

Lessons Learned from 10 Years of Network Analysis R&D for Defense and Intel Customers

Thayne Coffman

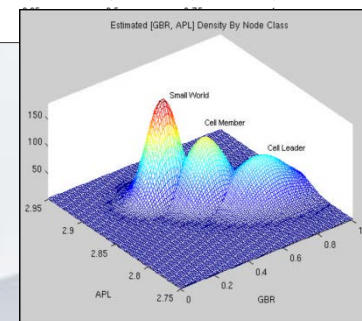
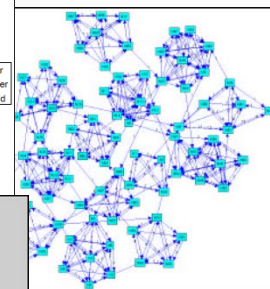
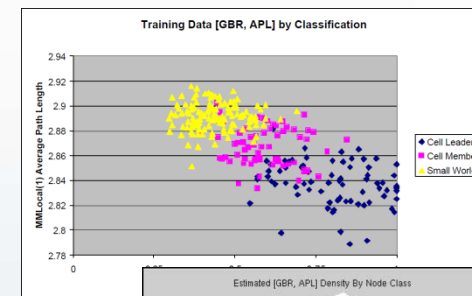
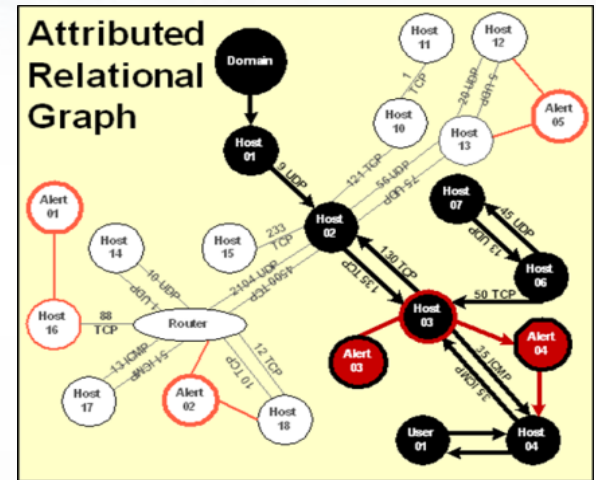
FloCon 2012

Austin, TX



The Speaker's Perspective

- 21CT
 - 12 years old, 90 ppl., Austin/SA/DC
 - Broad-spectrum R&D for DoD & IC
 - Now focused on applying LYNXeon™ graph analytics to flow data for USG & commercial
- Me
 - CS, AI, signal processing, pattern classification
 - 10 years @ 21CT: research, mgmt, strategy
 - Work marries graphs, signals, cyber, SNA, classification
- “Network” analysis == social or cyber
- Nobody is omniscient



Executive Summary

1. Analysts need tools that enable flexible workflows
2. Analysts need tools that run mid-complexity analytics
3. Anomaly detection is worth continued investment, but it will never be the whole answer

The screenshot shows the LYNXEON interface with a list of analytics on the left and a detailed view of a threat pattern on the right. The list includes:

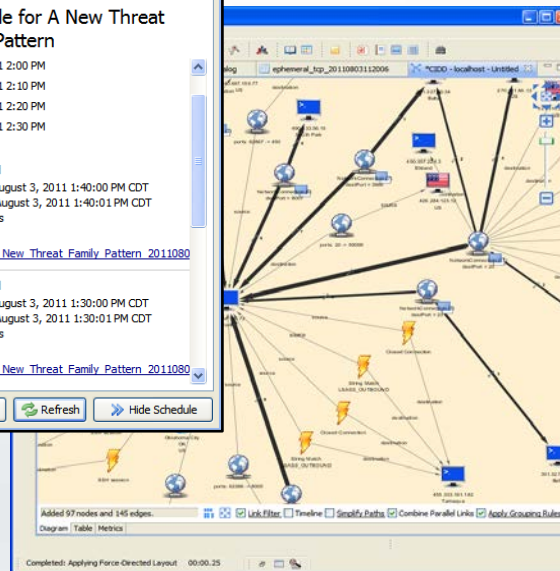
Name	Description
A New Threat Family Pattern	Finds new threat discovered by Bob
Ephemeral TCP Connections	Search for TCP connections from internal to external hosts that are using high ports.
Exfiltration Connections	Exfiltration connections are identified by looking for connections sending over 1 MB of traffic, where the sent/received ratio is 10 or over, and duration of connections are over 1 second.
FTP Exfiltration Connections	Search for potential exfiltration of data via FTP communications from compromised hosts. Look for event activity to identify the potentially exploited hosts, followed by external FTP transfers.
FTP Exfiltration Connections (Temporal)	Search for potential exfiltration of data via FTP communications from compromised hosts. Look for event activity to identify the potentially exploited hosts, followed by external FTP transfers. Enforces the temporal ordering of events before the FTP connection.
Invalid IP Packets	Search for connections exchanging invalid packet sizes for the given protocols.

The detailed view shows a "Schedule for A New Threat Family Pattern" with the following details:

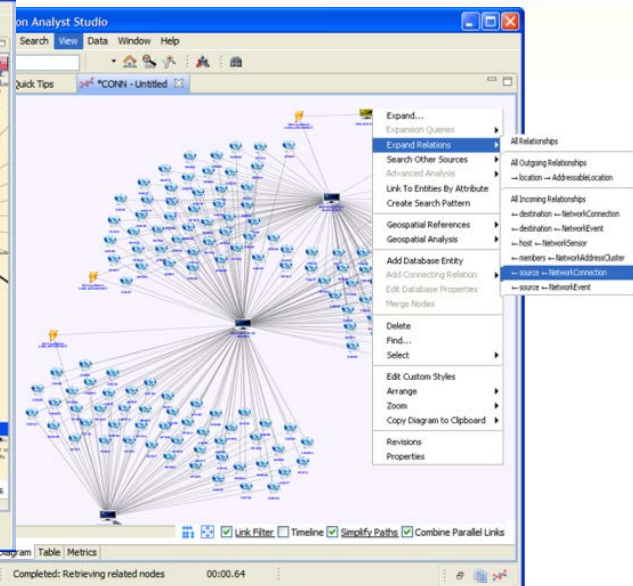
- Completed
- Started: August 3, 2011 2:00 PM
- Finished: August 3, 2011 2:20 PM
- Elapsed: 1s
- Results: A New Threat Family Pattern_20110803112006

Below this, another completed entry is shown:

