# FROM FALSE POSITIVES TO ACTIONABLE ANALYSIS:

## BEHAVIORAL INTRUSION DETECTION, MACHINE LEARNING AND THE SOC

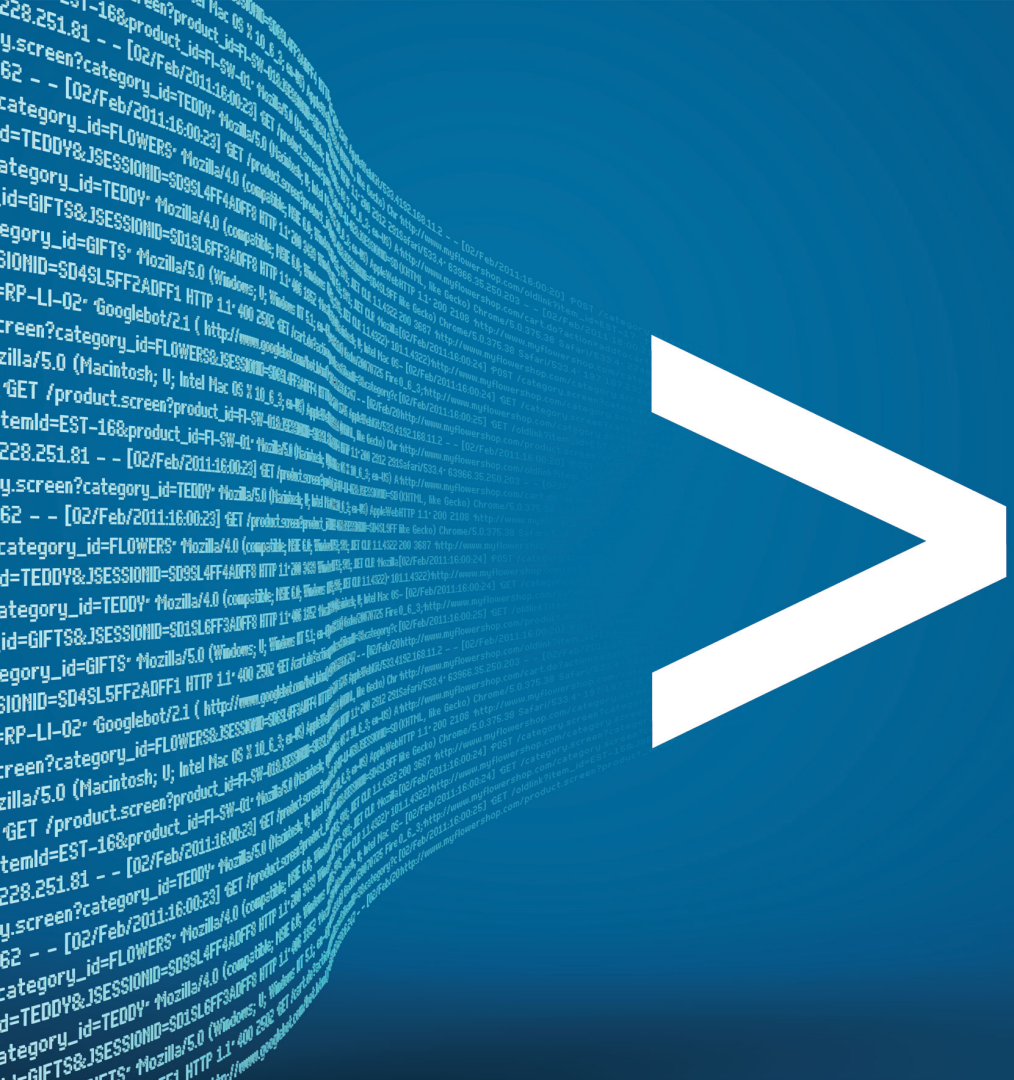
Joseph Zadeh

@JosephZadeh

splunk>

# Agenda

- Introduction
- Cybersecurity Analytics ROI
- Lambda Security: Defensive Architectures
- Behavioral Intrusion Detection and “Artificial” Intelligence
- Q&A



# Introduction

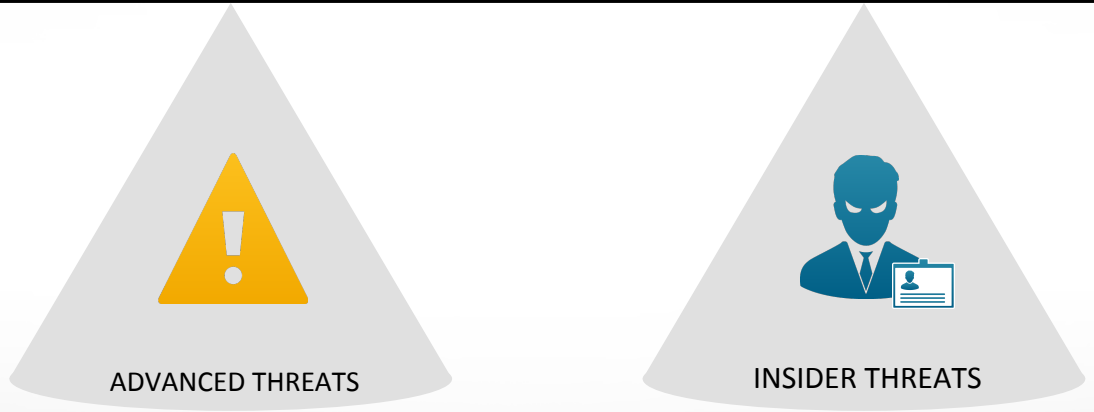
splunk>

# Splunk Acquires Caspida

**Extends Security Analytics Leadership by Adding Behavioral Analytics to Better Detect Advanced and Insider Threats**

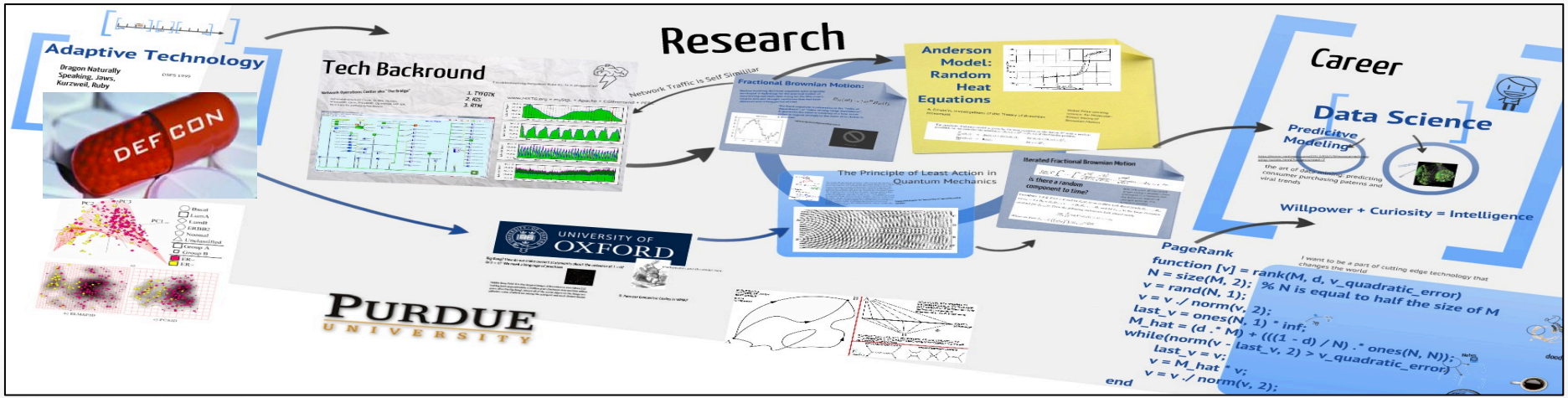
**Come see us at the Black Hat Booth #347**

splunk>enterprise + ES Splunk App for Enterprise Security + Machine Learning

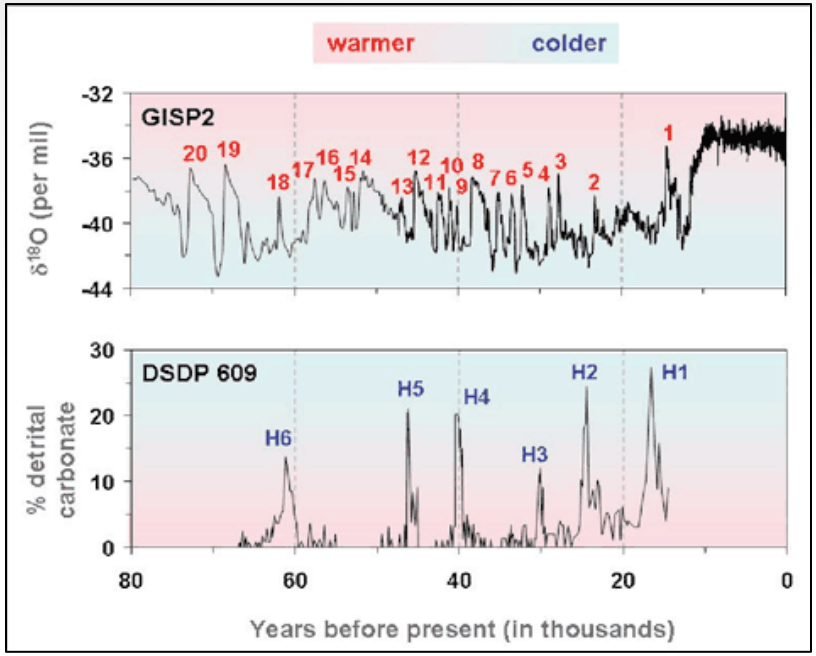


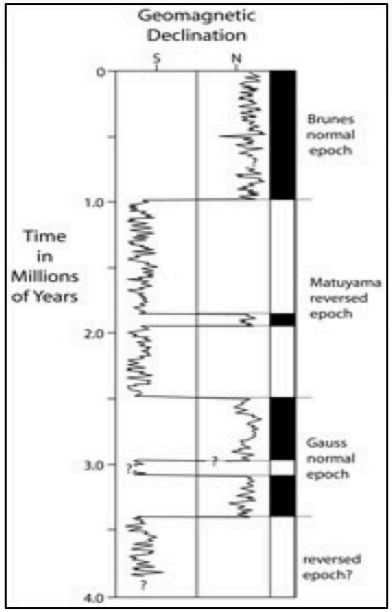
# Research Background

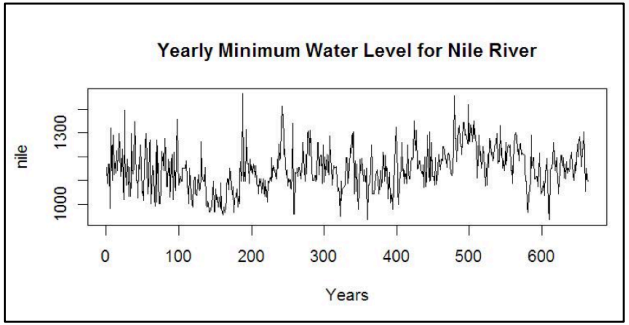
- Security Research
  - First security talk ever attended @ Decfon 8: Jon Erickson "Number Theory, Complexity Theory, Cryptography, and Quantum Computing"
    - I am still trying to understand that talk…
  - Late 90's: Traffic baselines and Layer 2 behavioral profiling using MRTG and first generation NMS
  - Recently: Consultant on cybersecurity analytics projects the last few years standing up custom solutions
- Related Research:
  - Fractal random walks: predicting time series
    - Human Behavior (stocks), Physical Processes (heat, magnetism)
    - Molecular kinetics (Brownian motion, quantum mechanics)
    - Stochastic Fast Dynamo, Stochastic Anderson Equation
    - Time as a random process



# ML: Predicting Time Series



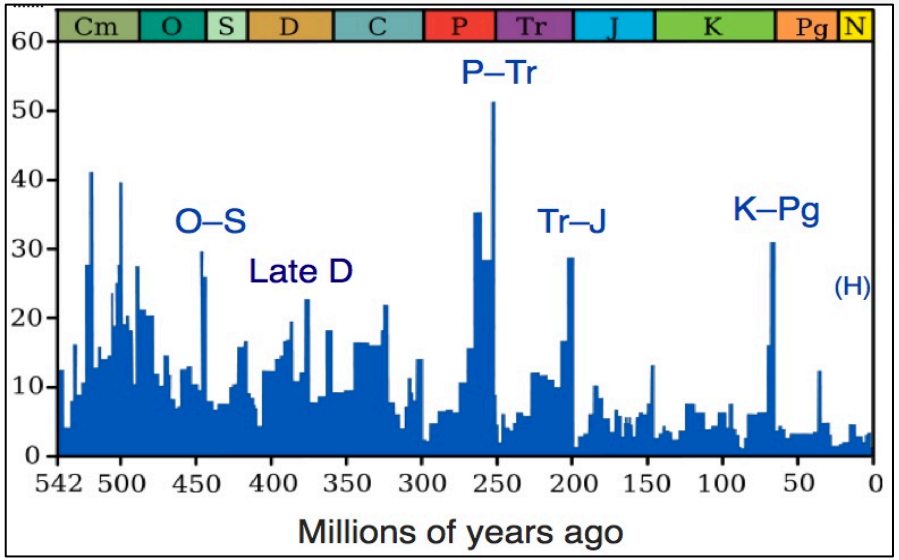



# A.I. and the “Big” Picture

Can Strong AI help predict..

- The dynamo effect?
- Pole shifts?
- Extinction events?
- Military escalations?
- Cybersecurity catastrophes?



# Cybersecurity Defense: Failings and Motivations

Mudge, "How a Hacker Has Helped Influence the Government - and Vice Versa" Blackhat 2011

9,000 Malware Samples Analyzed
- 125 LOC for Average Malware Sample
- Stuxnet = 15,000 LOC (120x average malware sample LOC)
- 10,000,000 = Average LOC for modern firewall/security stack

**Key Takeaway: For one single offensive LOC defenders write 100,000 LOC**
- 120:1 Stuxnet to average malware
- 500:1 Simple text editor to average malware
- 2,000:1 Malware suite to average malware
- 100,000:1 Defensive tool to average malware
- 1,000,000:1 Target operating system to average malware

Bruce Schneier, "The State of Incident Response by Bruce Schneier" Blackhat 2014

- G. Akerlof, "The Market for Lemons: Quality Uncertainty and the Market Mechanism"

**Key Takeaway: Security is a lemons market!**

- Prospect theory "As a species we are risk adverse when it comes to gains and risk taking when it comes to losses"

**Key Takeaway: We don't buy security products until it is too late!**

# A Philosophy of Defense

"Once you understand The Way broadly, you can see it in all things."

