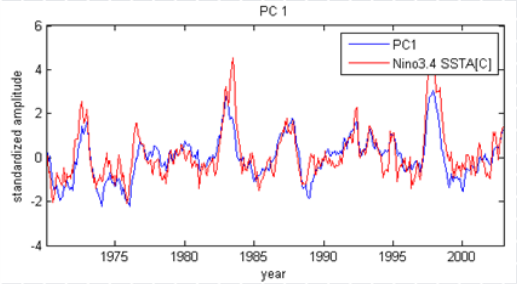


Answer: Seven ports in four different groups are the strongest matches

Patterns : Principal Component Analysis







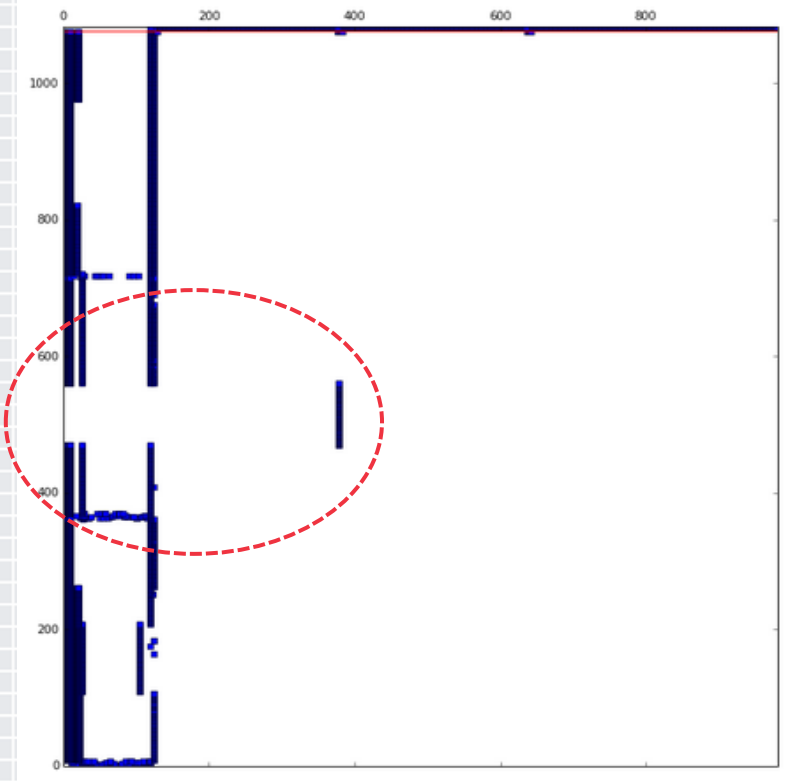
> The Use of PCs to summarize … climatological fields have been found to be so valuable that it is almost routine.
>
> *- Joliffe, Principal Component Analysis*

PCA Results

Question:

Are there any anomalous patterns in this data?

Answer: One source IP is talking to several destination IP's that do not exist *(horizontal scan)*



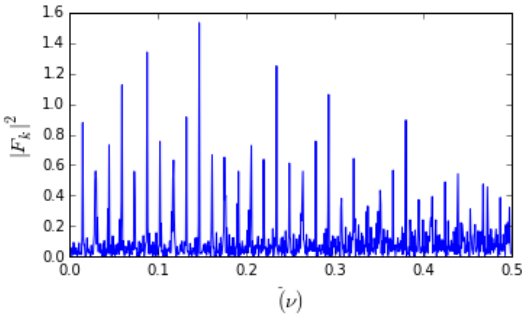
Signal Analysis: Fast Fourier Transform

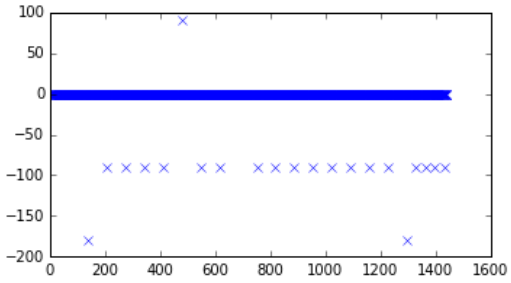
- Represent flow data as a function of sines and cosines *(waves)*
- Jump from time domain to frequency domain *(and back)*
- Easily filter noise from signal, or remove other frequencies