- Completed
- Started: August 3, 2011 1:40:00 PM CDT
- Finished: August 3, 2011 1:40:01 PM CDT
- Elapsed: 1s
- Results: A New Threat Family Pattern_20110803114000



LYNXeon

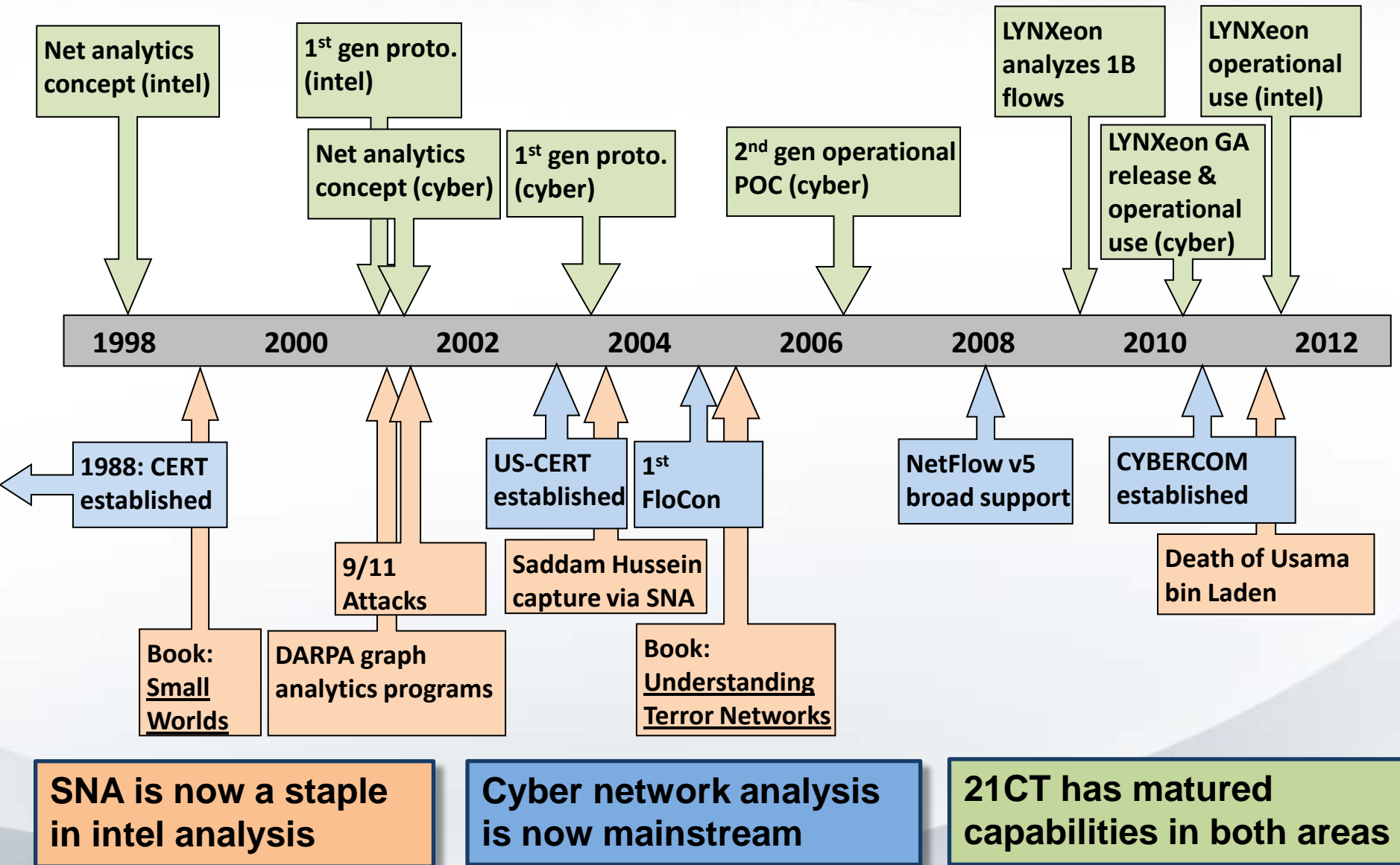


Briefing Roadmap

- 1. Analysts need tools that enable flexible workflows**
2. Analysts need tools that run mid-complexity analytics
3. Anomaly detection is worth continued investment, but it will never be the whole answer