— Miyamoto Musashi, Book of Five Ring 1643

# A Philosophy of Defense

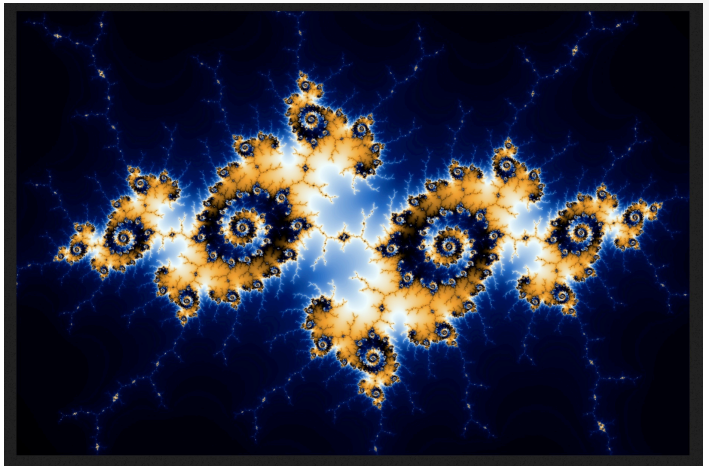
"Once you understand The Way broadly, you can see it in all things."
— Miyamoto Musashi, Book of Five Ring 1643

- Musashi was undefeated samurai (60 duels)
- Throughout the book, Musashi implies that the way of the Warrior, as well as the meaning of a "True strategist" is that of somebody who has made mastery of many art forms away from that of the sword…
- Such a philosophy is fractal – it has similar properties on many scales

# Fractal Defense

"Once you understand The Way broadly, you can see it in all things."
— Miyamoto Musashi, Book of Five Ring 1643

- A fractal is a natural phenomenon or a mathematical set that exhibits a repeating pattern that displays at every scale.
- Security is a combination of detection, protection and response
- Fractal defense: a philosophy for scaling detection, protection and response up and down the IT ecosystem



# Fractal Defense: Example

From our white paper “Defense at scale: Building a Central Nervous System for the SOC”

**Behavioral Indicator of Compromise (IOC)**

$$\mathbb{P}(W_1, \ldots, W_n) = \mathbb{P}(W_1) \prod_{i=2}^{n} \mathbb{P}(W_i | W_{i-1}, \ldots, W_1)$$

- Leverage expressive ways to target TTP’s (tactics, techniques and procedures) that are reusable and scalable across many use cases/ behaviors

- Can incorporate classical signatures into a probabilistic scoring mechanism

# Fractal Defense: Example

From our white paper “Defense at scale: Building a Central Nervous System for the SOC”

**Behavioral Indicator of Compromise (IOC)**

$$\mathbb{P}(W_1, \ldots, W_n) = \mathbb{P}(W_1) \prod_{i=2}^{n} \mathbb{P}(W_i | W_{i-1}, \ldots, W_1)$$

**Model 1.** *Let $W$ be a string of length $n$ such that $W$ is composed of characters $W_1, \ldots, W_n$ in increasing order. For example if $W =$ google then $W_1 = g, W_2 = o, W_3 = o, W_4 = g, W_5 = l, W_6 = e$. Apply the Bayes "chain rule" to get*

$$\mathbb{P}(W_1, \ldots, W_n) = \frac{\mathbb{P}(W_2|W_1)}{\mathbb{P}(W_2)\ldots\mathbb{P}(W_n)} \prod_{i=1}^{n} \left[ \lambda_1 \frac{1}{\#(w_i)} + \lambda_2 \frac{\#(w_{i-1}w_i)}{\#(w_{i-1})} + \lambda_3 \frac{\#(w_{i-2}w_{i-1}w_i)}{\#(w_{i-2}w_{i-1})} \right] \tag{1.1}$$

# Fractal Defense: Example

From our white paper “Defense at scale: Building a Central Nervous System for the SOC”

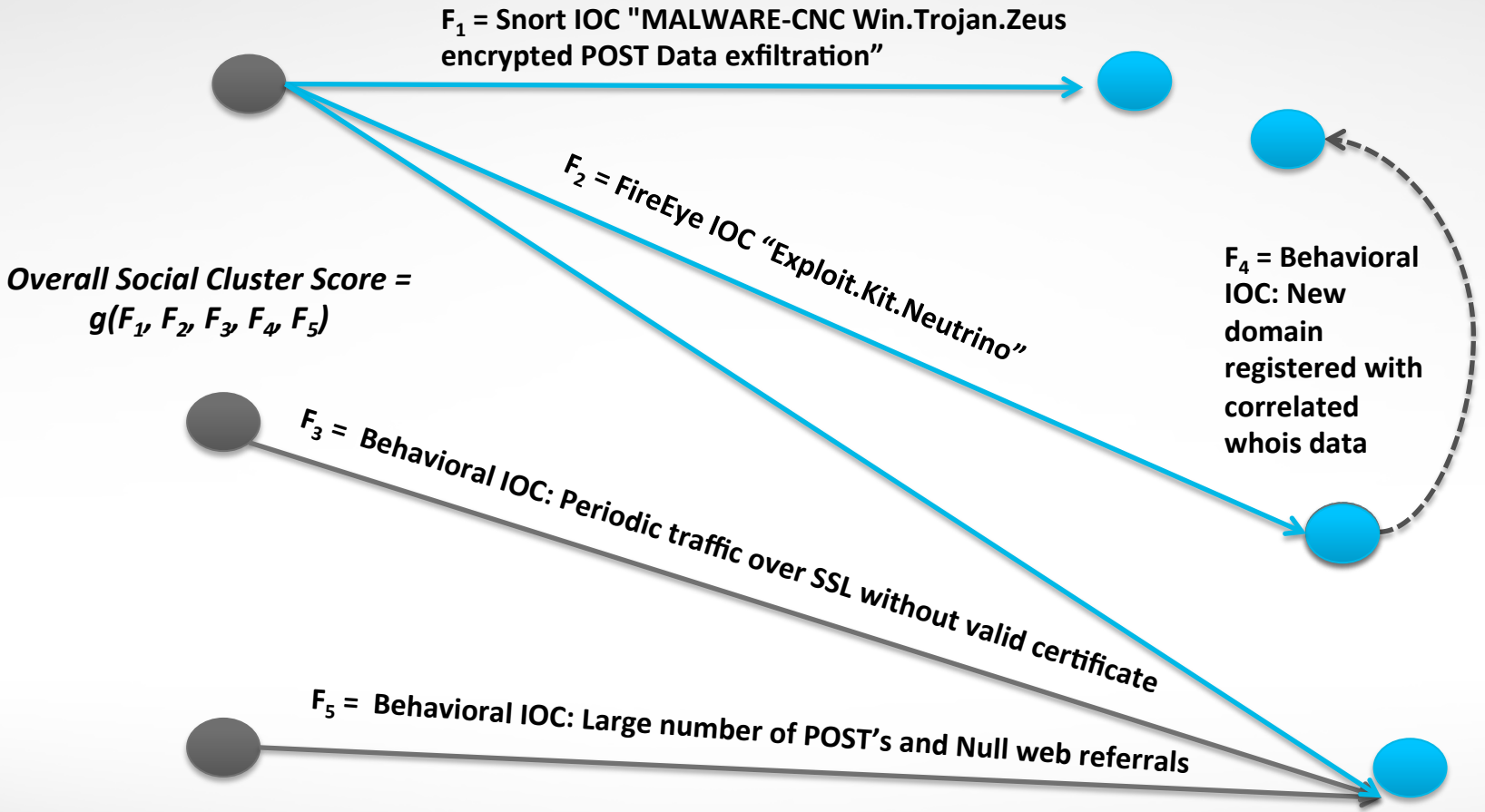
**Behavioral Indicator of Compromise (IOC)**

$$\mathbb{P}(W_1, \ldots, W_n) = \mathbb{P}(W_1) \prod_{i=2}^{n} \mathbb{P}(W_i | W_{i-1}, \ldots, W_1)$$

**Model 2.** *Let $W$ be any sequential data representing for instance an incoming stream of bytes or content types from proxy logs. For example let $W_1 = text/html$,*
*$W_2 = application/java\text{-}archive, W_3 = application/octet\text{-}stream$.*

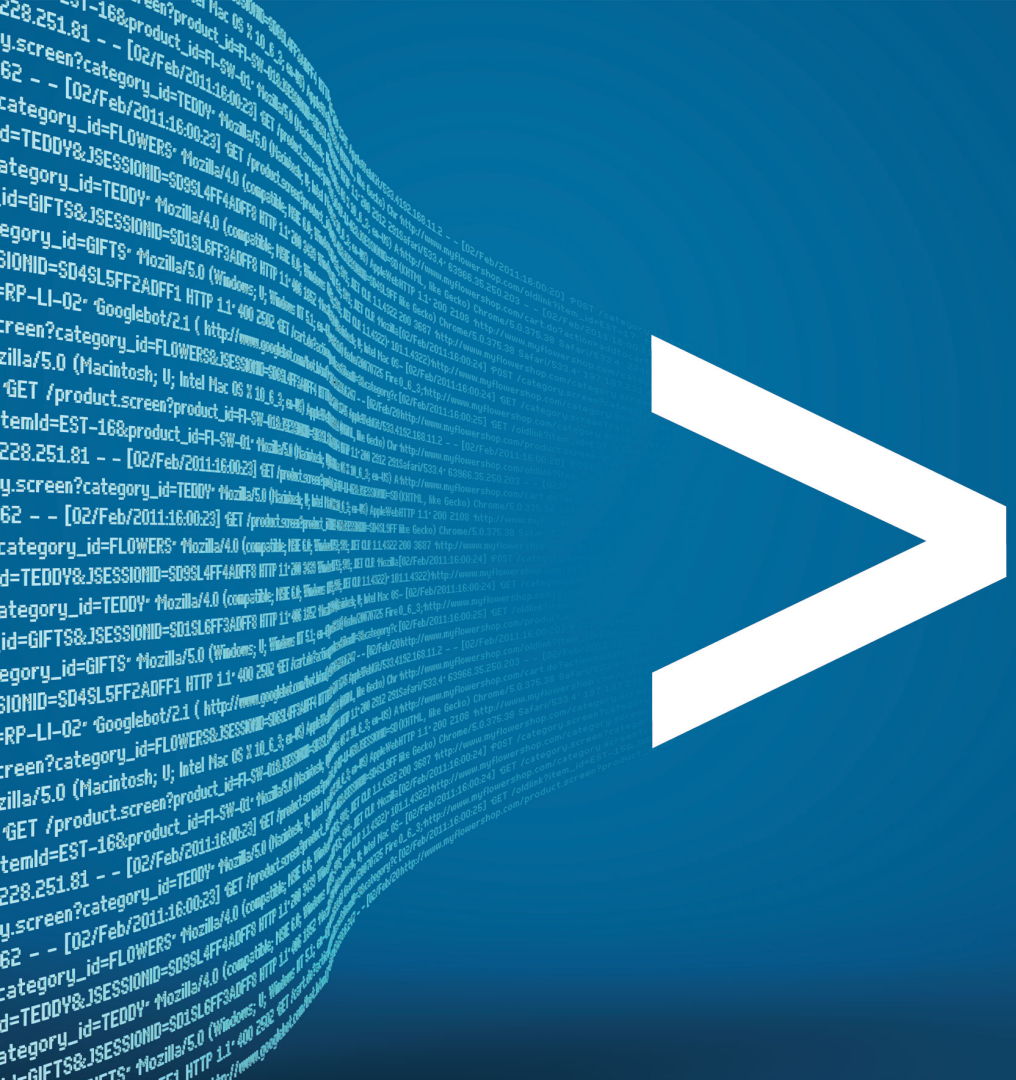
$$\mathbb{P}(W_1, \ldots, W_n) = \frac{\mathbb{P}(W_2|W_1)}{\mathbb{P}(W_2)\ldots\mathbb{P}(W_n)} \prod_{i=1}^{n} \left[ \frac{1}{2} \frac{1}{\#(w_i)} + \frac{1}{2} \frac{\#(w_{i-1}w_i)}{\#(w_{i-1})} \right] \tag{1.4}$$

# Central Nervous System Approach



# Sound Bytes

- **Fractal Defense**: Reuse logic (and code) across different security use cases. Make behavior based IOC's map to adversary tactics, techniques and procedures for better scalability.
- **Cybersecurity Analytics ROI**: Make security requirements functional by setting realistic benchmarks based on your own data
- **Lambda Architecture**: a generic problem solving system built on immutability and hybrid batch/real-time workflows



# Cybersecurity Analytics ROI

splunk>

# Cybersecurity Analytics

- Motivation
  - Number one mistake I see researchers making is modeling the "Continuum of behaviors" vs. modeling a discrete security use case
  - Help rank order use cases for management/researchers without security background (ranking should coincide with security expert intuition)

- Possible attack behaviors are infinite!
  - Intractable dimensionality
  - "Project the problem down to finitely many sub problems"

- Anomaly Detection != Actionable Intelligence

# Cybersecurity Analytics Roadmap

- Step 1: Make a grab bag of your favorite use cases/gaps of the threat surface
  - Model primitives: actionable units or single behaviors
- Step 2: Determine existing coverage and cost of impact per use case (APPROXIMATE! Unless you are have been logging costs of security events internally similar to MS…)
- Step 3: Build ROI Graph
  - What is your formula for ROI?
- Step 4: Rank Order
  - Rank by determining which ordering provides additional value to minimizing risk
  - Our example uses the added structure of LAN and WAN but you can complicate things further by trying to incorporate adversary capabilities, point solution metrics, etc…

# Enumerate Threat Surface

# Use Cases: Insider/LAN Threats

- **PtH/PtT**
- **Time of Day Model**
- **Lateral Reconnaissance**
- **Pop @ Risk**
- **Passive DNS**
- **Data Store Exfiltration**
- **Two Factor Attack**
- **Exploit Kits**
- **Crowd sourced Executable Classification**
- **MITM**
- **Telecommuter Ground Speed/ Triangulation**
- **Data mart reconnaissance/mapping**
- **VIP Asset Profiling**
- **User to Group Behavior Metrics**
- **User Access Pattern Models**
- **Shadow IT misconfiguration and gap profiling**
- **Beachhead/DMZ attack graph modeling**

# Enumerate Threat Surface

# Use Cases: External/WAN Threats

- **Web Referral Graph**
- **Time of Day Model**
- **Heartbeat Beacon Detection**
- **SSL Side Channel Analysis**
- **Watering Hole Analytic**
- **Passive DNS AI**
- **Predictive Blacklisting**
- **URL Relative Path Tokens**
- **Edit Distance Classification**
- **Exploit Kits Analytics**
- **Pseudo Random Domain Detection**
- **Executable Graph Classifier**
- **DNS Tunneling**