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos nx + b_n \sin nx\right) \quad x \in (-\pi, \pi]$$

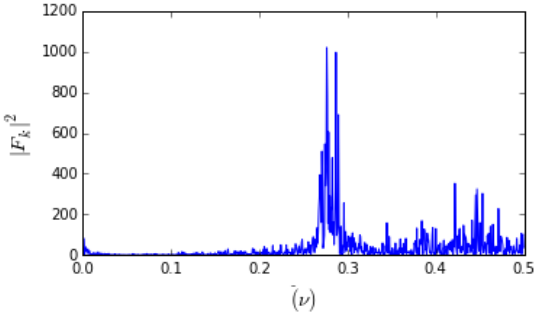
Signal Analysis - FFT

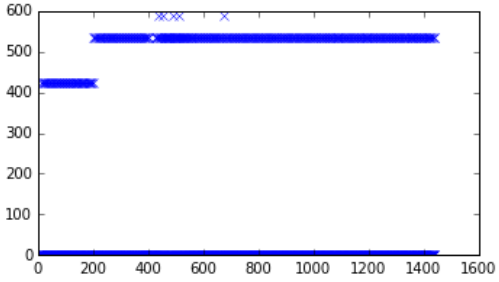
172.0.0.1 -> 172.30.0.3





10.170.148.64 -> 172.20.0.3





Visual Analytics: IPython and D3

IP[y]:

```
In [46]: %%javascript
         var viz = 'files/ipython/network.html'
         parent.document.getElementById('vizView').contentWindow.location.re
```

Possible Datasets

- Network Security
  - Firewall, Proxy, DNS, DHCP, SMTP, Active Directory, Netflows
- Platform Security
  - Antivirus, Antispyware, Host Intrusion, Firewall, Integrity
- Application Security
  - Whitelisting, Integrity
- Software Defined Infrastructure
  - Schedulers, Orchestration, Attestation, Integrity

What is Netflow?