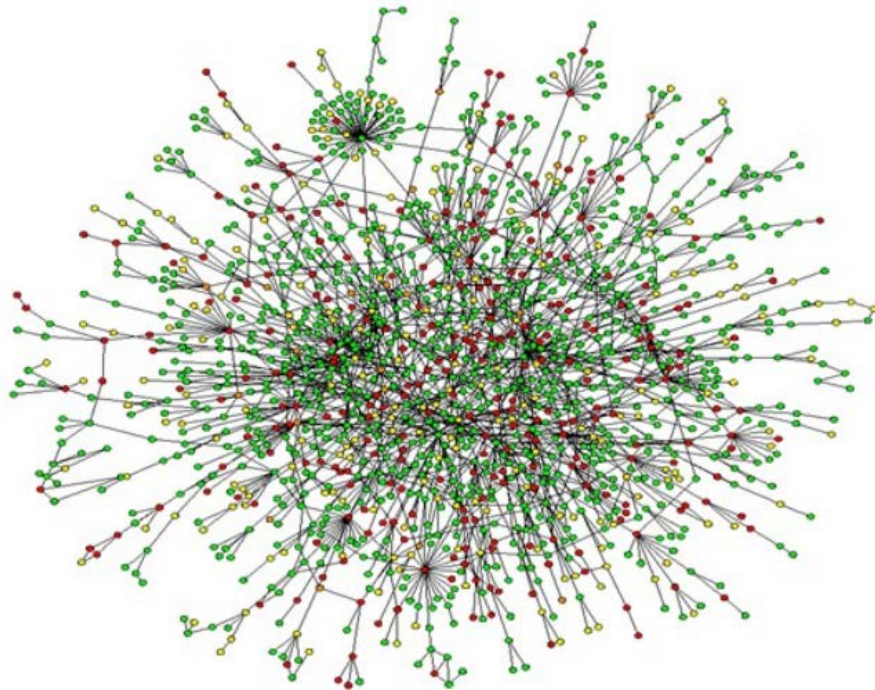


Network Analytics for Intel. & Cyber



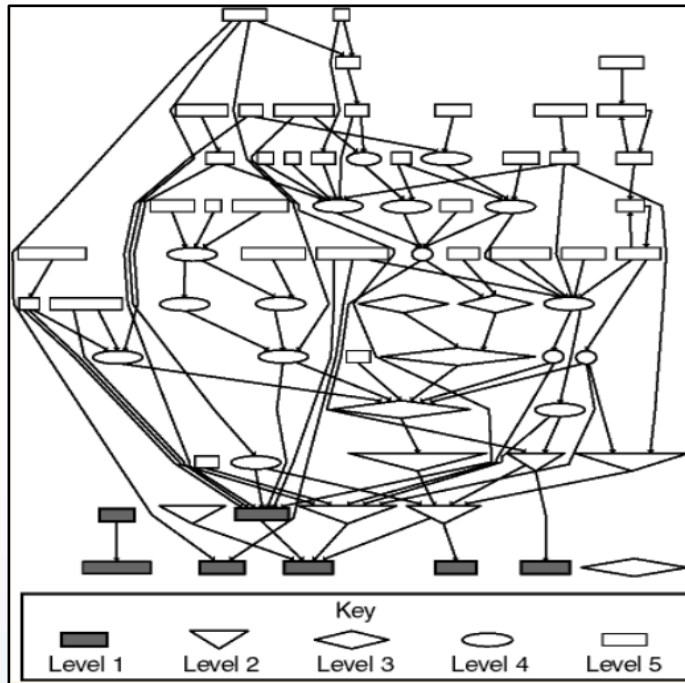
Lesson 1: The Problem

- Too much data to search & understand unaided
(Severe challenges in even automated processing)
- Too many attacks to run to ground
- Urgent need for deeply buried answers



Lesson 1: Doing it Wrong

- Try to take the analyst out of the loop
- Massive, inflexible, automated, integrated data mining “solutions”
- Fixed workflows built around standing queries



≠

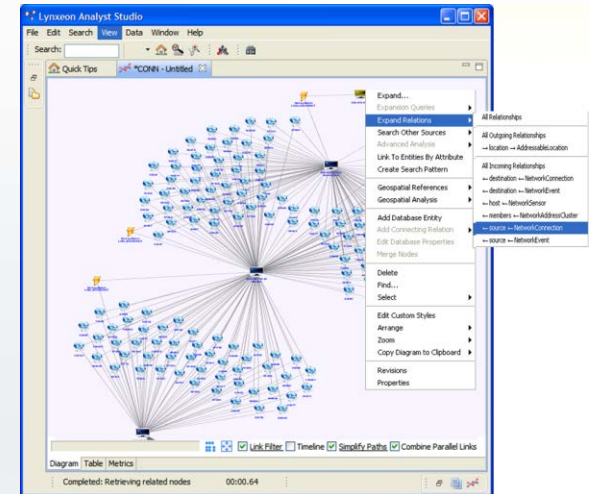
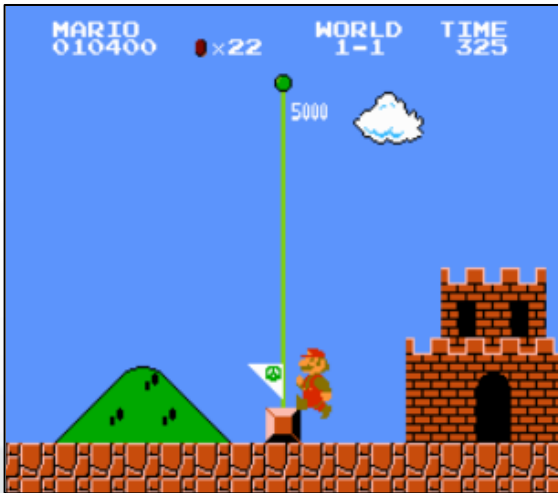


- $\{P(F+) = 0.001\% \} \cdot \{10^9 \text{ flows}\} = 10^4 \text{ false positives. Now what?}$

Lesson 1: Doing it Right

Analysts need tools that enable flexible workflows.

- Embrace an analyst-centric iterative process
 - Avoid hardcoded analytics & workflows
 - Sandbox tools – i.e., platforms
 - Minimize timespan of: ideas/workflows → prototype analytics → reusable tools
 - Distill, mature, scale, apply, integrate, catalog, and share analytics