# Enumerate Threat Surface

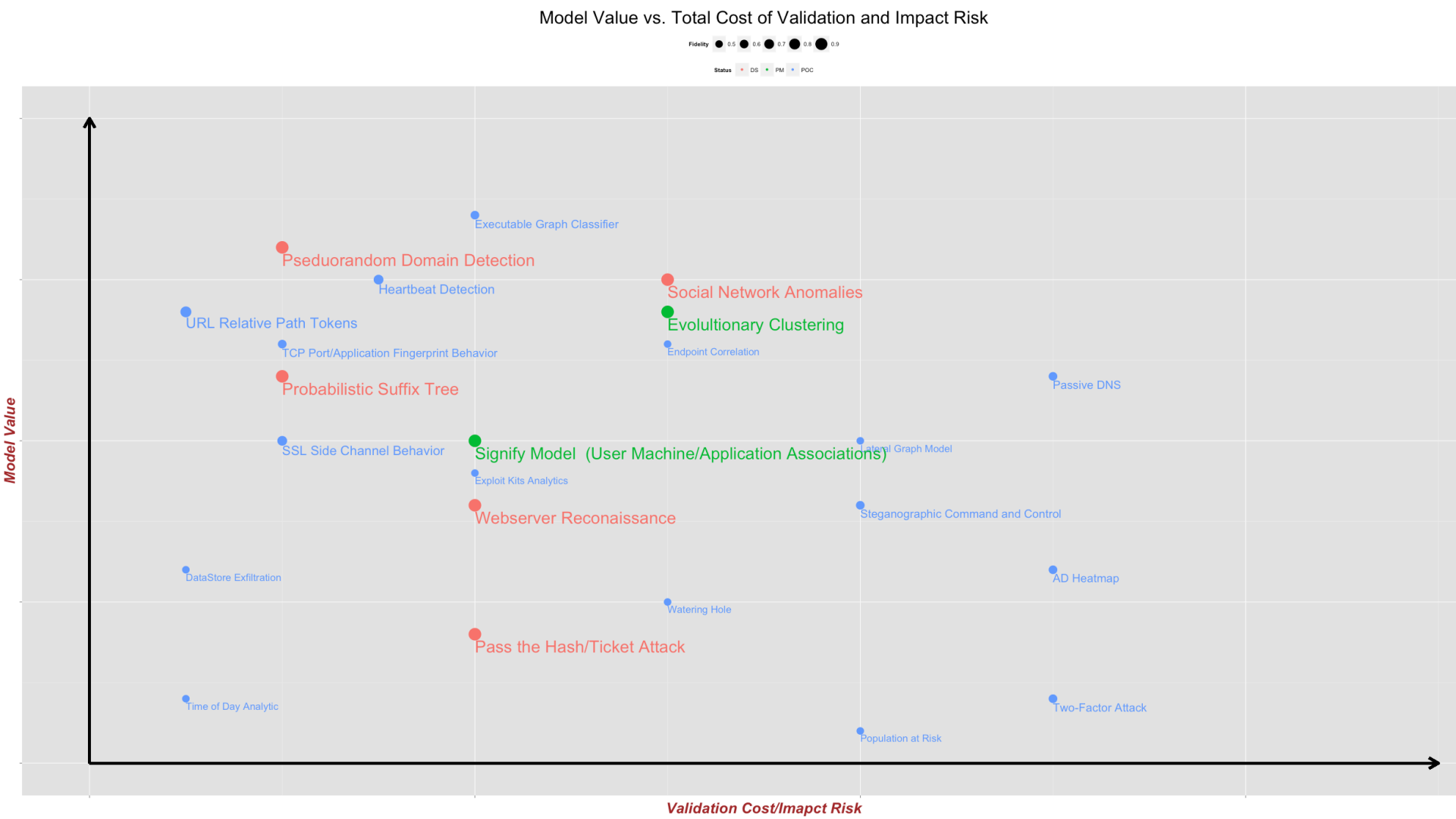
# Use Cases: IoT and Shadow IT

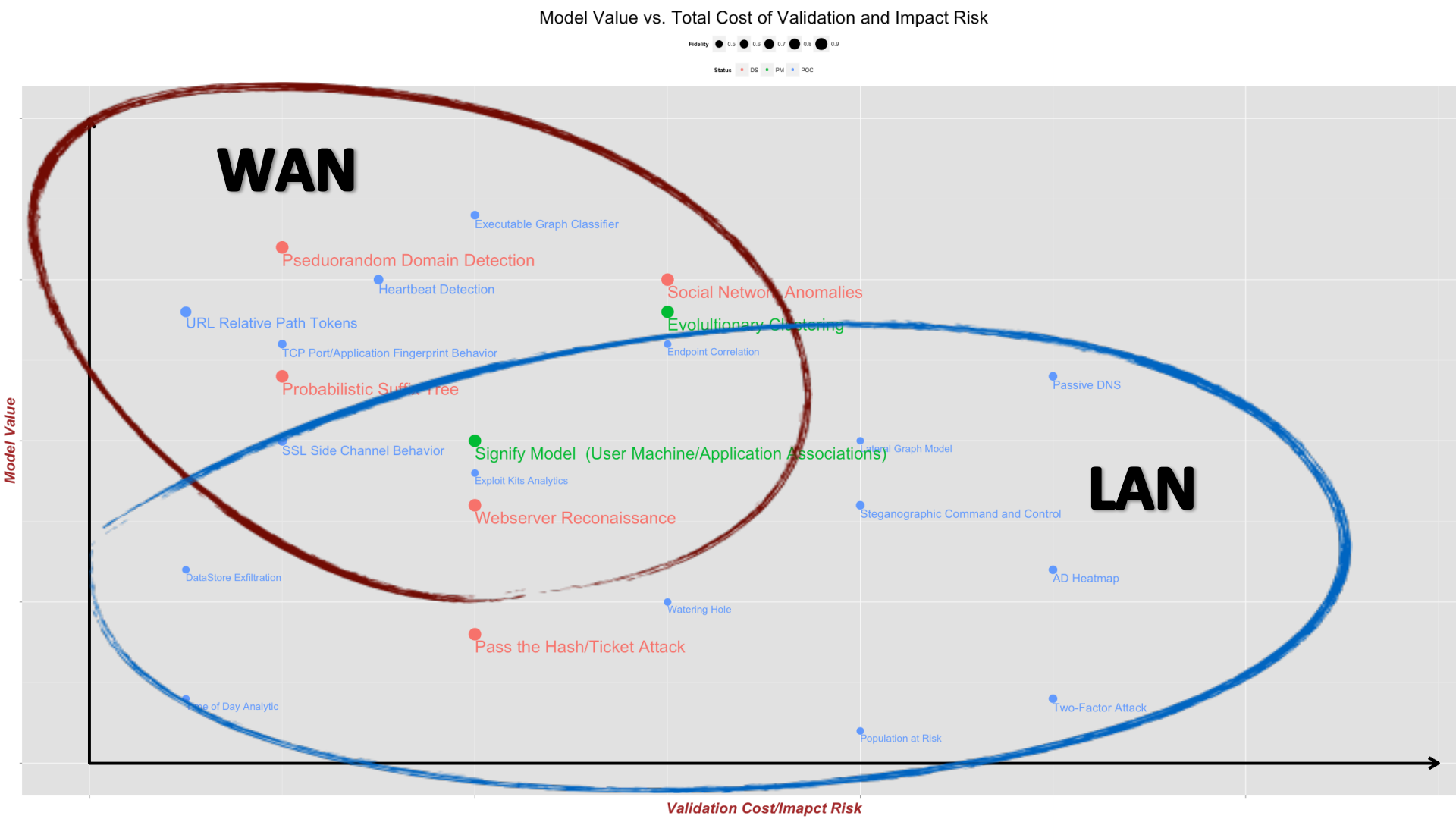
- **Embedded Systems Behavioral Profiles**
- **BYOD Population Analysis**
- **Passive mobile application classifier**
- **Mobile app store profiling**
- **Common environment baselines for mobile users**
- **Mobile Beacon Detection**
- **SSL Side Channel Analysis**
- **Credential compromise**
- **Rogue Device Detection**
- **HVAC Controller Attacks (BacNet)**

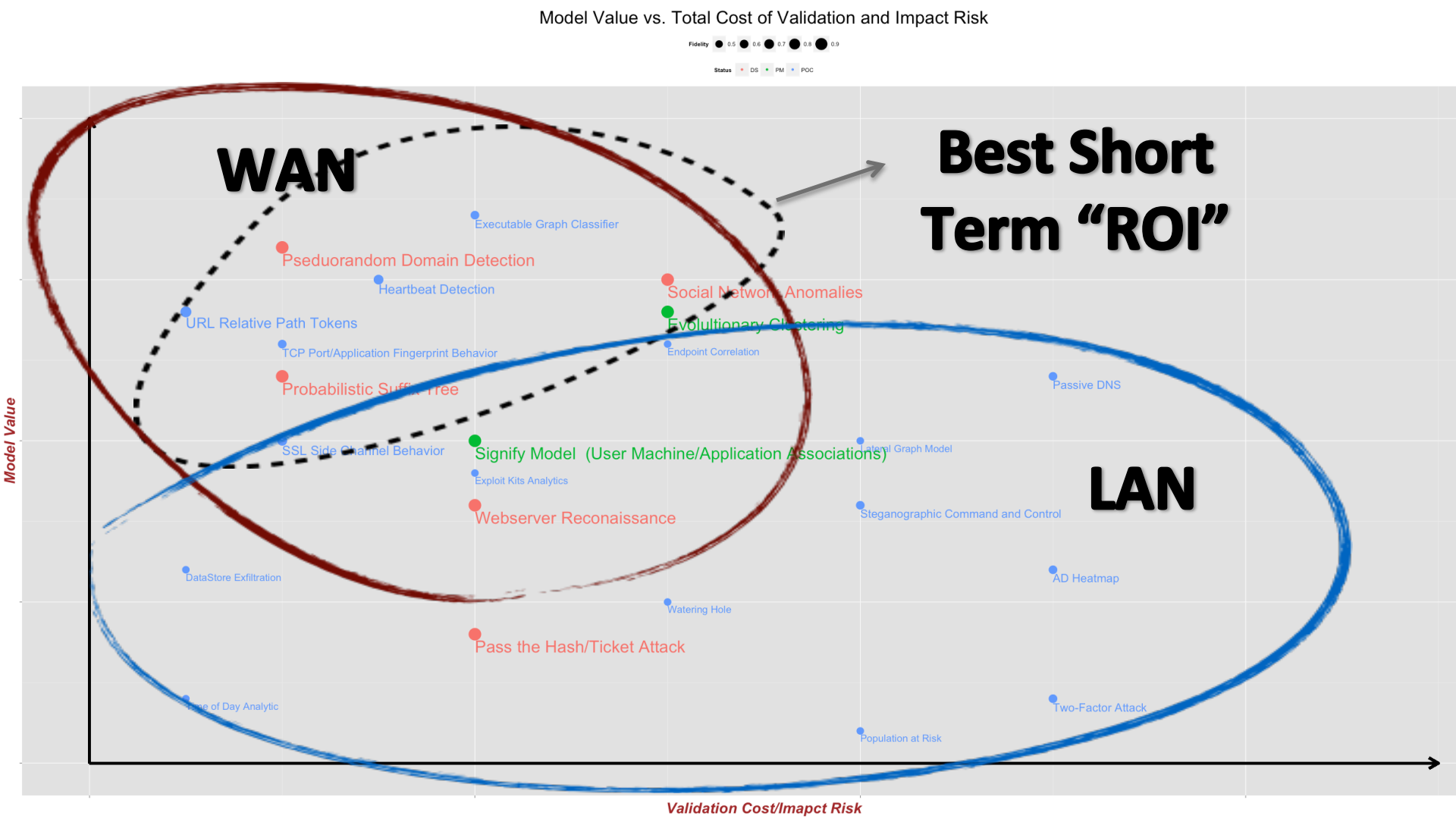
# Security Analytics ROI

- What is the intrinsic value of each model we build?
  - False Positive/True Positive ratios, AUC, etc.
  - Cost of Validating the Model
  - What is the risk to the organization for missing the threat?
  - Find Net New Threats!
- How to prioritize what analytic models to invest in
  - Dimensions: Impact Risk, Cost of Validation, Cost of Investment, Cost of Maintenance, Adversary Model

**$ CYBERSECURITY ROI =**
**$ IMPACT RISK – ($ COST OF INVESTMENT + $ COST OF VALIDATION)**

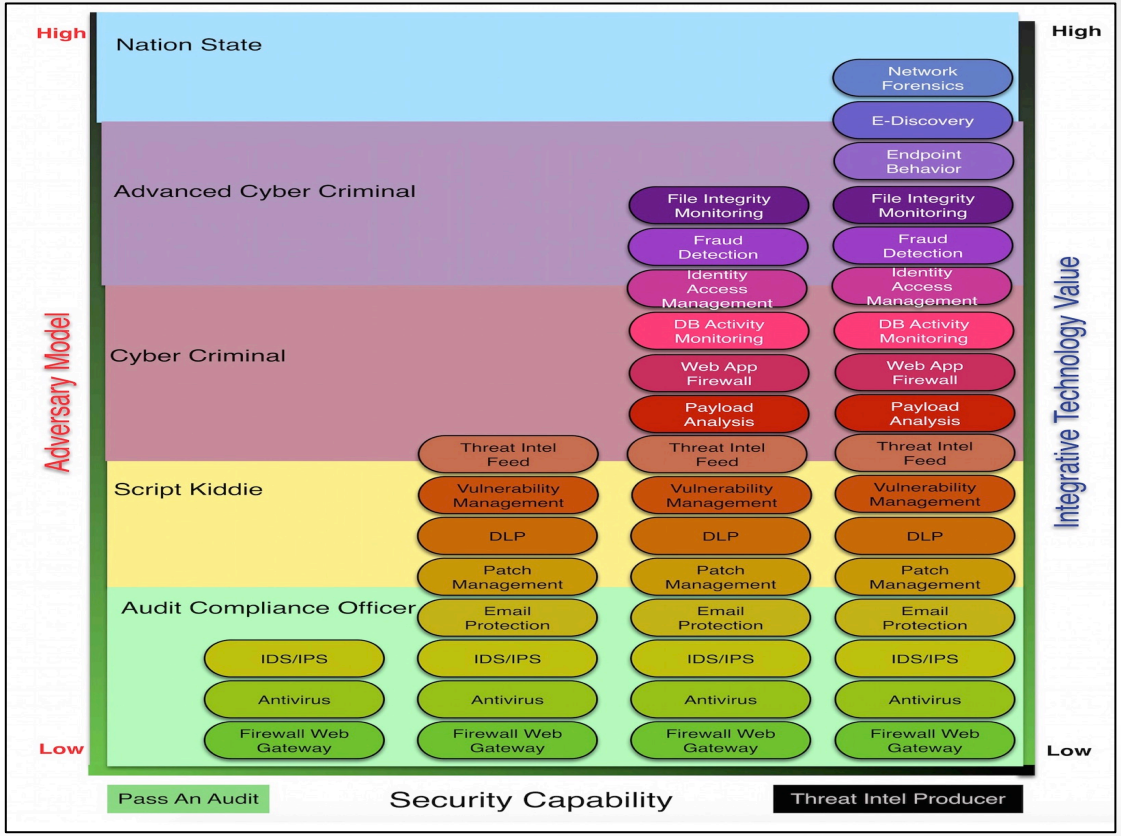

Model Value vs. Total Cost of Validation and Impact Risk
Fidelity 0.5 0.6 0.7 0.8 0.9
Status DS PM POC
Executable Graph Classifier
Pseduorandom Domain Detection
Heartbeat Detection
Social Network Anomalies
URL Relative Path Tokens
Evolultionary Clustering
TCP Port/Application Fingerprint Behavior
Endpoint Correlation
Probabilistic Suffix Tree
Passive DNS
SSL Side Channel Behavior
Signify Model (User Machine/Application Associations)
Exploit Kits Analytics
Webserver Reconaissance
Steganographic Command and Control
DataStore Exfiltration
AD Heatmap
Watering Hole
Pass the Hash/Ticket Attack
Time of Day Analytic
Two-Factor Attack
Population at Risk
Model Value
Validation Cost/Imapct Risk


Model Value vs. Total Cost of Validation and Impact Risk
Fidelity 0.5 0.6 0.7 0.8 0.9
Status DS PM POC
WAN
LAN
Executable Graph Classifier
Pseduorandom Domain Detection
Heartbeat Detection
Social Network Anomalies
URL Relative Path Tokens
Evolultionary Clustering
TCP Port/Application Fingerprint Behavior
Endpoint Correlation
Probabilistic Suffix Tree
Passive DNS
SSL Side Channel Behavior
Signify Model (User Machine/Application Associations)
Exploit Kits Analytics
Webserver Reconaissance
Steganographic Command and Control
DataStore Exfiltration
AD Heatmap
Watering Hole
Pass the Hash/Ticket Attack
Two-Factor Attack
Population at Risk
Model Value
Validation Cost/Imapct Risk


Model Value vs. Total Cost of Validation and Impact Risk
Fidelity 0.5 0.6 0.7 0.8 0.9
Status DS PM POC
WAN
Best Short Term "ROI"
LAN
Executable Graph Classifier
Pseduorandom Domain Detection
Heartbeat Detection
Social Network Anomalies
URL Relative Path Tokens
Evolutionary Clustering
TCP Port/Application Fingerprint Behavior
Endpoint Correlation
Probabilistic Suffix Tree
Passive DNS
SSL Side Channel Behavior
Signify Model (User Machine/Application Associations)
Exploit Kits Analytics
Webserver Reconaissance
Steganographic Command and Control
DataStore Exfiltration
AD Heatmap
Watering Hole
Pass the Hash/Ticket Attack
Two-Factor Attack
Population at Risk
Model Value
Validation Cost/Imapct Risk

# Adversary Capabilities and the Threat Surface

- Attacker capability vs. Security Capability is an important dimension to consider when prioritizing new solutions/ analytics
- Try to handle the low hanging fruit to more complex adversary behavior by road mapping custom analytics based on gaps in existing and future technology solutions
- It is a mistake to model the most complex adversaries first