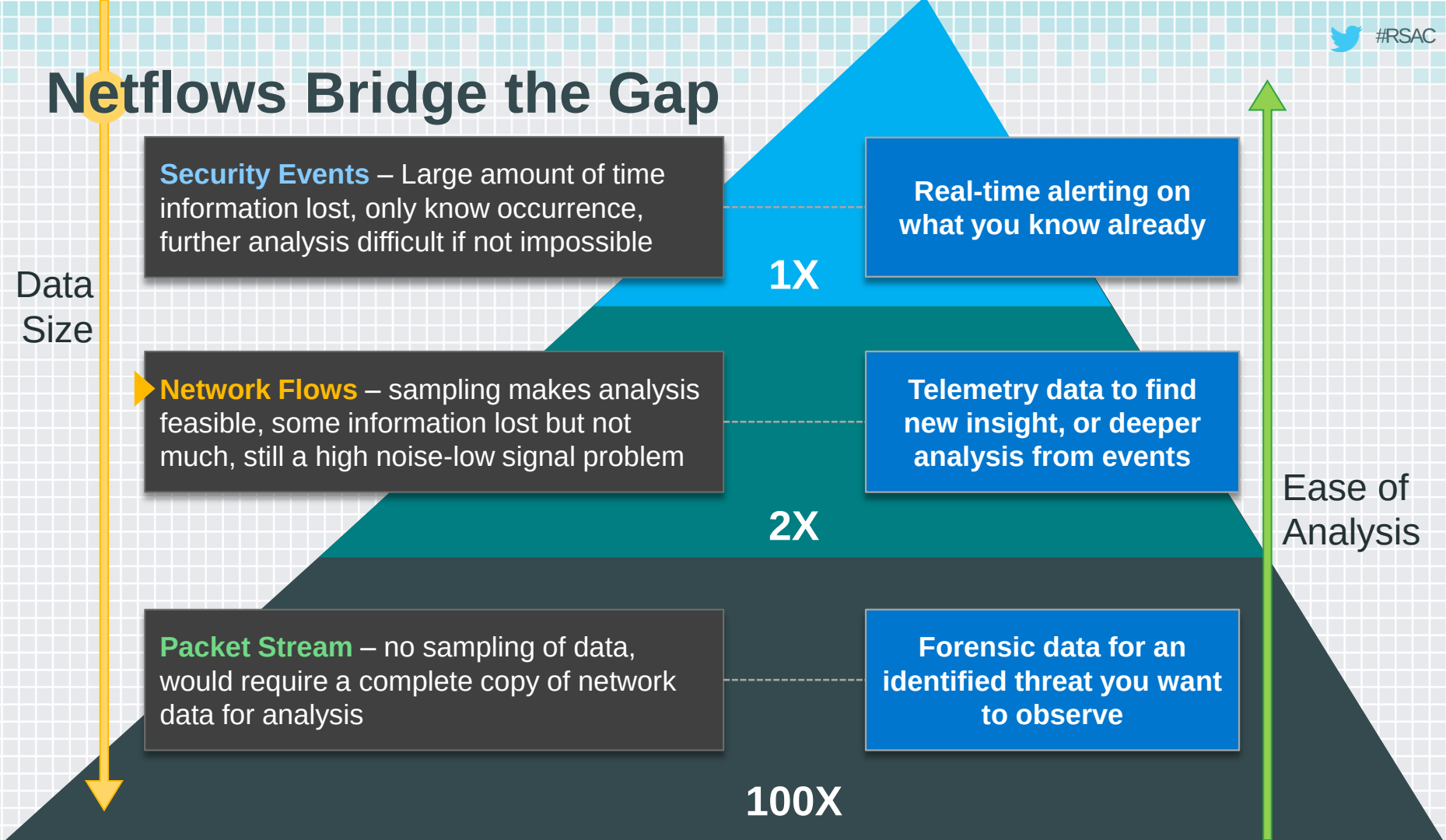
Network Accounting Protocol *(routers, switches)*

"Phone bill for network traffic"

Anything that crosses a network boundary creates a Netflow record - Including security threats and attacks

Time Started	Time Ended	Total Duration	Source Address	Destination Address	Source Port	Destination Port	Protocol	Flag	Packets	Bytes
7/7/2014 15:50	7/7/2014 15:51	52.416	208.43.253.206	192.198.147.164	443	60922	TCP	.AP...	10	1265
7/7/2014 15:50	7/7/2014 15:51	56.352	208.43.253.206	192.198.147.165	443	61701	TCP	.AP...	12	1518
7/7/2014 15:50	7/7/2014 15:51	60.256	202.188.66.182	192.198.147.39	50559	443	TCP	.AP...	11	736
7/7/2014 15:50	7/7/2014 15:51	56.48	54.230.149.156	192.198.147.165	80	59578	TCP	.AP.S.	267	365062
7/7/2014 15:50	7/7/2014 15:51	60.096	202.79.203.99	192.198.147.38	5286	443	TCP	.AP...	8	542
7/7/2014 15:50	7/7/2014 15:51	59.488	175.137.26.41	192.198.147.38	65422	443	UDP	.....	61	11077
7/7/2014 15:50	7/7/2014 15:51	51.712	208.43.253.206	192.198.147.164	443	64504	TCP	.AP...	11	1466
7/7/2014 15:50	7/7/2014 15:51	60.16	202.79.203.99	192.198.147.39	36513	443	TCP	.AP...	11	736
7/7/2014 15:50	7/7/2014 15:51	60	203.90.242.126	192.198.147.165	80	53787	TCP	.AP.S.	83	53632
7/7/2014 15:50	7/7/2014 15:51	56.096	23.67.71.136	192.198.147.164	80	53946	TCP	.AP...	8	2416
7/7/2014 15:50	7/7/2014 15:51	50.752	208.43.253.198	192.198.147.164	443	50486	TCP	.AP...	10	1265
7/7/2014 15:50	7/7/2014 15:51	51.744	208.43.253.206	192.198.147.164	443	58504	TCP	.AP...	10	1265
7/7/2014 15:50	7/7/2014 15:51	60.192	202.79.203.111	192.198.147.38	7860	443	TCP	.AP...	10	696
7/7/2014 15:50	7/7/2014 15:51	60.128	175.144.57.1	192.198.147.38	49669	443	TCP	.AP...	430	97625
7/7/2014 15:50	7/7/2014 15:51	44.704	50.112.120.33	192.198.147.165	80	63072	TCP	.AP.S.	18	11916
7/7/2014 15:50	7/7/2014 15:51	59.968	203.90.242.126	192.198.147.164	80	64014	TCP	.AP...	34	21360
7/7/2014 15:50	7/7/2014 15:51	57.44	208.43.253.198	192.198.147.165	443	58749	TCP	.AP...	12	1518