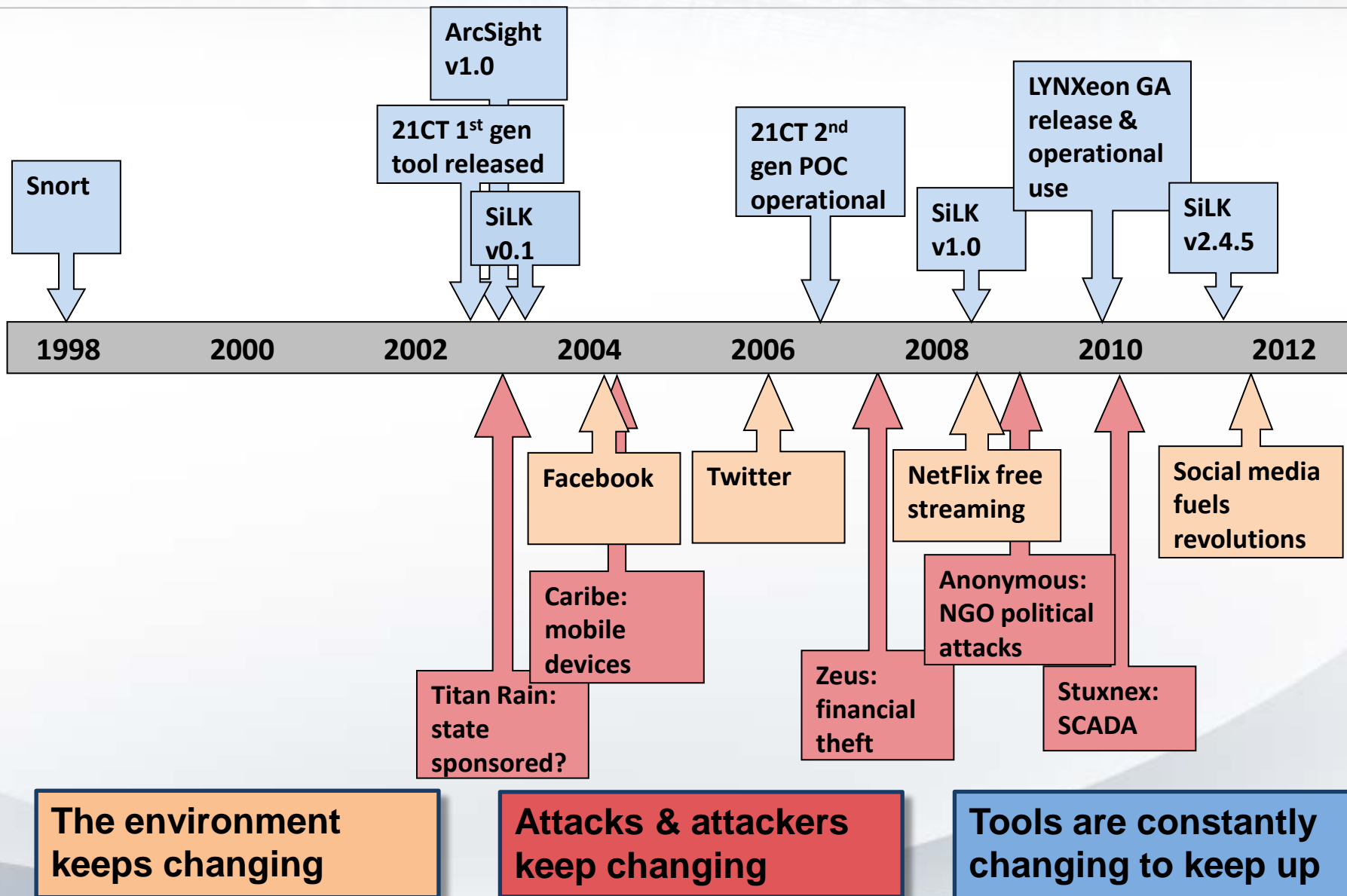


Briefing Roadmap

1. Analysts need tools that enable flexible workflows
- 2. Analysts need tools that run mid-complexity analytics**
3. Anomaly detection is worth continued investment, but it will never be the whole answer



Cat and Mouse in a Changing World



Lesson 2: The Problem

- Unexpected changes in environment and attacks
- Signatures only catch what they're looking for
- Anomaly detection doesn't fill all the gaps "yet"

Morris Worm

Caribe

Melissa

Stuxnet



Project
Chanology

Titan Rain

Simile

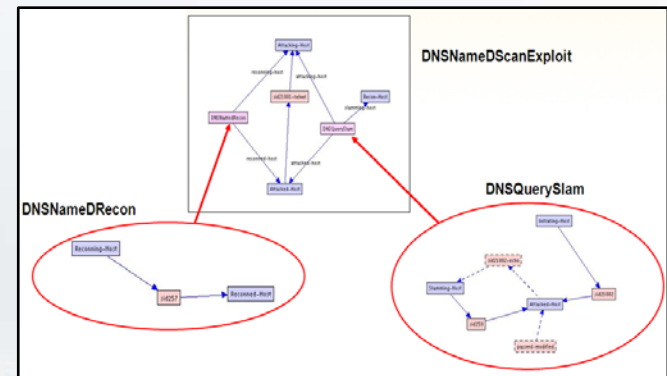
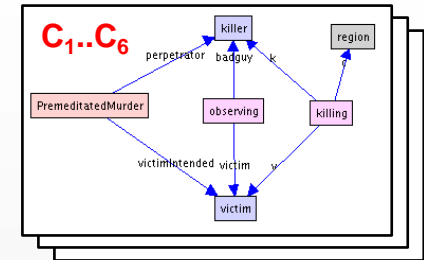
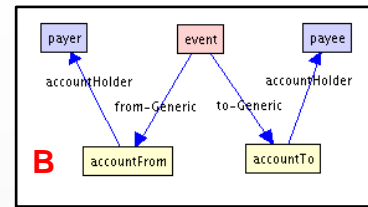
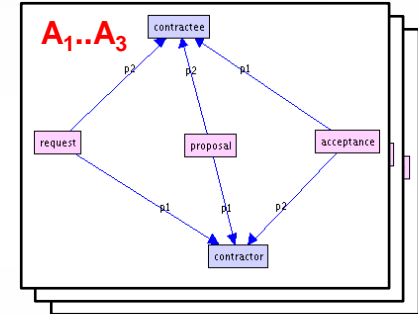
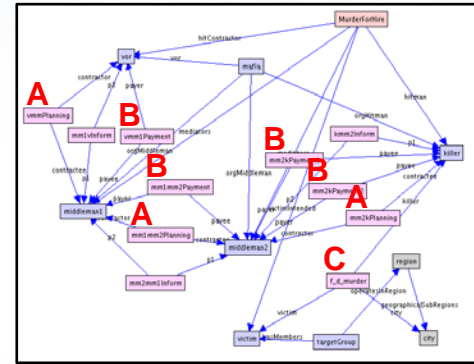
ILOVEYOU