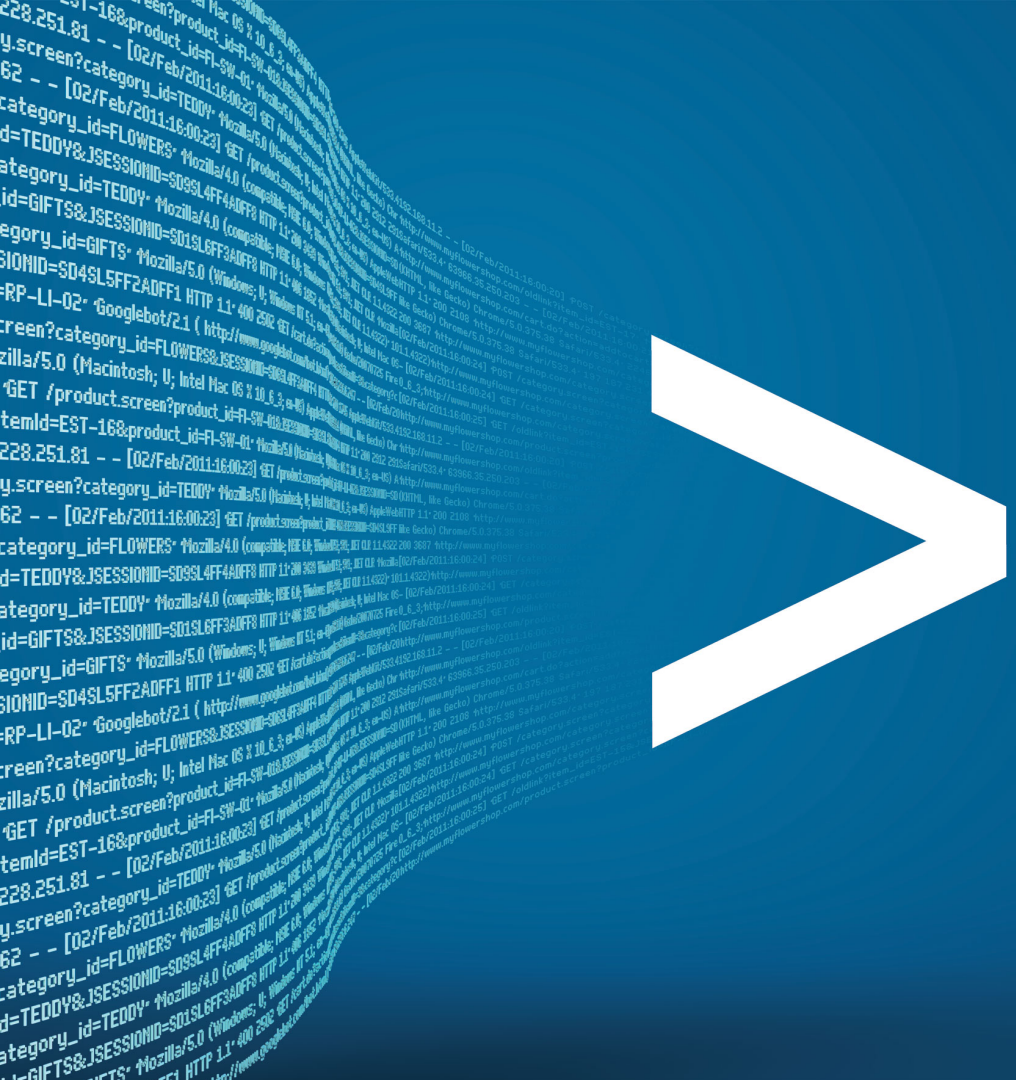
# Nextgen Benchmarks

- DARPA, Predict.org spearheading the collection and annotation of complex data sets for security research
  - Skaion 2006 DARPA Dataset
  - Contagio Malware Dump
  - CTU University: CTU-13 Dataset. A Labeled Dataset with Botnet, Normal and Background traffic

- Evidence Collection and the SOC
  - Most important workflow that is missing from large scale Intel/telemetry sharing across organizations.,
- Public Repos / OSINT

# Functional Requirements!

- Real problem in security is requirements are non-functional
- Benchmark next gen product by isolating the sub problems and holding specific metrics accountable to real world data
- How do you rank order the value of a cybersecurity analytic?



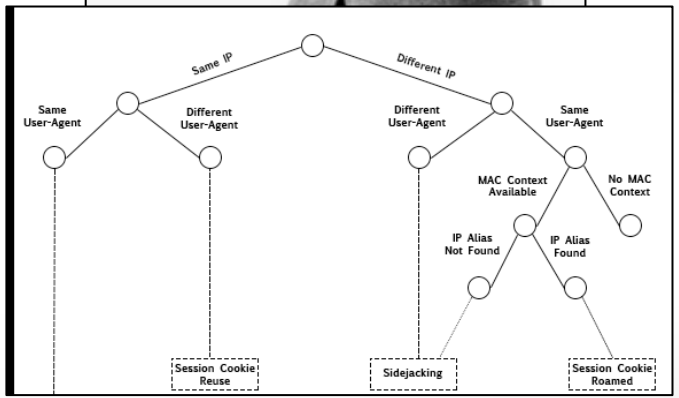
# Defensive Architectures

splunk>

# Why Build A Defensive Tool?

- Incident Response Is Hard Work! What can we automate?

**A security analyst is an oracle whose input is evidence and whose output is True Positive, False Positive, True Negative or False Negative**




  - The list of possible questions is large but typically the flow is a type of decision tree for example

# Why Build A Defensive Tool?

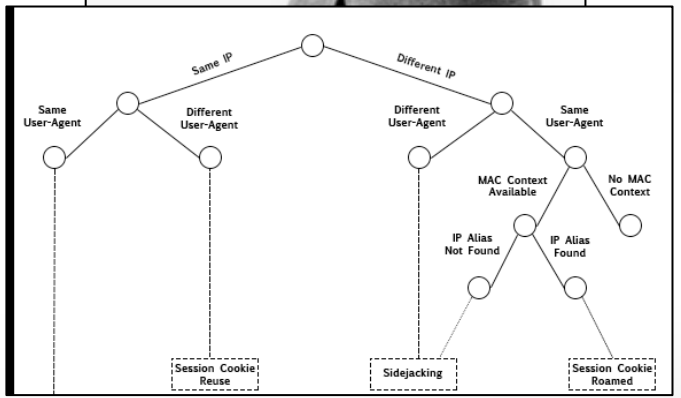
Security Oracle Workflow

**Example 1:**

Evidence => Periodic Communication => LAN to WAN Data =>WAN URL has Bad Reputation => Correlate with VT => True Positive

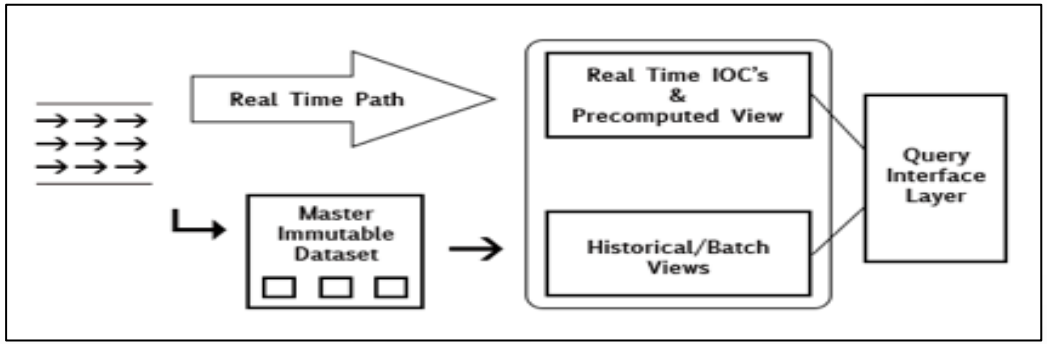
**Example 2:**

Evidence => Potential C2 Domain => LAN to WAN Data => WAN URL is new Google IP => False Positive



# Lambda Security

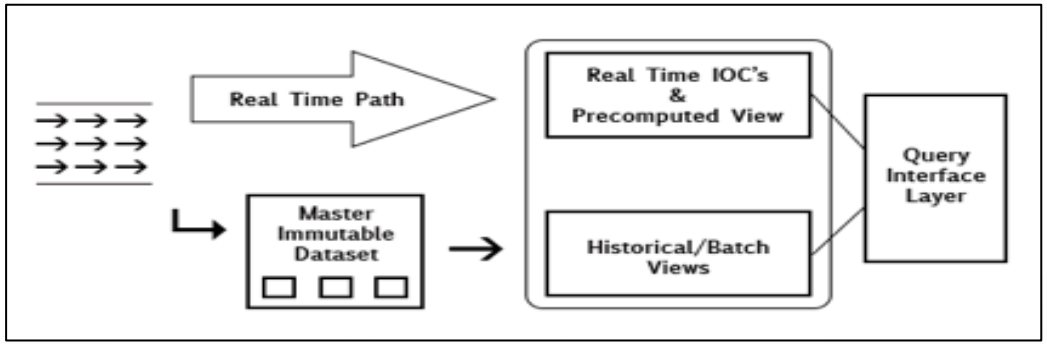
- Lambda Architecture: batch + real time computing paradigm
- Minimizes the complexity in historical computations overcoming bottlenecks SOC has experienced operating first gen SIEMs
- Data model that is append-only, distributed and immutable is optimized for security centric workflows and analyst queries



# Lambda Security

- Architecture is described by three simple equations:

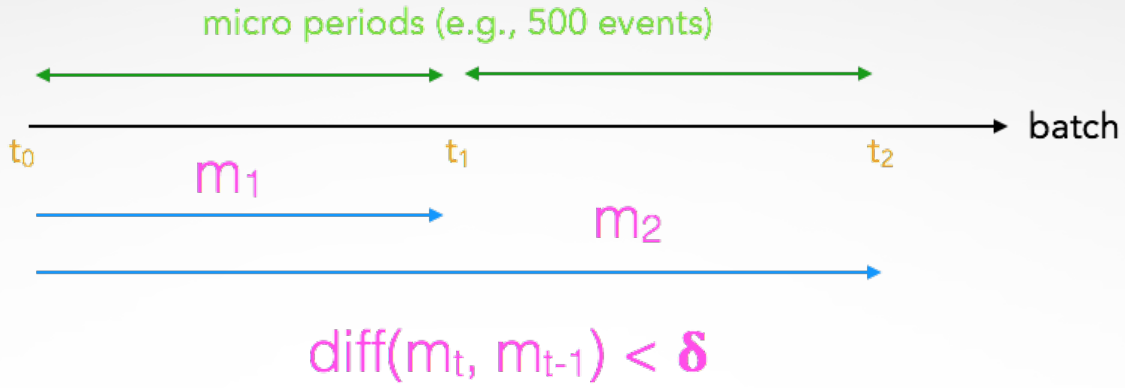
**batch view = function(all data)**
**realtime view = function(realtime view, new data)**
**query = function(batch view, realtime view)**



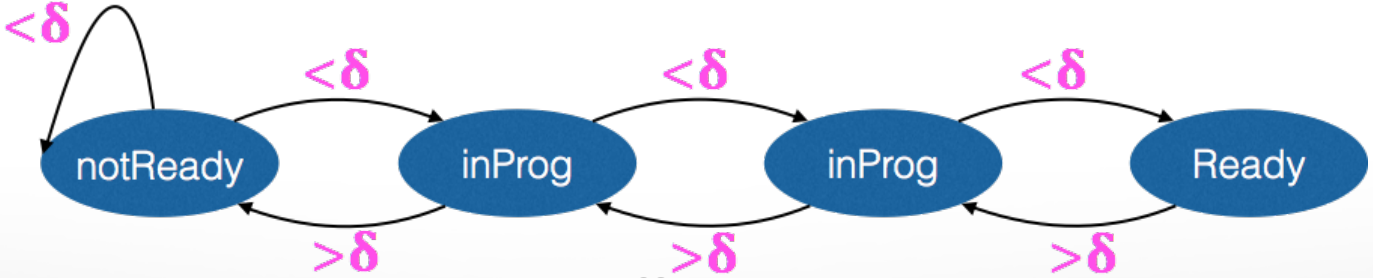
# Lambda Security

- Lambda architecture provides a design paradigm for a scalable central nervous system for the SOC whose components include
  - Machine learning based ETL(Extract/Transform/Load)
  - Distributed crawlers
  - Automated identity/session resolution and fingerprinting
  - Formal evidence collection protocol for automated labeling of incident response data
  - Analytics Metrics and establishing benchmarks for heterogeneous data

# When is a model ready?




A model stabilization algorithm:



# Lambda Firewalls?!

Manage the paths accordingly start building lambda workflows into Everything!!!

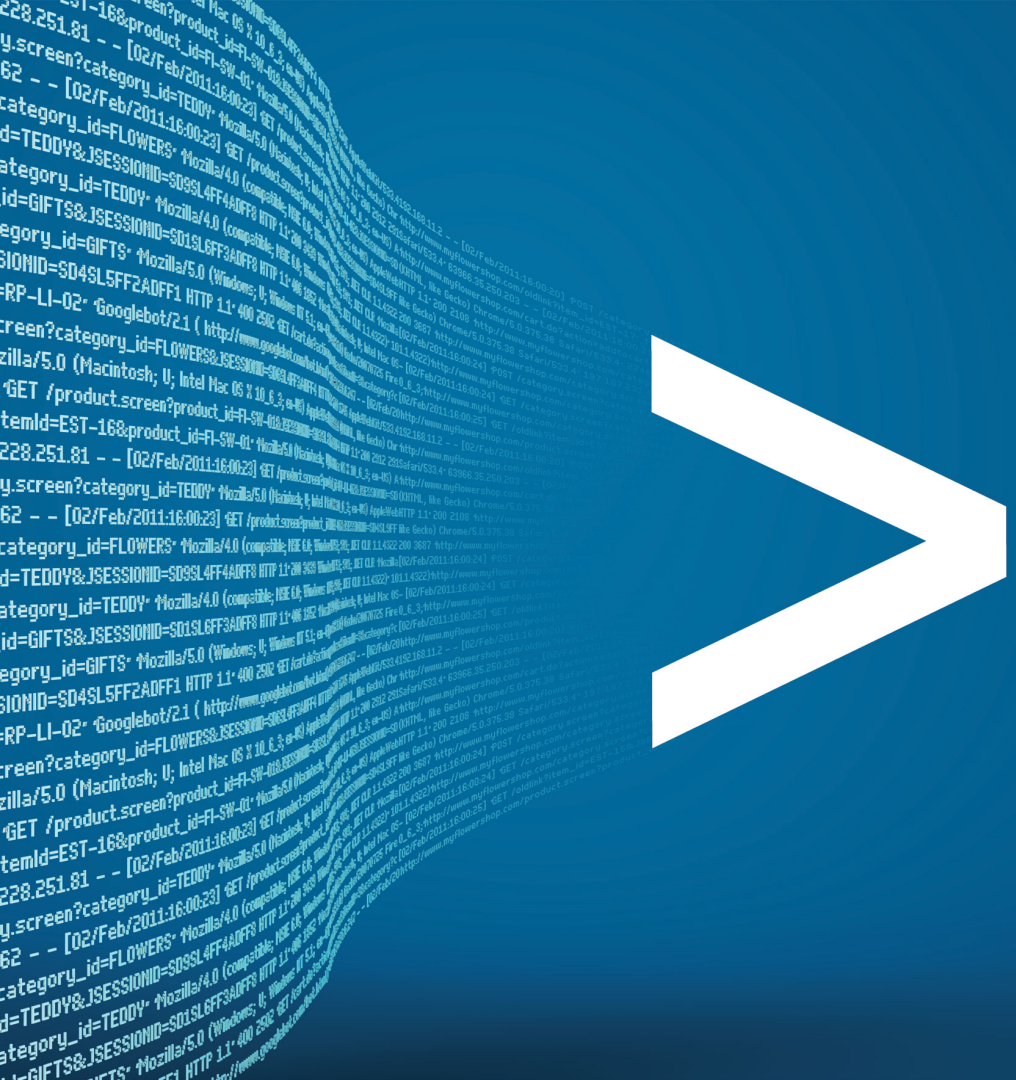
- Lambda firewall
  - Statistical whitelist computation aspect (fuzzy ACL's)
  - Path for signatures and sequential behaviors that is more expressive than PCRE
- Central nervous system approach to blending signals
  - Defense should scale up and down the size of organization: a properly engineered central nervous system should be able to protect SMB market as well as large scale deployments

# Complexity Class P-Complete and NC

- The Complexity Class P-Complete and NC
  - NC => parallelizable
- Some problems don't parallelize well!!
  - P-Complete => Inherently Sequential
  - Any problem where you have to maintain state across nodes: Circuit Value Problem, Linear programming
  - Streaming models are usually harder to maintain than batch models

In complexity theory, the notion of **P-complete** decision problems is useful in the analysis of both:

1. which problems are difficult to parallelize effectively, and;
2. which problems are difficult to solve in limited space.