Netflows Bridge the Gap



Additional Context

- Internal IP address ranges
- Roles of known IP addresses *(e.g. proxy server, web server)*
- Geolocation information
- Security device policies *(e.g. firewall rules)*

Actionable Insight

- Summary views and drill downs make investigations and incident response easier.
- Filtering “noise” will make the machine learning smarter and learn over time

Suspicious Connects (showing top 150)

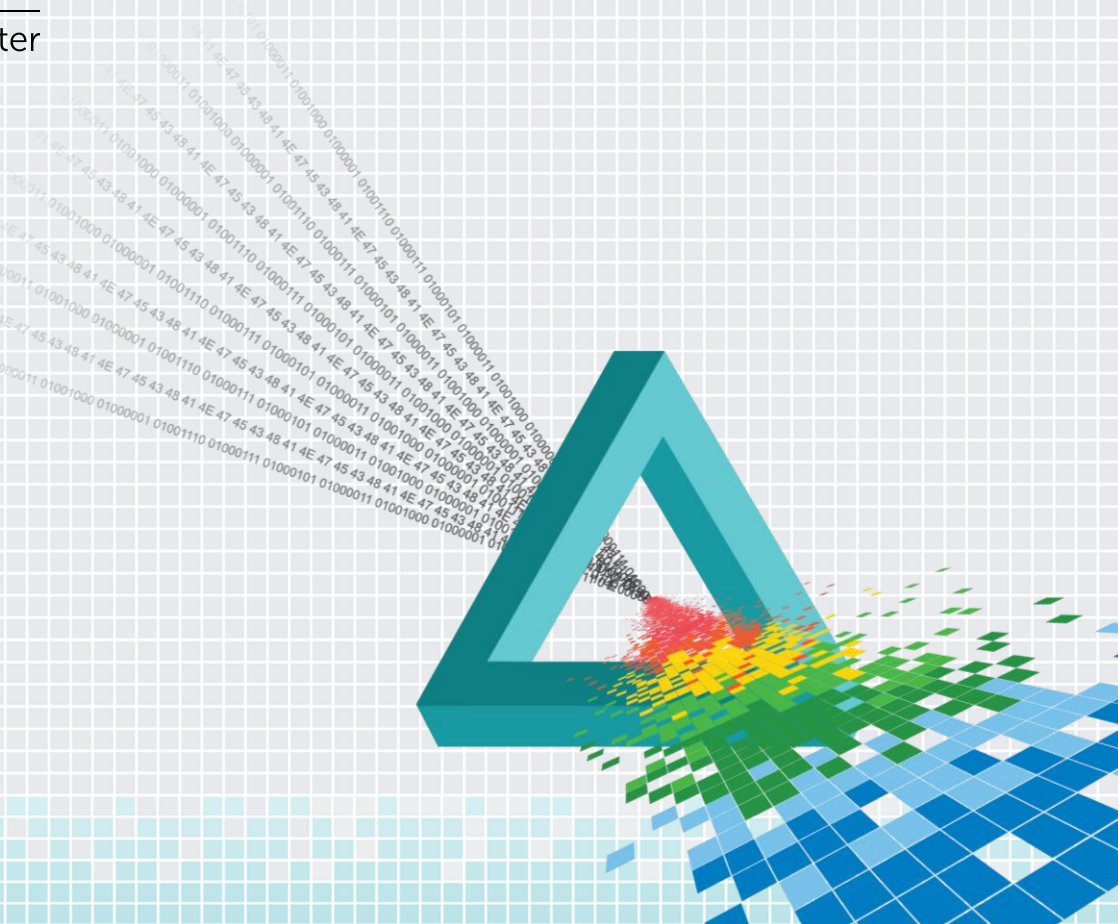
Suspicious connections are listed below in ranked order. Mouse over a connection to highlight it in the connection graph on the right. Click the connection to generate all the log details in the detail view.