nimda

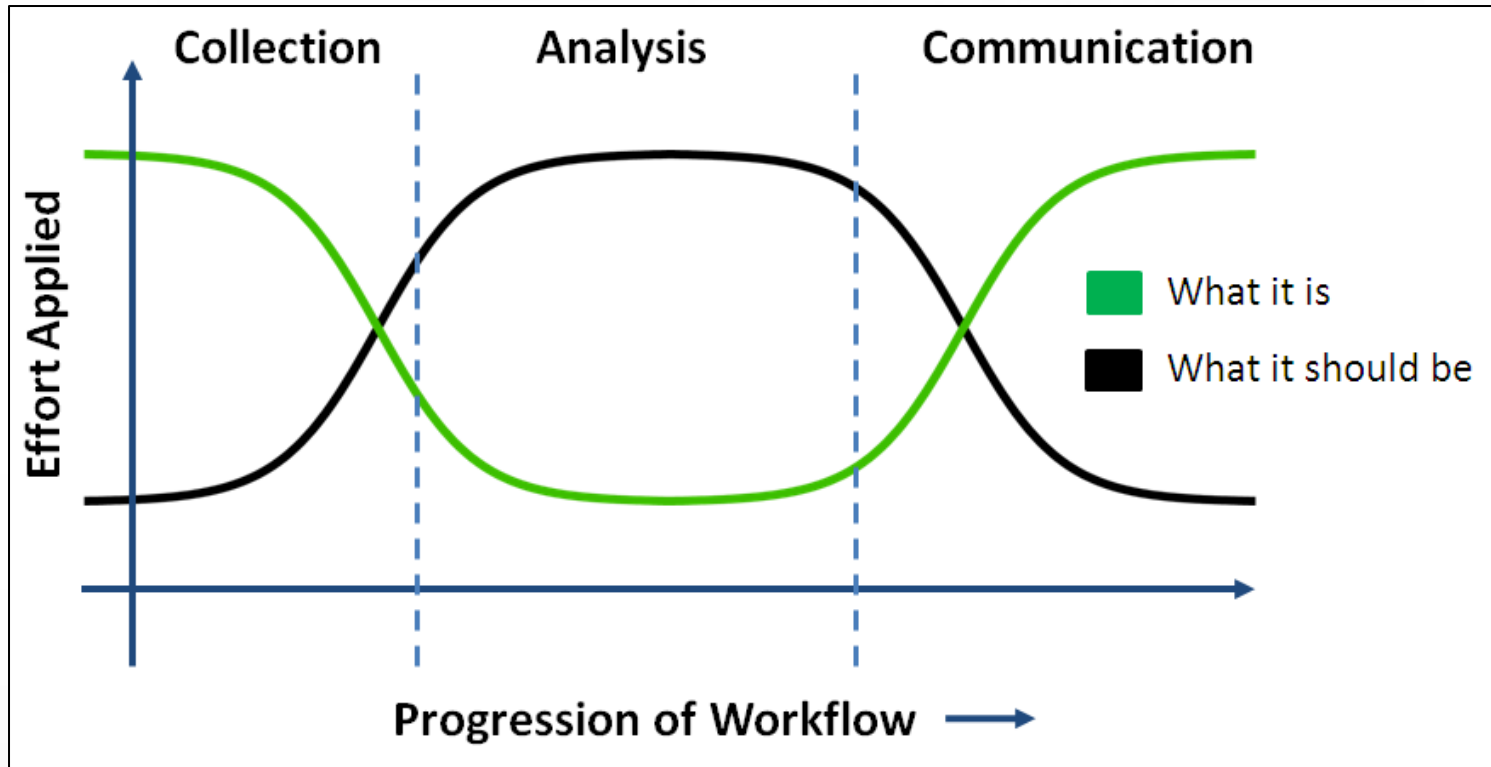


Lesson 2: Doing it Wrong

- Try to make your signatures flexible
- Contract murders example
 - 10^4 - 10^5 elements to search
 - Multi-level complex patterns
 - Matches 1.3M variations
 - ...and inexact matching
- That's flexible enough, right?



The Intelligence Analysis Bathtub

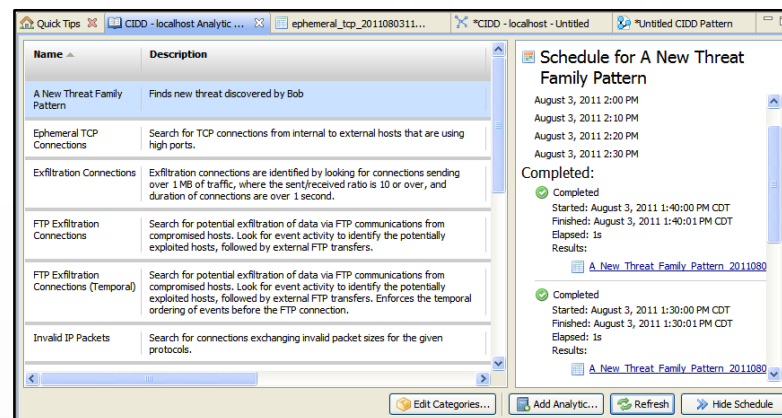


- Massive systems = accept the bathtub (but don't say that)
- "Flexible patterns" = accept the bathtub (but don't say that)
- How do we really invert the bathtub?

Lesson 2: Doing it Right

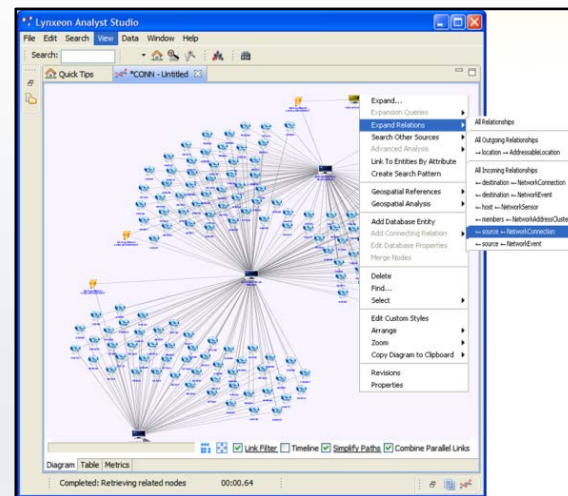
Analysts need tools that run mid-complexity analytics.

- Too small = return to overload
- Just right = simple correlations
- Too big = never flexible enough
- Combine with flexible workflows
 - Bite-sized fast & scalable analytics
 - Analyst builds ad hoc analysis chains based on task, attack, & data exploration
 - Run, see results, augment/pivot, repeat
- Embrace and enable the analyst in the loop



The screenshot shows a software interface with a table of threat patterns and a sidebar. The table has two columns: 'Name' and 'Description'. The sidebar shows a 'Schedule for A New Threat Family Pattern' with a list of dates and times, and a 'Completed' status with details like 'Started: August 3, 2011 1:40:00 PM CDT' and 'Elapsed: 1s'.

Name	Description
A New Threat Family Pattern	Finds new threat discovered by Bob
Ephemeral TCP Connections	Search for TCP connections from internal to external hosts that are using high ports.
Exfiltration Connections	Exfiltration connections are identified by looking for connections sending over 1 MB of traffic, where the sent/received ratio is 10 or over, and duration of connections are over 1 second.
FTP Exfiltration Connections	Search for potential exfiltration of data via FTP communications from compromised hosts. Look for event activity to identify the potentially exploited hosts, followed by external FTP transfers.
FTP Exfiltration Connections (Temporal)	Search for potential exfiltration of data via FTP communications from compromised hosts, followed by external FTP transfers. Enforces the temporal ordering of events before the FTP connection.
Invalid IP Packets	Search for connections exchanging invalid packet sizes for the given protocols.