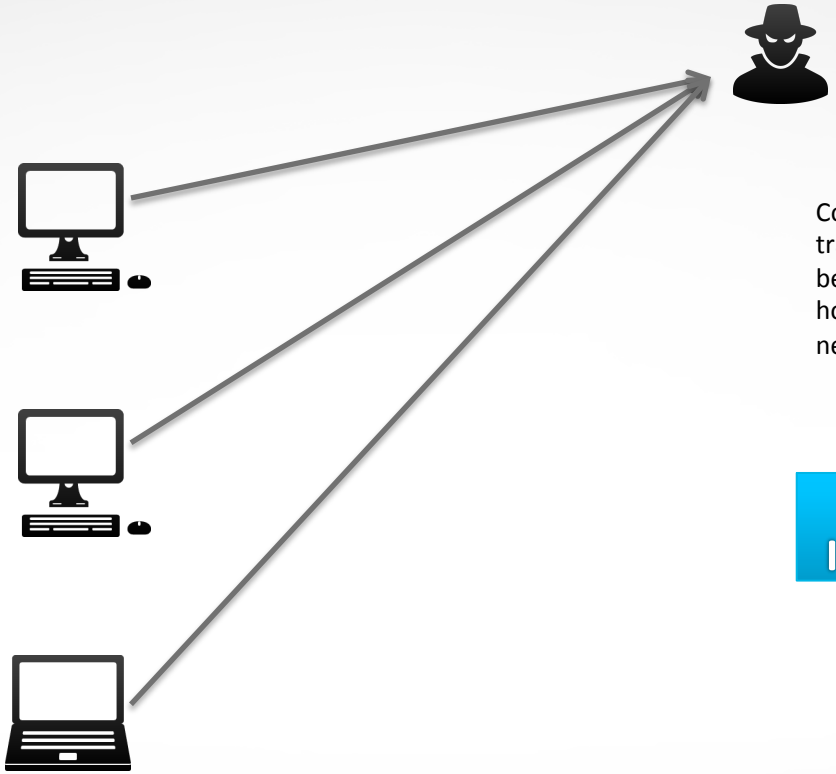
# Behavioral Intrusion Detection: Next Gen Signatures

splunk>

# Cybersecurity and Graph Mining

- Dynamic Temporal Graphs
  - Social Network of Communications forms a dynamic graph that evolves over time
  - Given a graph structure we can leverage state of the art graph mining techniques to detect anomalous graph patterns
    - Anomalous Clicks
    - Rare Sub-Structures
    - Rare Paths
- Anomalies in graphs can be easy to identify algorithmically
  - PageRank
  - Graph Cut/Partitioning
  - Random Walk Driven Label Propagation

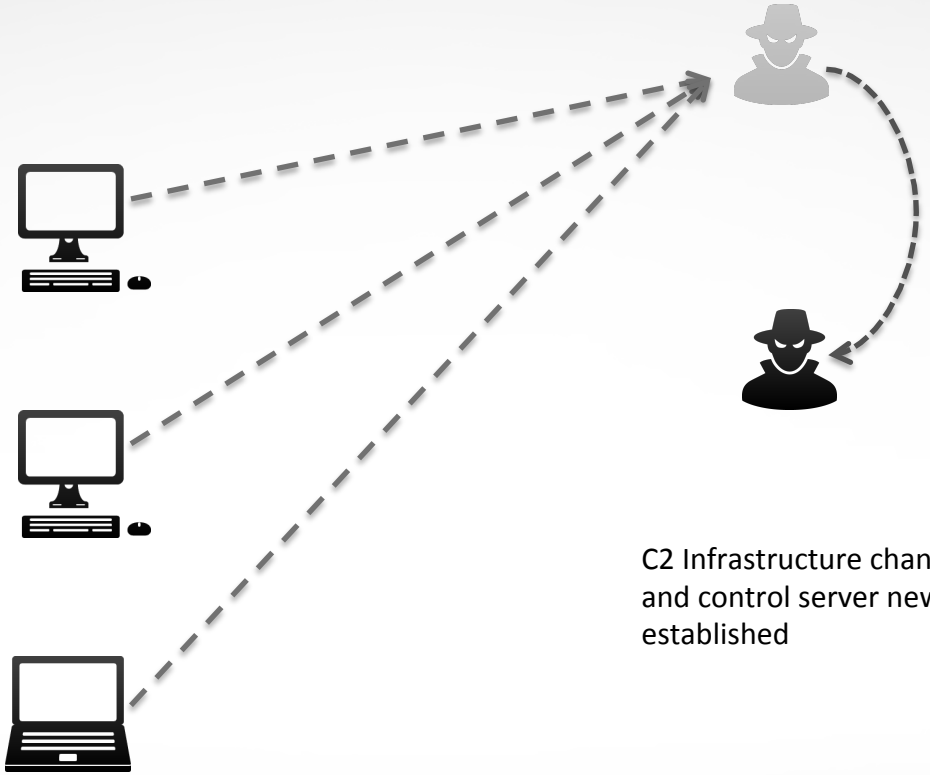
# Behavioral IOC: Mobile C2



Command and Control (C2) traffic has been established between compromised hosts inside the corporate network and C2 servers

**Time t_0**
**IP: 66.253.41.67**

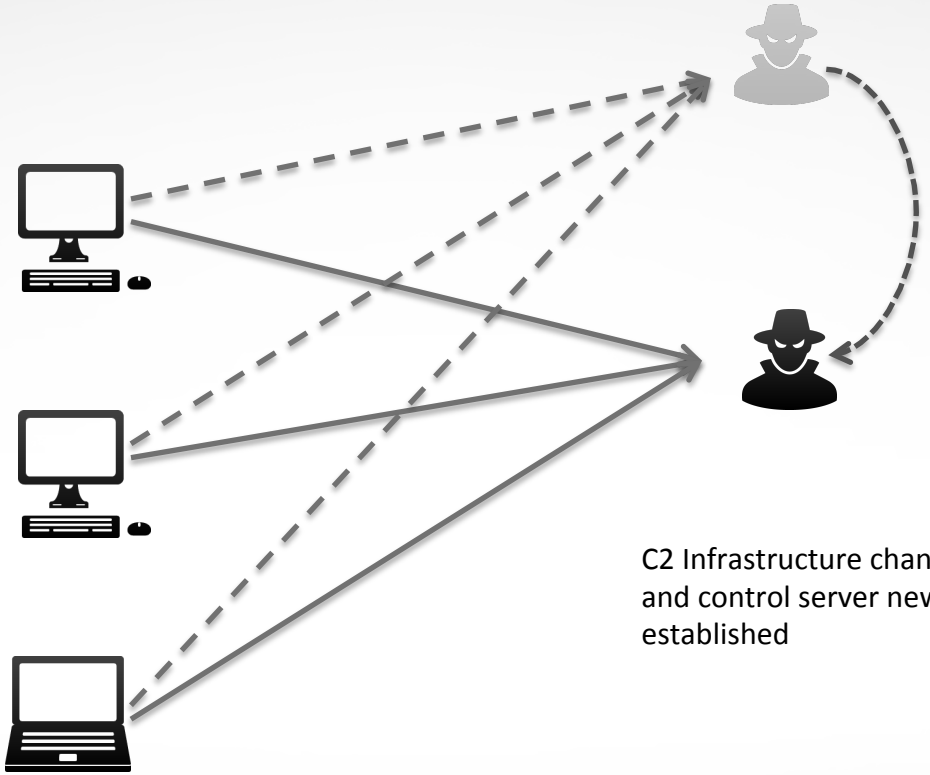
# Behavioral IOC: Mobile C2



**Time t_1**
**IP: 47.99.1.63**

C2 Infrastructure changes locations of command and control server new communication path is established

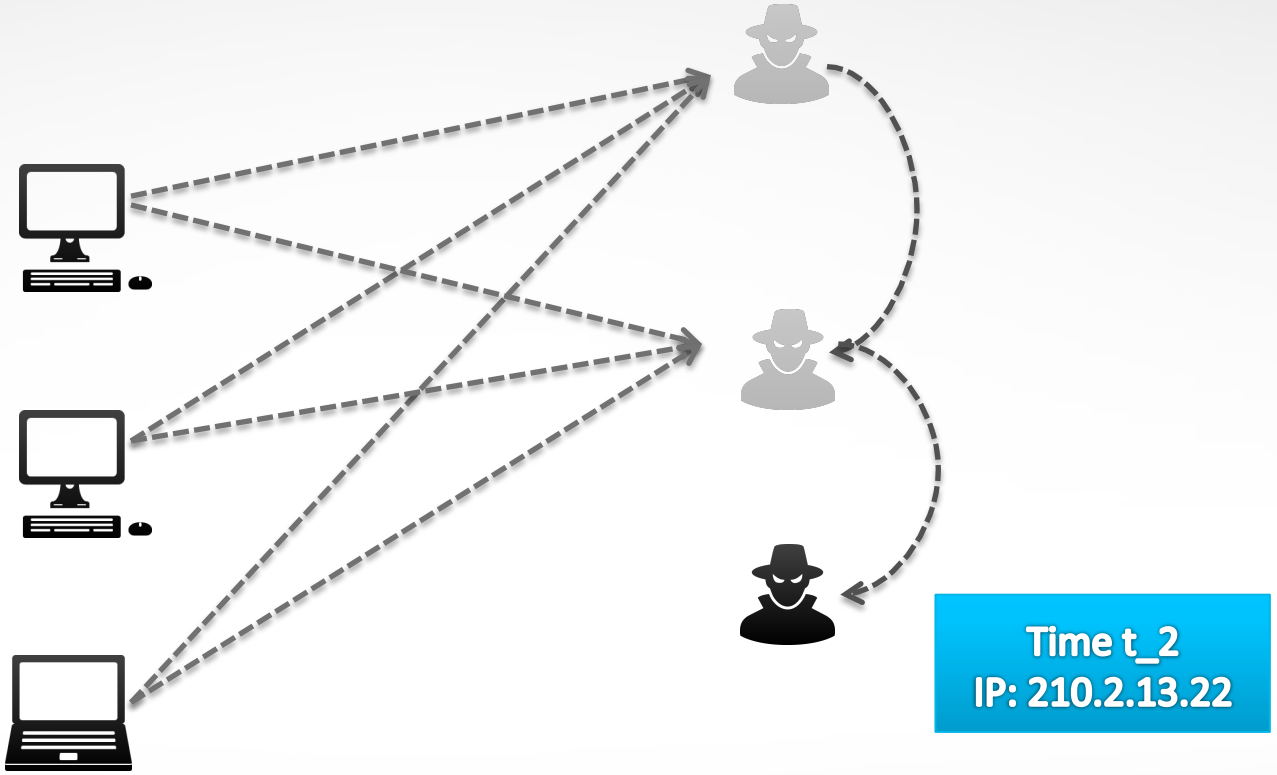
# Behavioral IOC: Mobile C2



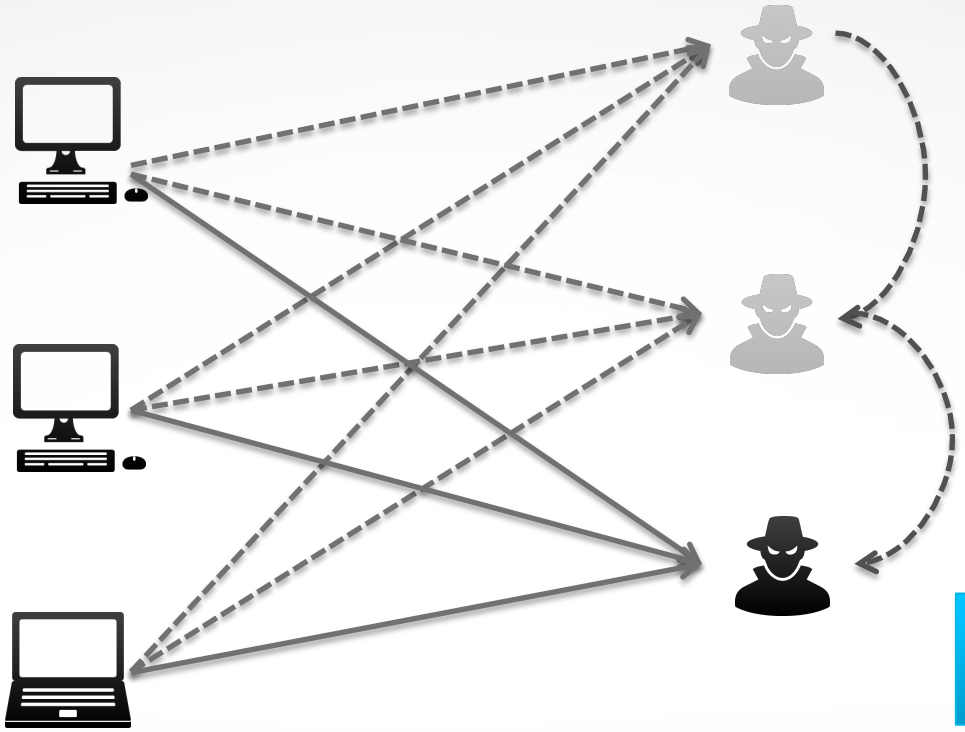
**Time t_1**
**IP: 47.99.1.63**

C2 Infrastructure changes locations of command and control server new communication path is established

# Behavioral IOC: Mobile C2



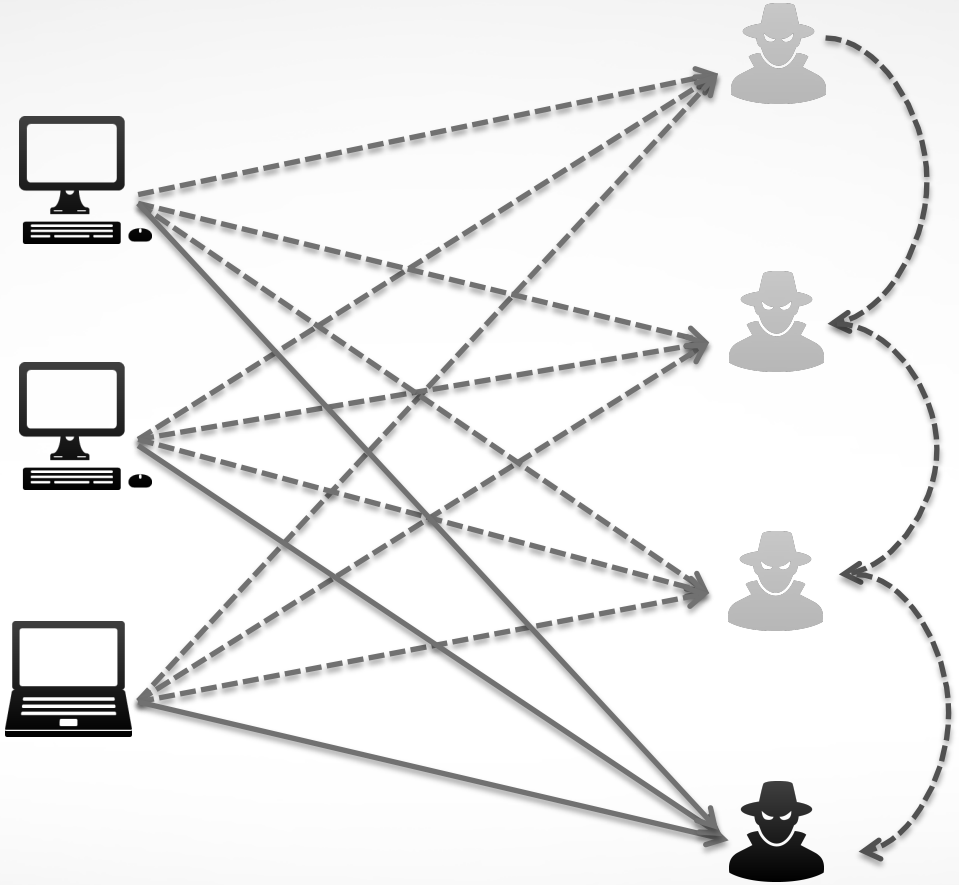
# Behavioral IOC: Mobile C2



At each time step (typically a day or two) the C2 Infrastructure changes locations of command and control via this “Fluxing” behavior. A subset of these type of graph patterns is known as “Fast Fluxing”

**Time t_2**
**IP: 210.2.13.22**

# Behavioral IOC: Mobile C2



The constant mobility of command and control infrastructure will continue this IP/Domain fluxing movement until detected