tstart	srcIP	dstIP	sport	dport	proto	ipkt	ibyt
2014-07-08 10:10:40	134.191.242.7	174.139.113.58	123	1806	UDP	3	108
2014-07-08 09:49:13	192.198.156.3	174.139.113.58	2974	123	UDP	2	72
2014-07-08 09:49:13	192.198.156.2	174.139.113.58	2974	123	UDP	2	72
2014-07-08 09:49:13	192.198.156.1	174.139.113.58	2974	123	UDP	2	72
2014-07-08 09:49:15	192.198.156.133	174.139.113.58	2981	123	UDP	2	72
2014-07-08 22:30:47	134.134.73.2	202.109.253.101	59405	137	UDP	9	702
2014-07-08 23:15:22	134.134.73.2	202.109.253.101	59441	137	UDP	61	4758
2014-07-08 18:02:58	134.134.139.70	198.175.80.171	22	63684	TCP	348115	18467411
2014-07-08 18:04:57	134.134.139.70	198.175.80.171	53457	22	TCP	337746	437745779
2014-07-08 22:44:21	134.134.73.2	202.109.253.101	59405	137	UDP	16	1248
2014-07-08 22:44:21	134.134.73.2	202.109.253.101	59441	137	UDP	18	1404
2014-07-08 03:36:03	134.134.139.70	198.175.80.171	64425	22	TCP	350064	465426191

RSA Conference2015

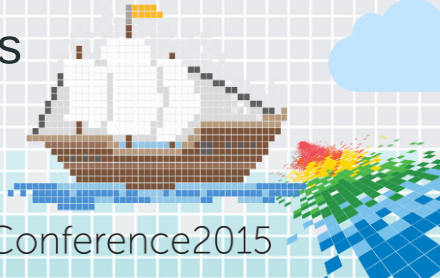
San Francisco | April 20-24 | Moscone Center

Demo



Future – Distributed Analytics

- Most “heavy lifting” will occur on premise
  - Don’t want to ship billions of events over wide area networks
  - May want multiple instances for a global environment
  - Cloud provider analytics are a likely trend to help with transparency
- Analytics will be combined and synthesized
  - On-premise and cloud analytics will be correlated
  - Analytic results will be shared across organizations to “raise all boats”
  - This will be a very collaborative activity across industries



How To Apply

- Download test data sets: VAST, CERT (http://vacommunity.org/VAST+Challenge+2013%3A+Mini-Challenge+3) (https://www.cert.org/insider-threat/tools/)
- Investigate what data you collect today that can be used to look for security threats!
  - Also are there other easy data sources to pull together and analyze?
- There is an abundant and extraordinary amount of free tools to start digging, learning and visualizing
  - "The democratization of data analytics and visualizations"
  - Gephi, Scilab, iPython, R, D3, NFDUMP,
- Tons of free courses online (e.g. Coursera Data Science)

Key Messages/Summary

1. The **scale** of compute has changed dramatically

2. We need a **platform** to store and process data at scale

3. We need algorithms and approaches to provide **insight**

4. We need **actionable** insights that solve hard problems