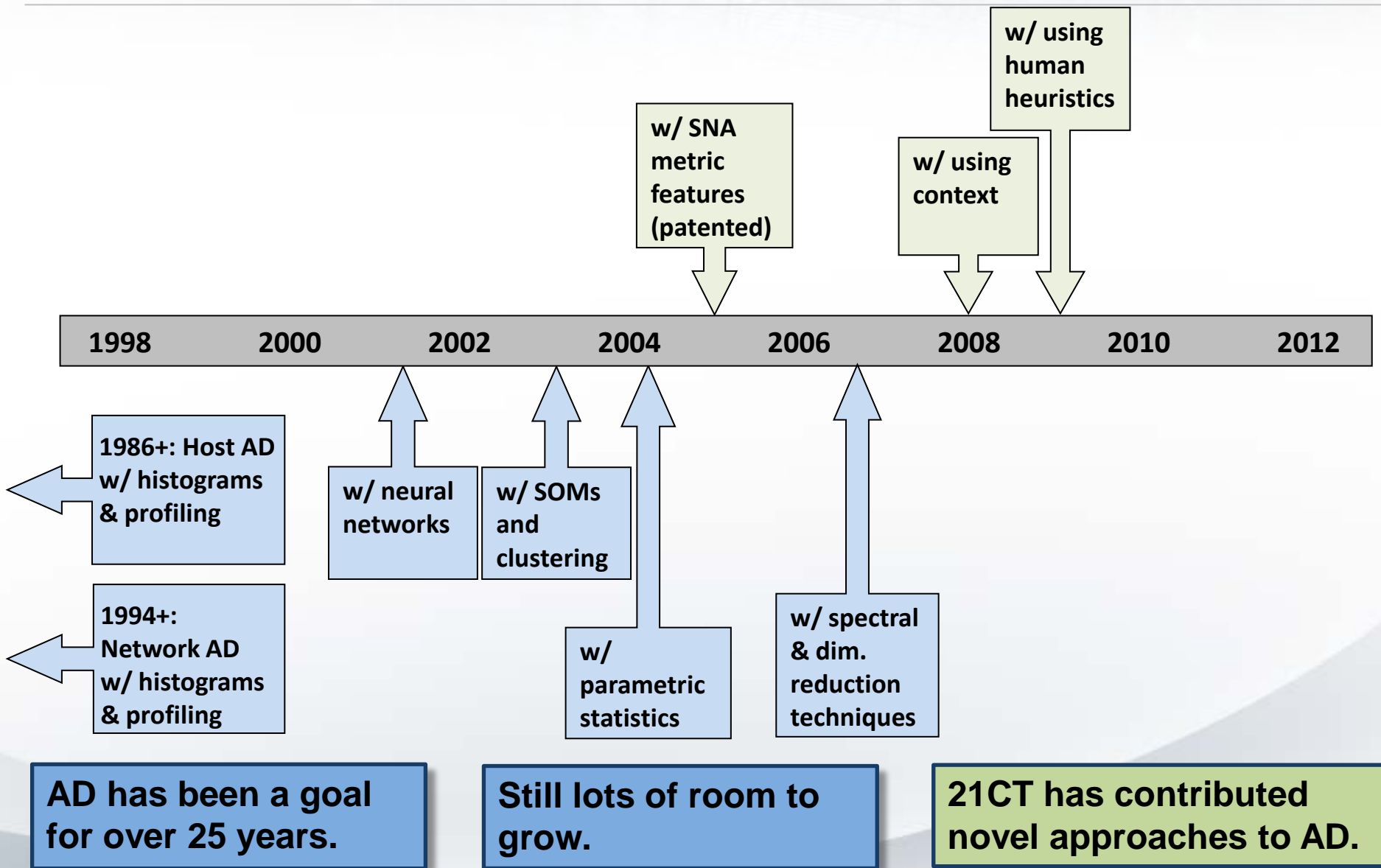


Briefing Roadmap

1. Analysts need tools that enable flexible workflows
2. Analysts need tools that run mid-complexity analytics
- 3. Anomaly detection is worth continued investment, but it will never be the whole answer**

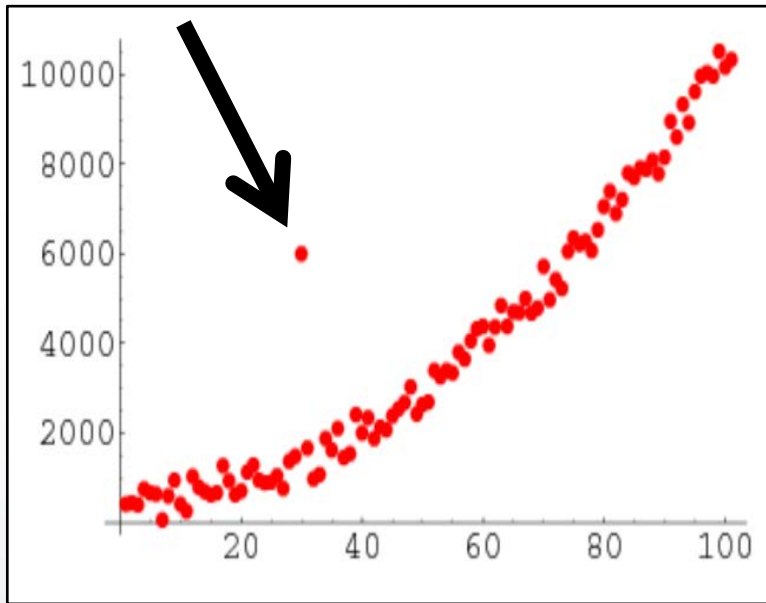


A Brief History of Time Anomaly Detection

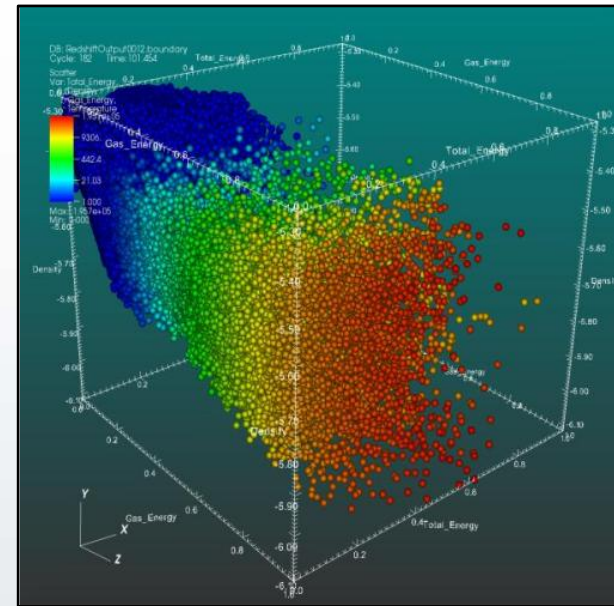


Lesson 3: The Problem

- Can anomaly detection fill the detection gap?
- Changing environments, tactics, attacks, and data
- Too much data, and too little
- The smart adversaries try to look normal



A.D. HAPPY!



A.D. SAD!