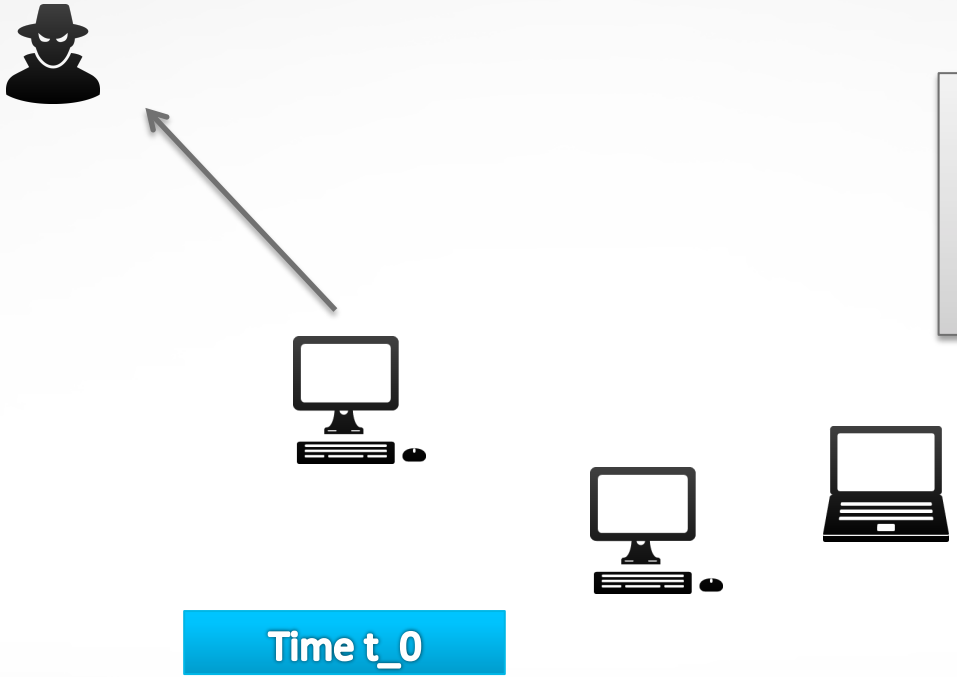
**Time t_n**
**IP: 82.21.4.6**

# Behavioral IOC: Perimeter Pivot

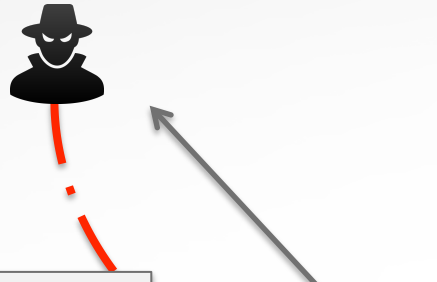
Command and Control (C2) traffic has been established between “Beachhead” and command and control operator

Time t_0

# Behavioral IOC: Perimeter Pivot



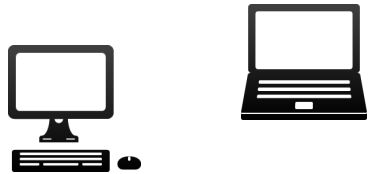
# Behavioral IOC: Perimeter Pivot



| Destination | Protocol | Length | Info |
|---|---|---|---|
| 37.59.5.67 | HTTP | 260 | POST /sears/sears.php HTTP/1.0 |
| 37.59.5.67 | HTTP | 260 | POST /sears/sears.php HTTP/1.0 |
| 37.59.5.67 | HTTP | 260 | POST /sears/sears.php HTTP/1.0 |

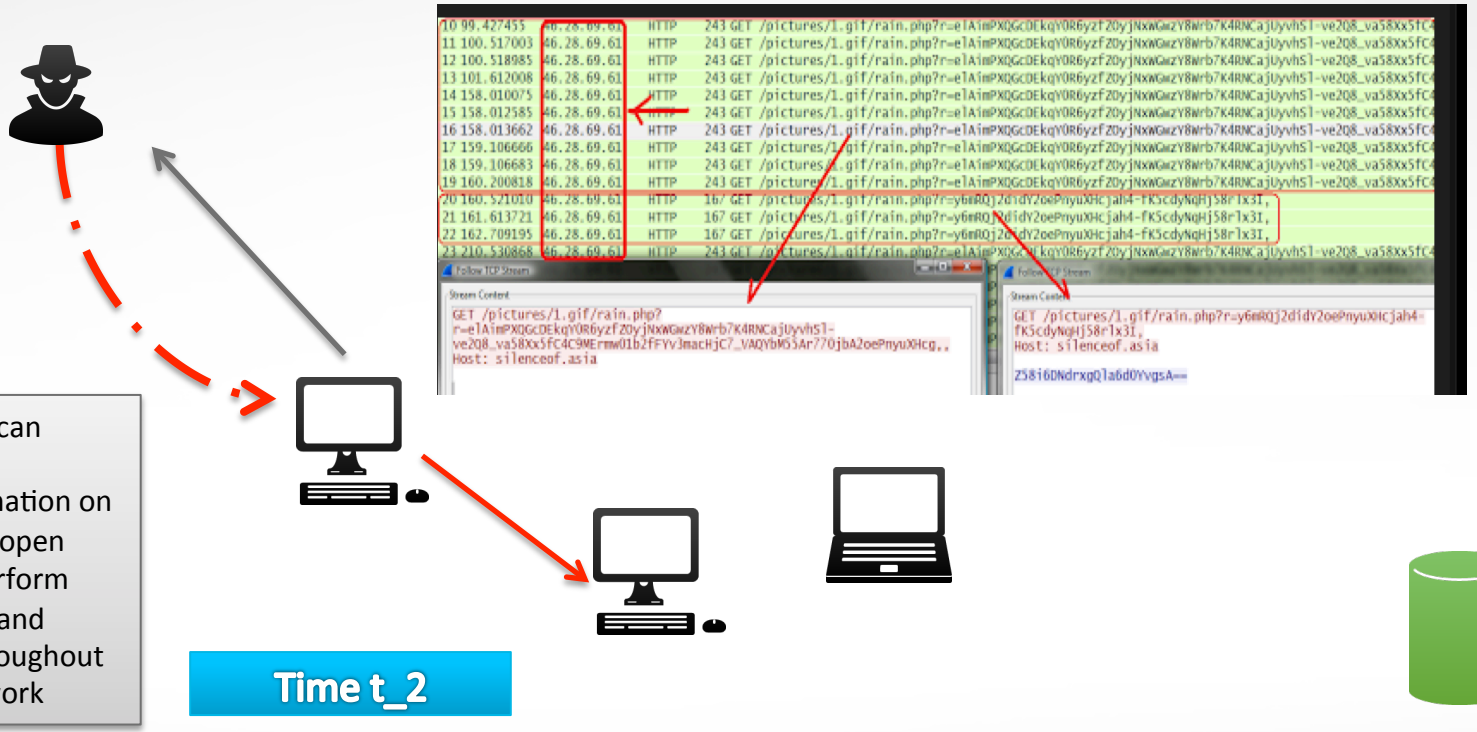
Obfuscated instructions get returned through an **Upstream** conversation embedded in PHP, .js, Flash, etc..

Commands obfuscated in this way can be through of as a hidden "Downstream Beacon"



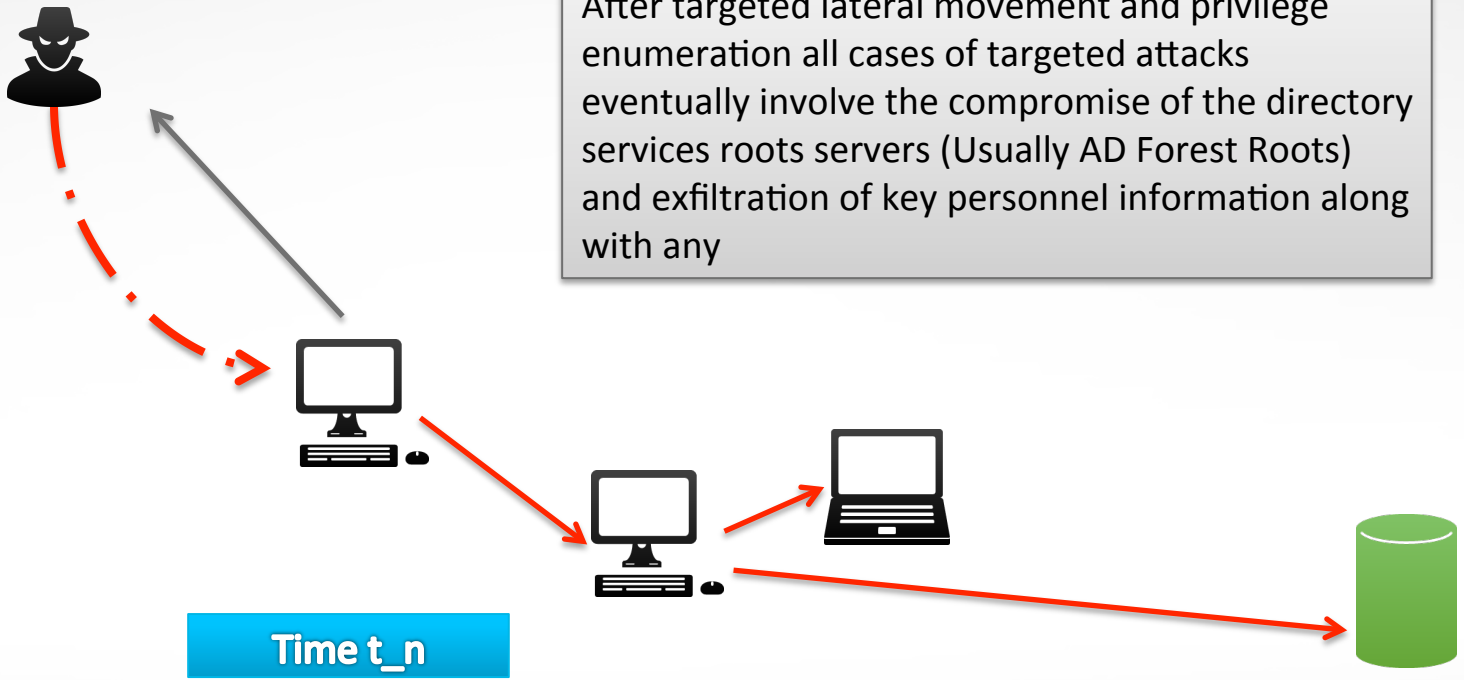
Time t_1

# Behavioral IOC: Perimeter Pivot

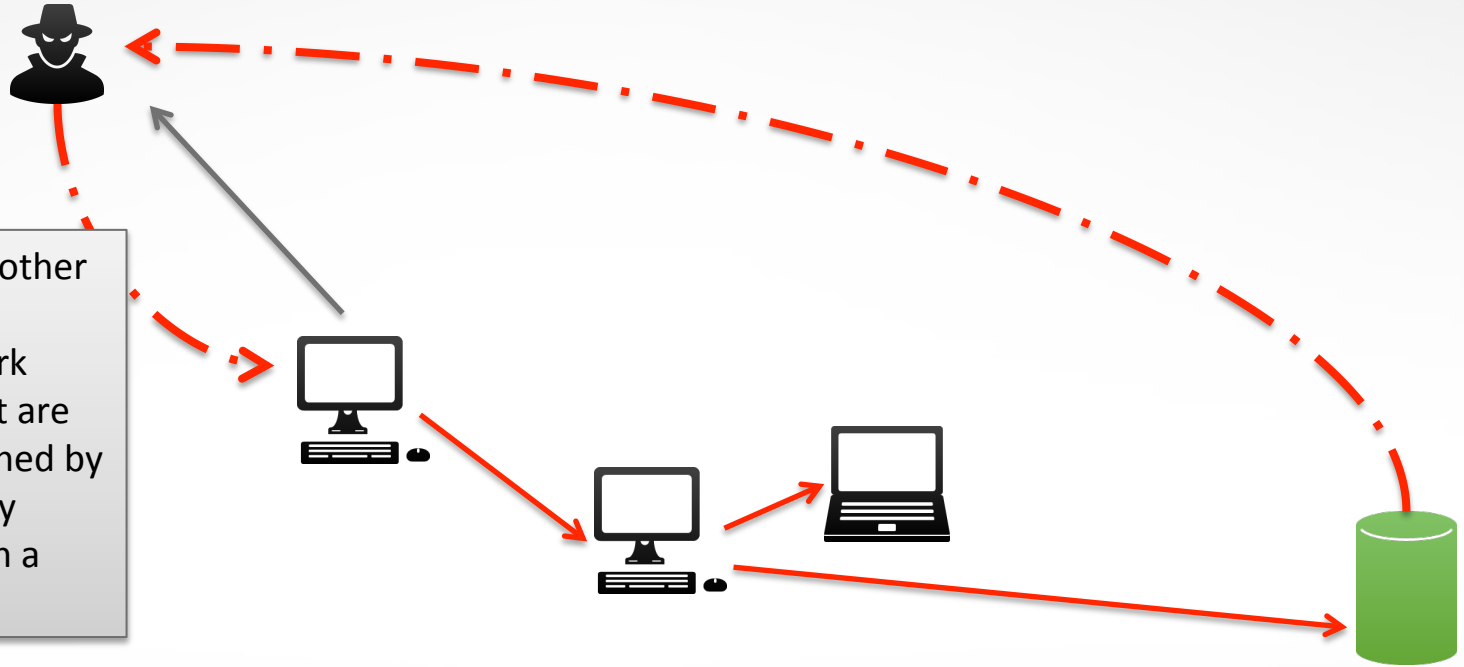



Embedded commands can signal infected asset to enumerate local information on the machine, attach to open network shares and perform lateral reconnaissance and privilege escalation throughout the compromised network

# Behavioral IOC: Perimeter Pivot



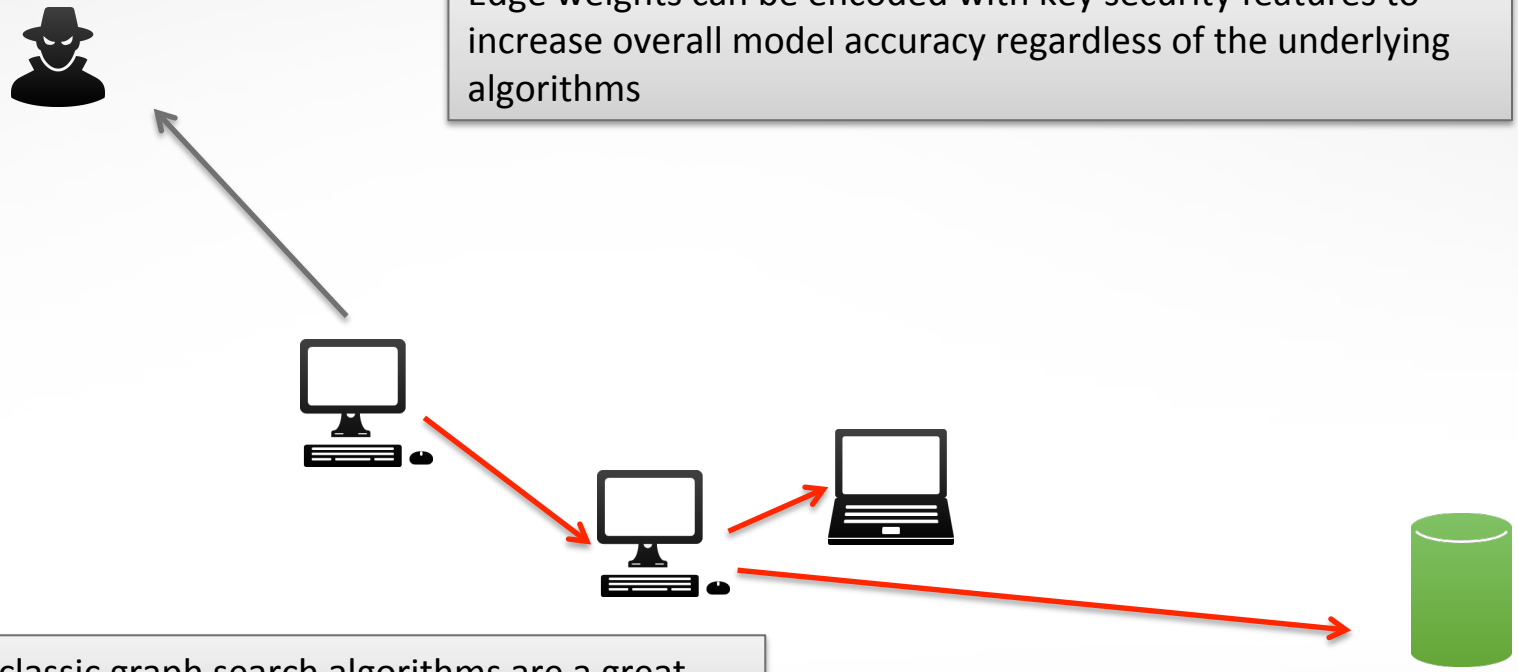
# Behavioral IOC: Perimeter Pivot



Exfiltration and other patterns have different network components but are usually constrained by the pictures they make as paths in a graph…