Lesson 3: Doing it Wrong

- Rely on AD as an auto-magic detector that finds (only) bad people
 - $P(F+)$ will never be zero
 - Many technical challenges remain: training data, generality, flexibility
- Accepts the bathtub, once again
- True generalized AD == a human, strong AI, or oracle

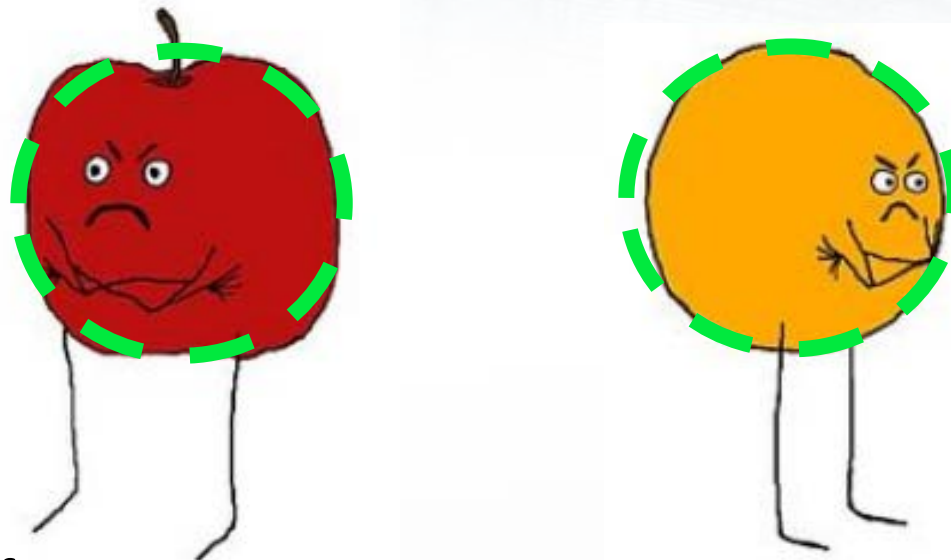


Lesson 3: Doing it Right

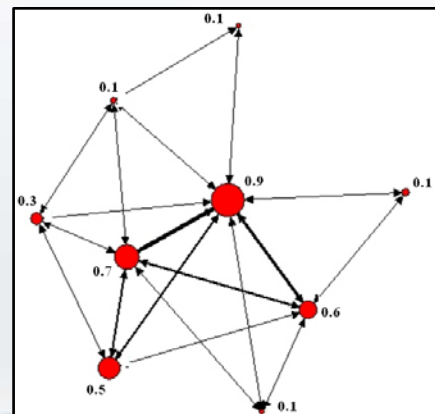
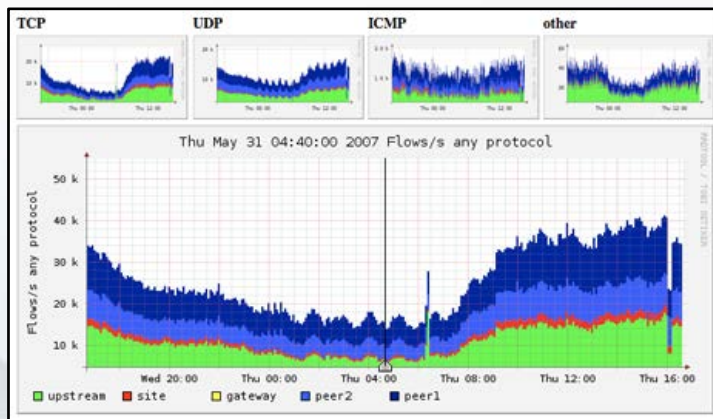
Anomaly detection is worth continued investment, but it will never be the whole answer.

- Inherent gaps point back to analyst-centric model
- Use for analyst cueing like other detectors
- Still lots of room to grow
- Consider these 4 ideas...

Lesson 3.1: Look for Better Features



- Traditional features == communication quantity
- Social network analysis metrics == communication structure



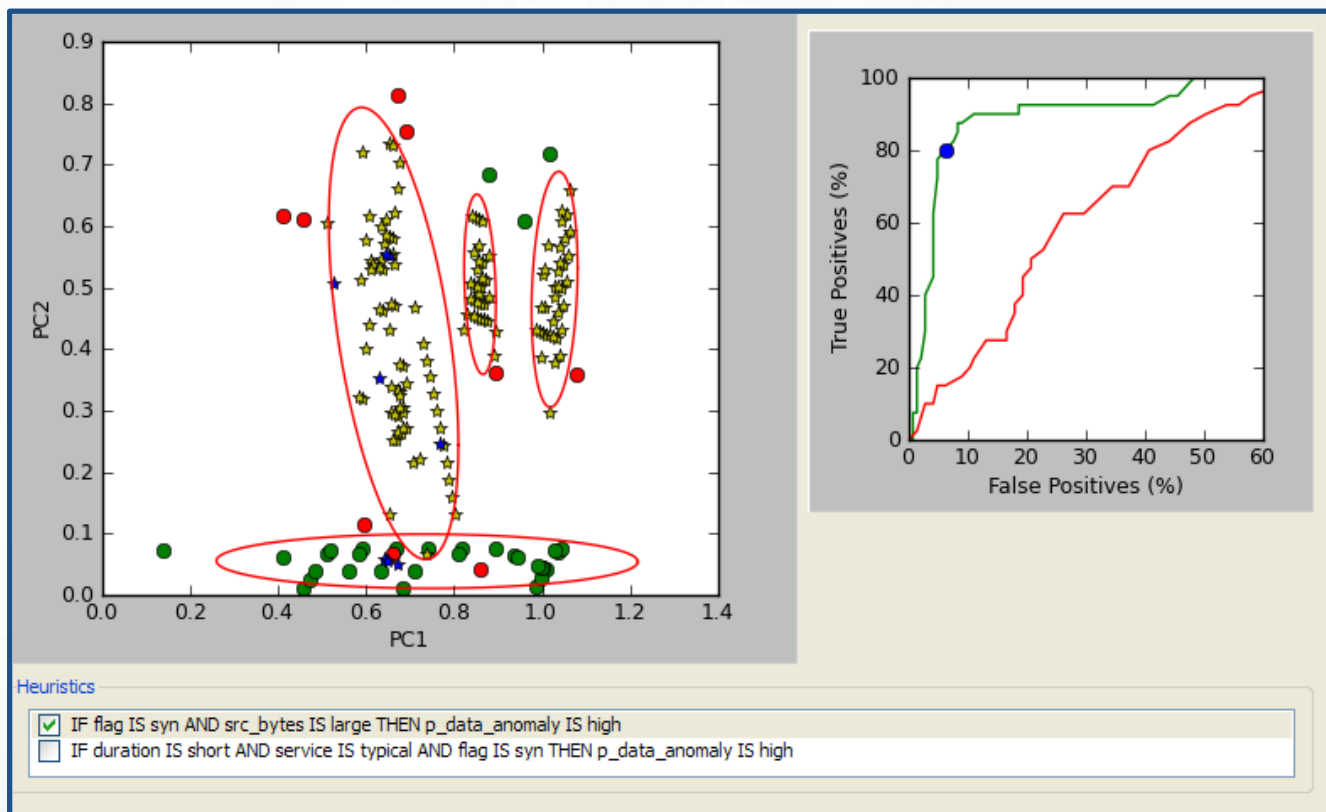
Lesson 3.2: Leverage Context



- Flexibly pull in external context data (hard)
- Condition training data
- Then cluster & group

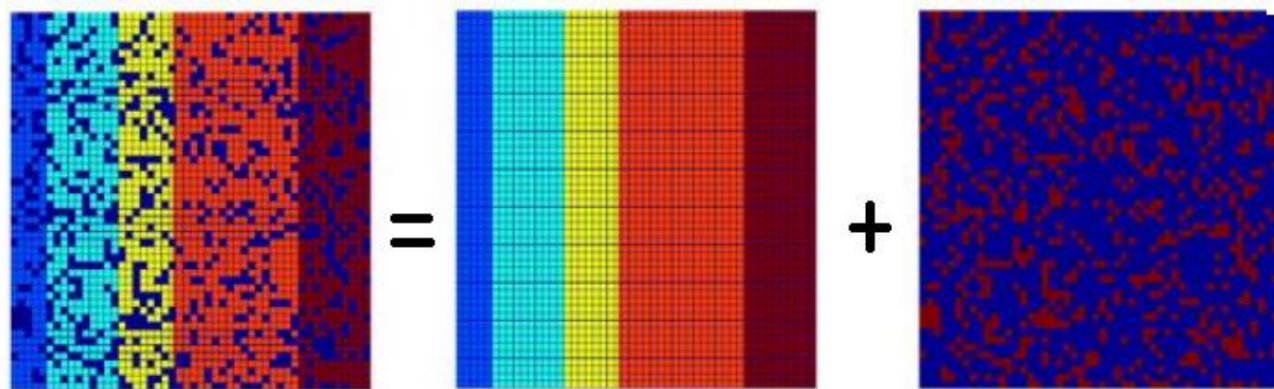
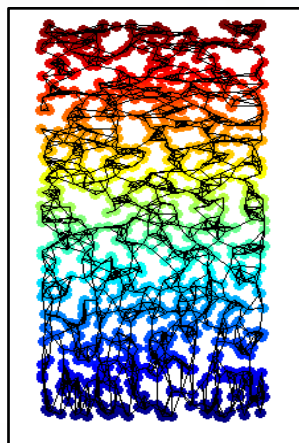
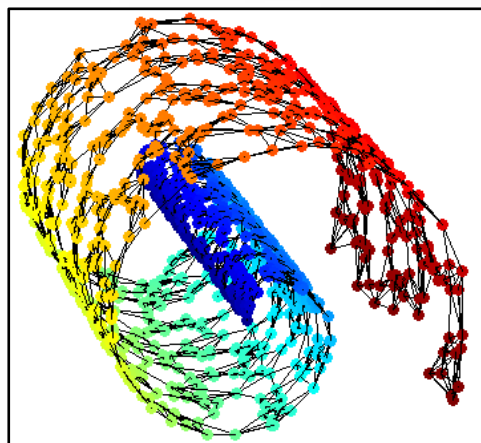
Lesson 3.3: Leverage Domain Expertise

21CT prototype
built under AFRL
anomaly detection
research effort



- Leverage analyst expertise to locally modify sensitivity
- Makes anomaly detection more adaptive

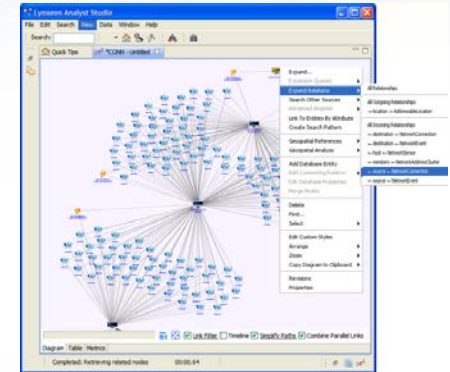
Lesson 3.4: Manage Dimensions and Data



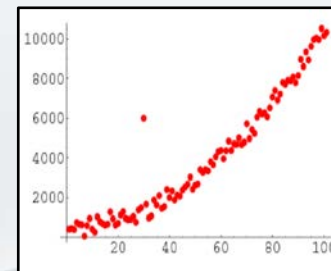
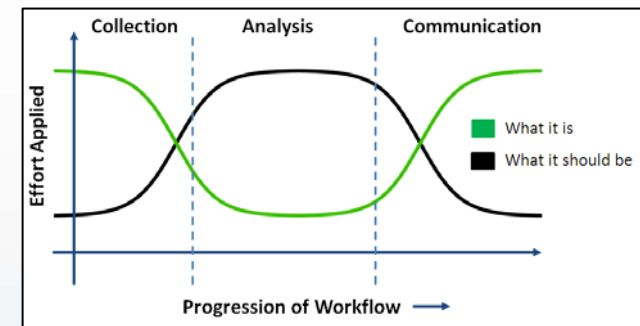
- Submanifold learning & dimensionality reduction
- Sparse representations, sparse matrix completion

Conclusions

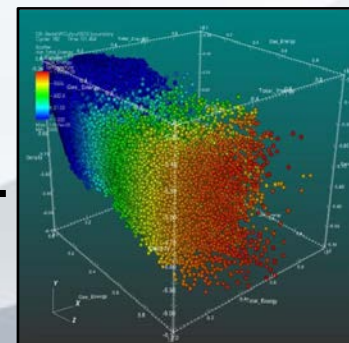
1. Analysts need tools that enable flexible workflows
 - Human must be inside the loop, and needs help
 - One workflow will never fit all
2. Analysts need tools that run mid-complexity analytics
 - Hand-in-hand with flexible workflows
 - Truly inverts the bathtub
3. Anomaly detection is worth continued investment, but it will never be the whole answer
 - Lots of room to grow and value to add
 - But full AD means a human or strong AI



LYNXeon



vs.



Questions & Discussion

For future questions, contact:
Dr. Thayne Coffman
Chief Technology Officer
21CT
tcoffman@21technologies.com