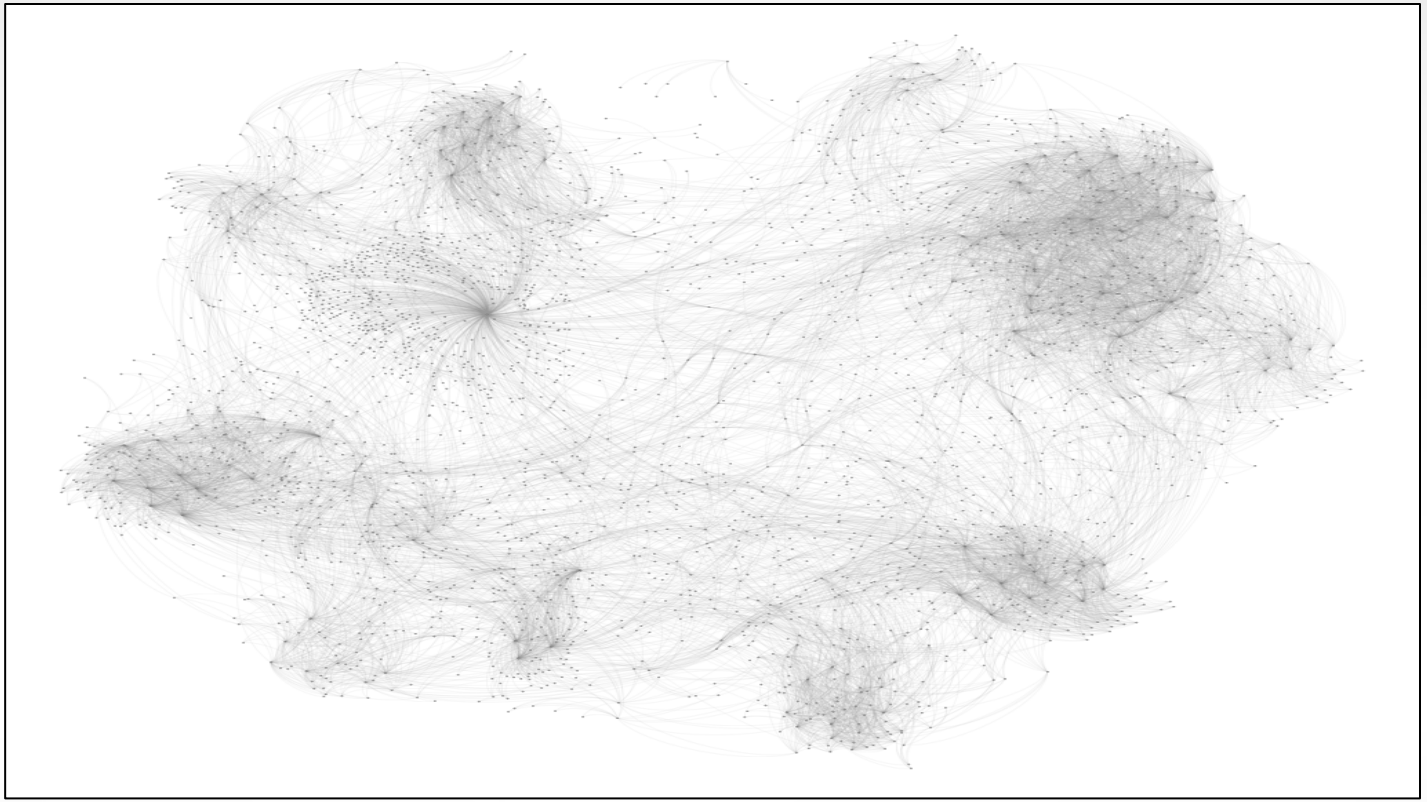
# Behavioral IOC: Perimeter Pivot

Edge weights can be encoded with key security features to increase overall model accuracy regardless of the underlying algorithms

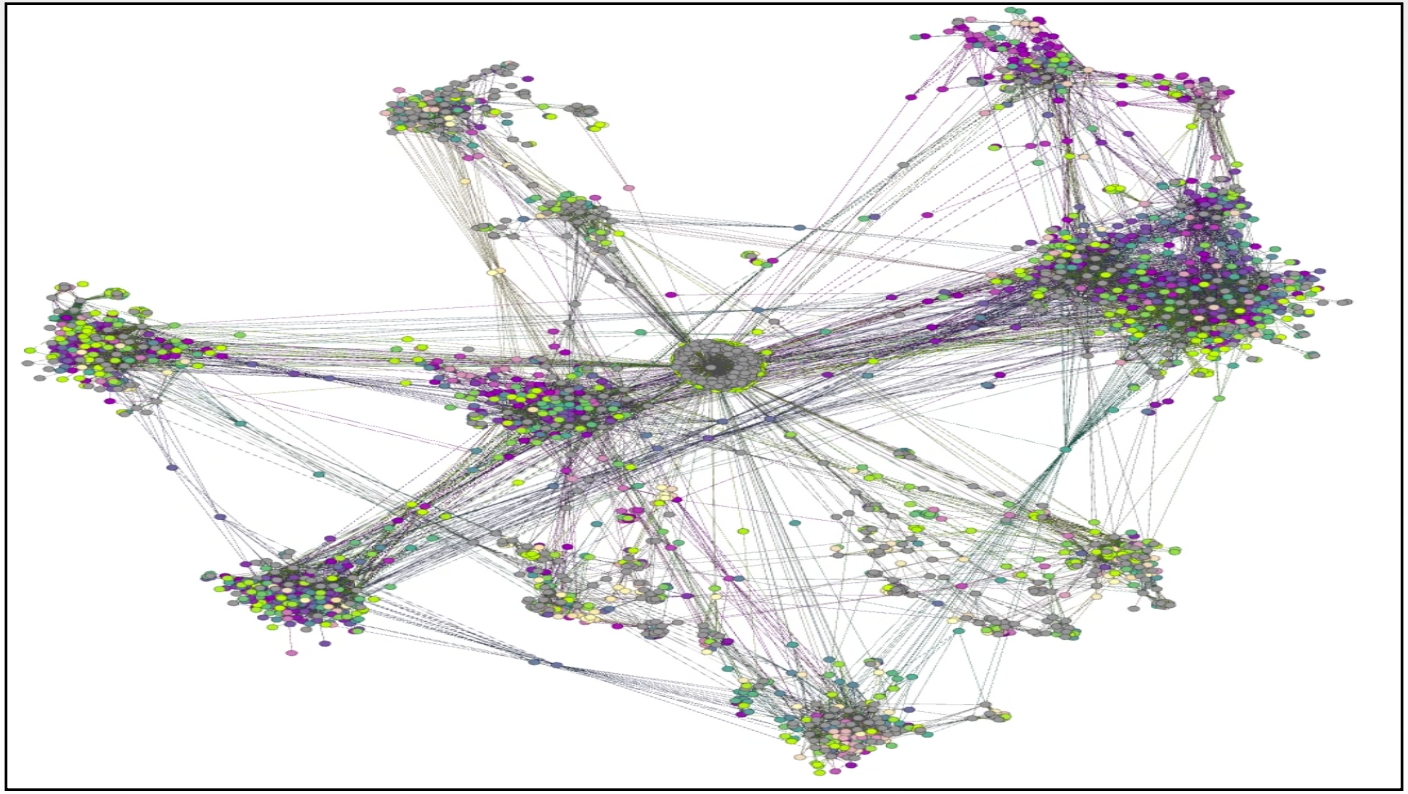


BFS/DFS + Other classic graph search algorithms are a great examples of algorithms useful in detecting this graph signature

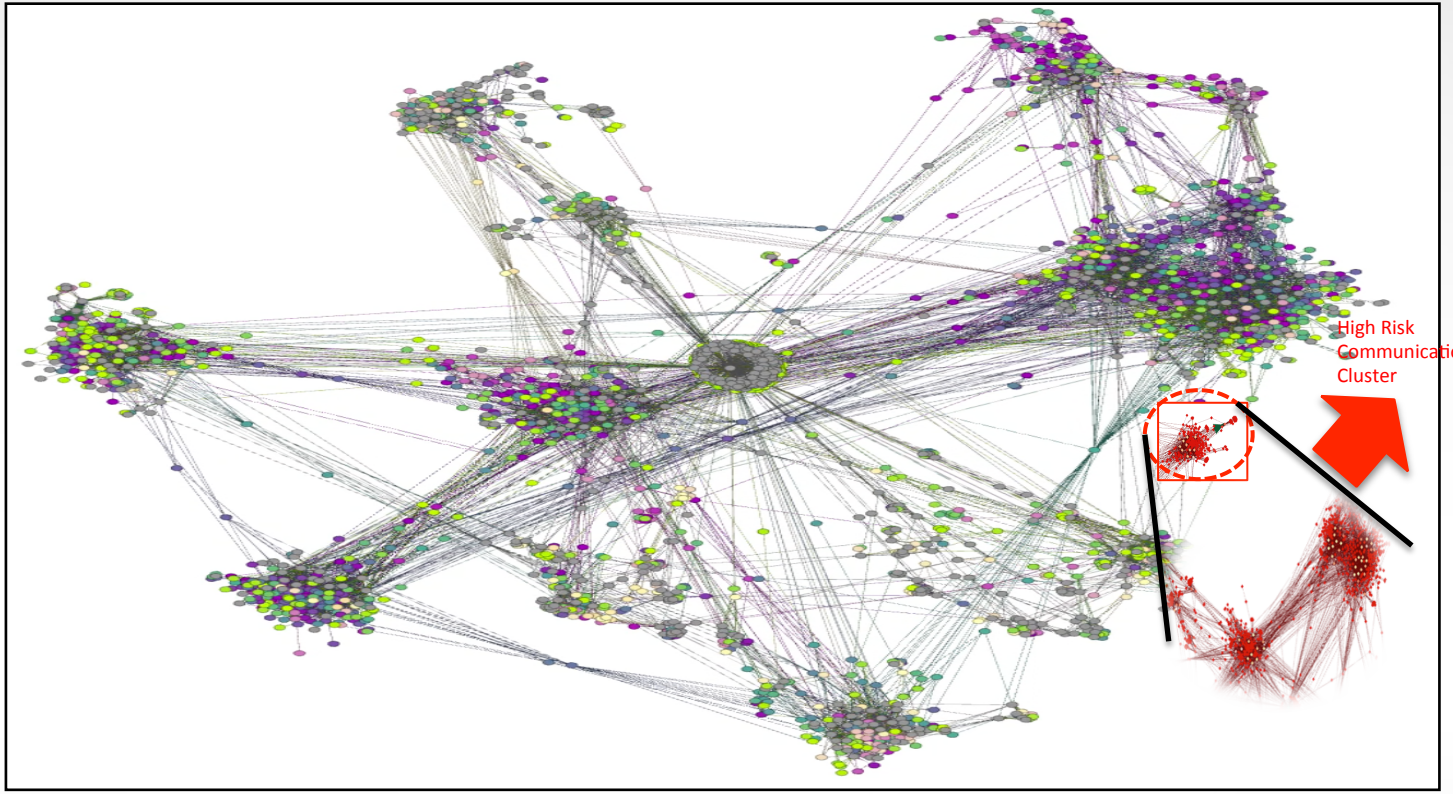
# Fractal Defense:  Batch + Real Time



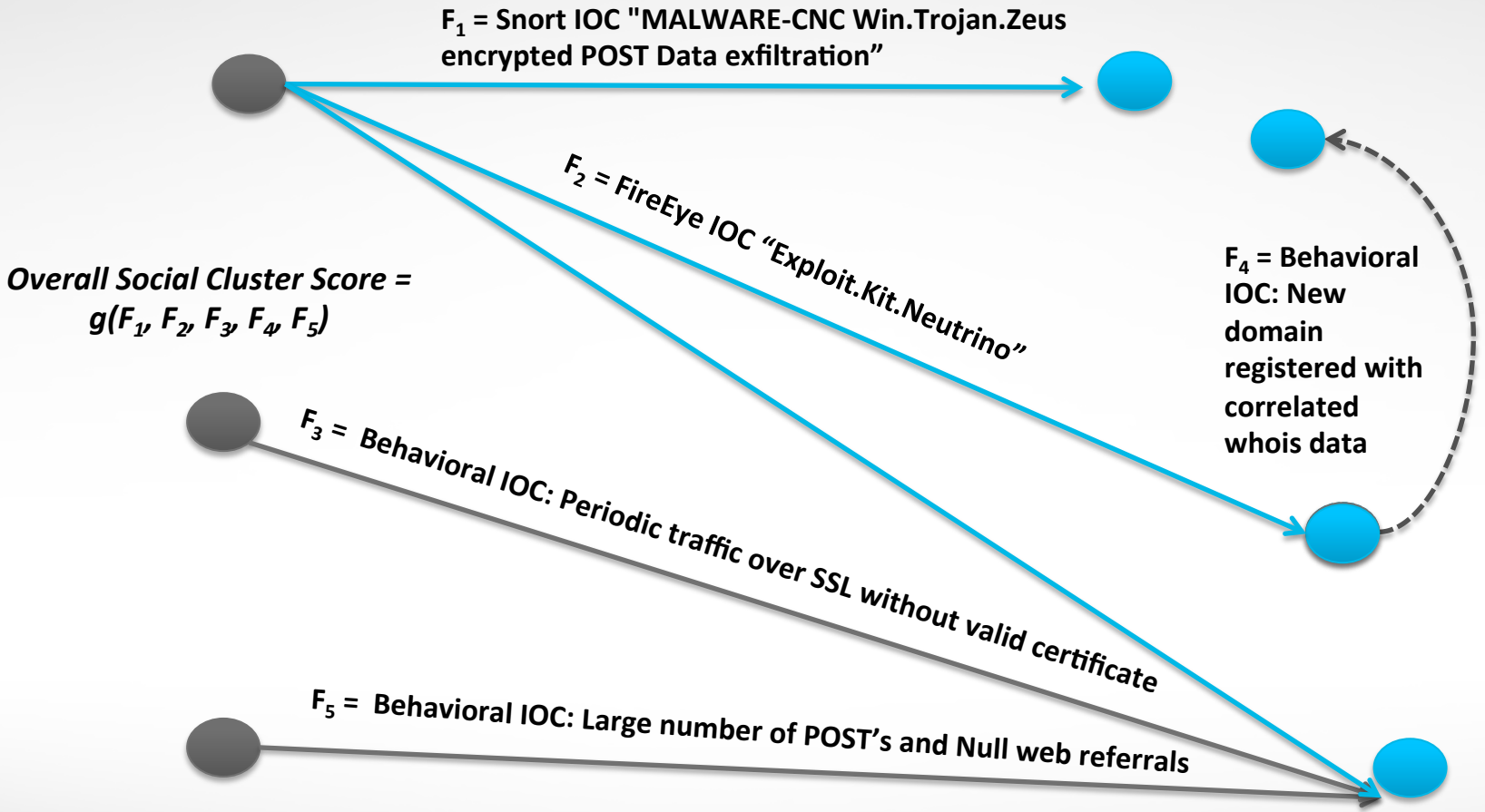
# Batch + Real Time: Identity Resolution

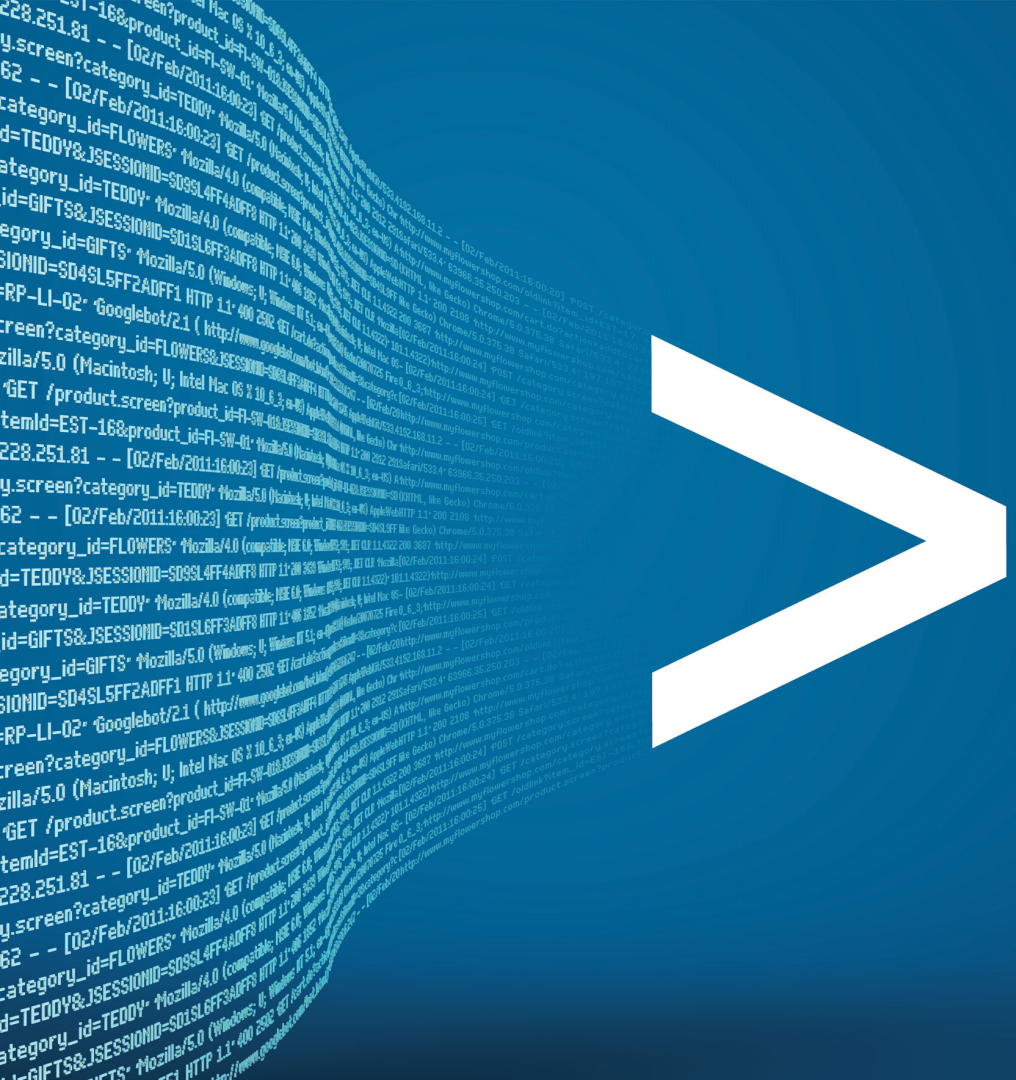


# Batch + Real Time: Updates



# Central Nervous System Approach





# Artificial Intelligence and Defense

splunk>

# Machine Learning and Cybersecurity

- Specific Challenges
  - No Ground Truth: Machine Learning works best in the case of large amount of labeled examples
  - Concept/Adversarial Drift: Labels change over time
- Lack of Labeled Data => “No Free Lunch”
  - Wolpert, D.H., Macready, W.G. (1997), "No Free Lunch Theorems for Optimization", *IEEE Transactions on Evolutionary Computation* **1**, 67.
  - Wolpert, David (1996), "The Lack of *A Priori* Distinctions between Learning Algorithms", *Neural Computation*, pp. 1341-1390.

> The so called No-Free-Lunch principle is a basic insight of machine learning. It may be viewed as stating that in the lack of prior knowledge (or inductive bias), any learning algorithm may fail on some learnable task.

# Limits of Automated Intrusion Detection

- Travis Goodspeed: “Packets in Packets”
  - Paper showing any communication medium we can embed a covert language to avoid eavesdropping in open channels
- Can we programmatically answer the questions: “Does a communication contain steganography?”
  - Equivalent to checking if a computer program will halt?
- Polymorphic Malware => NFL

# A.I. and Meta-Theorems

- Is intelligence achievable in software (Strong AI)?
  - Scott Aronson: Unlikely software/hardware combinations are competing against 3 Billion years of evolution
- Keep a catalogue of deep results and curiosities
  - Gödel's Incompleteness
  - Church-Turning
  - Blum's Speedup Theorem
  - No free Lunch
  - One Learning Algorithm Hypothesis
  - Grover's/Shor's Algorithms
- Track Cutting Edge ML
  - Paper: "Building high-level features using large scale unsupervised learning"
    - Andrew Ng and Jeff Dean et al. (2012) ICML

# Halting Problem

**The Halting Problem**

**Theorem 1** *If we define language*

$$HALT = \{\langle \alpha, x \rangle | M_\alpha \text{ stops on input } x\}$$

*then this language is not accepted by any Turing Machine.*

- The problem is to determine, given a program and an input to the program, whether the program will eventually halt when run with that input
- The halting problem is famous because it was one of the first problems proven algorithmically undecidable. This means there is no algorithm which can be applied to any arbitrary program and input to decide whether the program stops when run with that input.

# Theoretical Backbone

- Classical Computation
  - Logical Consistency of Computer Languages (Church-Turing)
  - Physical Realization of Turning Machine (Church turning + Von Neumann)
  - Floating point representation with controllable error propagation
- Weak/Strong AI
  - Halting problem and No Free Lunch theorems => building intelligent software is “hard”
  - Current machine learning methods are a type of weak AI
- Distributed Computation
  - Complexity classes P-Complete, NC
  - CAP Theorem
  - Actor Models
  - Batch + Real-Time := Lambda Architecture

# Data Science in Cybersecurity

- What is a behavior mathematically?
  - Fraud in Cybersecurity manifests itself in infinitely many possibilities
- Automated identification of fraud in IT is in some sense equivalent to trying solve the halting problem on a Turning Machine
  - Computationally it is impossible to “enumerate” all possible behaviors

# Blackhat Sound Bytes

- **Fractal Defense**: Reuse logic (and code) across different security use cases. Make behavior based IOC's map to adversary Tactics, Techniques and Procedures for better scalability.
- **Cybersecurity Analytics ROI**: Make requirements functional by setting realistic benchmarks based on your own data and metrics
- **Lambda Architecture**: a generic problem solving system built on immutability and hybrid batch/real-time workflows

# Sources


[1] Nathan Marz and James Warren. *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications, 2013.

[2] Nathan Marz. How to beat the CAP theorem. nathanmarz.com/blog/how-to-beat-the-cap-theorem.html, December 30 2011.

[3] Anton Chuvakin. 9 Reasons Why Building a Big Data Security Analytics Tool is Like Building a Flying Car. blogs.gartner.com/anton-chuvakin/2013/04/15/9-reasons-why-building-a-big-data-security-analytics-tool-is-like-building-a-flying-car/, April 15 2013.

[4] R. Sommer and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *Security and Privacy (SP), 2010 IEEE Symposium on*, pages 305–316, May 2010.

[5] Peter (Mudge) Zatko. Black Hat USA: How a Hacker Has Helped Influence the Government and Vice Versa. youtu.be/kZk9fsQisI8, 2011.

[6] Monzy Merza. Active response: Automated risk reduction or manual action? www.rsaconference.com/events/us15/agenda/sessions/2012/active-response-automated-risk-reduction-or-manual, 2015.

[7] David Bianco. The Pyramid of Pain. detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html, January 17, 2014.

[8] DARPA. Anomaly detection at multiple scales. opencatalog.darpa.mil/ADAMS.html, 2014.

[9] Nathan Marz. Cassandra NYC 2011: The storm and cassandra realtime computation stack. youtu.be/cF8a_FZwULI, December 30 2011.

[10] Nathan Marz. Apache storm. youtu.be/ucHjyb6jv08, 2013.

[11] Tom White. *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 1st edition, 2009.

[13] Nathan Marz. Apache storm. storm.apache.org, 2014.

[14] Marius Kloft and Pavel Laskov. Online anomaly detection under adversarial impact. In *I*, volume 9 of *JMLR Proceedings*, pages 405–412. JMLR.org, 2010.

[15] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *Trans. Evol. Comp*, 1(1):67–82, April 1997.

[16] David H. Wolpert. The lack of a priori distinctions between learning algorithms. *Neural Comput.*, 8(7):1341–1390, October 1996.

[17] Michael I Jordan. Ama: Michael i jordan. www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan, 2014.

[18] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.

[19] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. Technical Report Arxiv report 1206.5533, Université de Montréal, 2012.

[20] Eric J. Humphrey, Juan Pablo Bello, and Yann LeCun. Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In Fabien Gouyon, Perfecto Herrera, Luis Gustavo Martins, and Meinard Müller, editors, *ISMIR*, pages 403–408. FEUP Edições, 2012.

[21] Jack W. Stokes, John C. Platt, Joseph Kravis, and Michael Shilman. ALADIN: Active Learning of Anomalies to Detect Intrusions. Technical Report MSR-TR-2008-24, Microsoft, March 2008.

[22] G. Rush, D.R. Tauritz, and A.D. Kent. Dcafe: A distributed cyber security automation framework for experiments. In *Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International*, pages 134–139, July 2014.

[23] Jurg. Kohlas and Paul-Andre Monney. Theory of evidence: A survey of its mathematical foundations, applications and computational aspects. *Zeitschrift Operations Research*, 39(1):35–68, 1994.

[24] Audun Josang, Javier Diaz, and Maria Rifqi. Cumulative and averaging fusion of beliefs. *Information Fusion*, 11(2):192 – 200, 2010.

[25] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38(2):325–339, 04 1967.

[26] Romain Fontugne, Pierre Borgnat, Patrice Abry, and Kensuke Fukuda. MAWILab: Combining Diverse Anomaly Detectors for Automated Anomaly Labeling and Performance Benchmarking. In *ACM CoNEXT '10*, Philadelphia, PA, December 2010.

[27] Hoda Eldardiry et. al. Multi-source fusion for anomaly detection: using across-domain and across-time peer-group consistency checks. *JoWUA*, 5(2):39–58, 2014.

# Sources


[28] Pamela Bhattacharya, Marios Iliofotou, Iulian Neamtiu, and Michalis Faloutsos. Graph-based analysis and prediction for software evolution. In *International Conference on Software Engineering (ICSE)*, Zurich, Switzerland, June 2012.

[29] Yi Yang, Marios Iliofotou, Michalis Faloutsos, and Bin Wu. Analyzing interaction communication networks in enterprises and identifying hierarchies. In *IEEE International Workshop on Network Science (NSW)*, June 2011.

[30] Brian Gallagher, Marios Iliofotou, Tina Eliassi-Rad, and Michalis Faloutsos. Homophily in application layer and its usage in traffic classification. In *IEEE INFOCOM*, San Diego, CA, USA, March 2010.

[31] Hesham Mekky, Ruben Torres, Zhi-Li Zhang, Sabyasachi Saha, and Antonio Nucci. Detecting malicious HTTP redirections using trees of user browsing activity. In Giuseppe Bianchi, Yuguang M. Fang, and Xuemin S. Shen, editors, *INFOCOM 2014, 33rd IEEE International Conference on Computer Communications*, pages 1159–1167, Los Alamitos, CA, USA, April 2014. IEEE.

[32] Damien Faya Hamed Haddadibi, Steve Uhligc, Liam Kilmartind, Andrew W. Mooreb Jerome Kunegise, and Marios Iliofotouf. Discriminating graphs through spectral projections. *Computer Networks*, 55:3458–3468, October 2011.

[33] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322:549–560, 1905.

[34] Patrick Billingsley. *Probability and Measure*. Wiley-Interscience, April 1995.

[35] H. Hurst. Long term storage capacity of reservoirs. *Transaction of the American society of civil engineer*, 116:770–799, 1951.

[36] B. B. Mandelbrot and J. W. van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.

[37] David Nualart. *The Malliavin calculus and related topics*. Probability and its applications. Springer, Berlin, Heidelberg, New-York, 2006.

[38] Diethelm Wuertz et. al. farma: ARMA time series modeling. cran.r-project.org/web/packages/fArma/index.html, June 24 2013.

[39] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 2(1):1–15, February 1994.

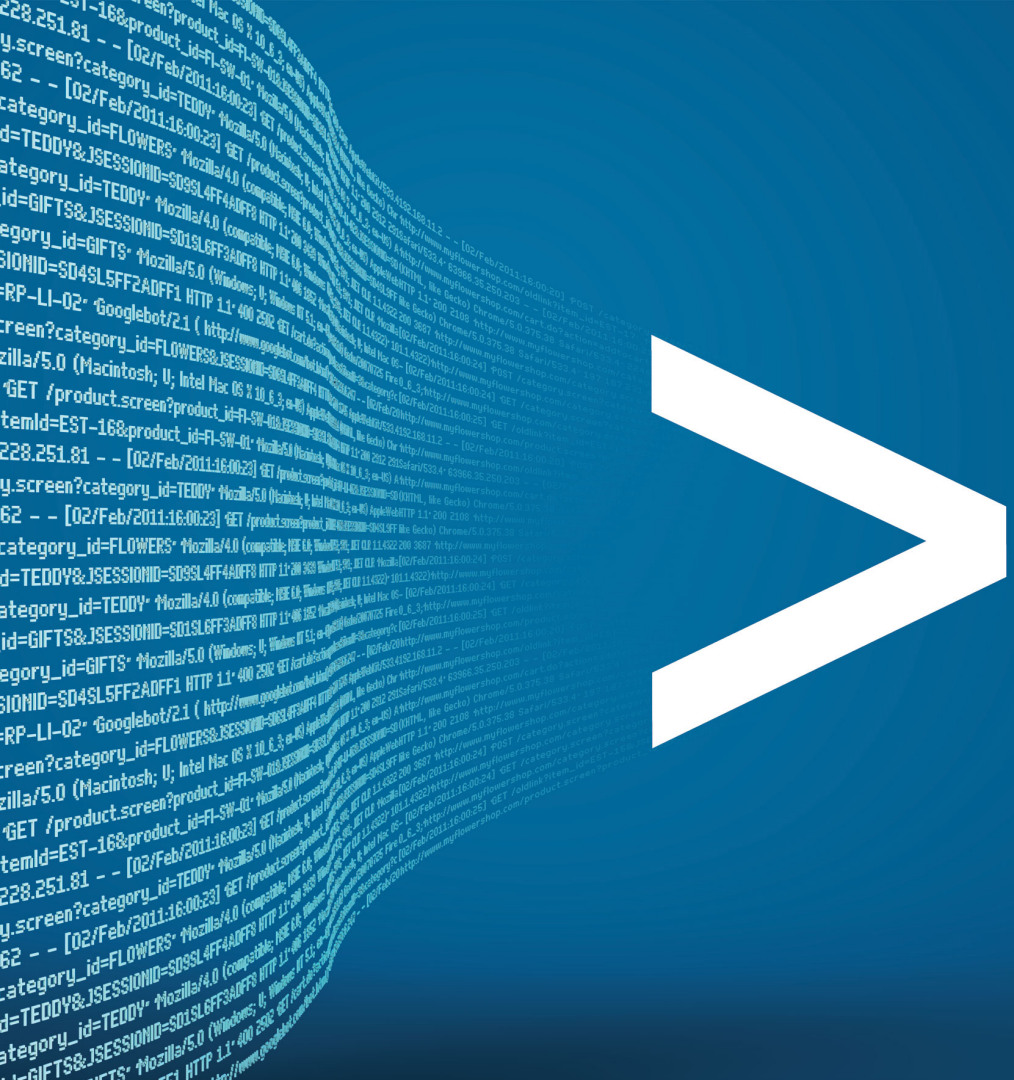
[40] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104, May 2000.

[41] Steffen Rendle. Factorization machines with libFM. *ACM Trans. Intell. Syst. Technol.*, 3(3):57:1–57:22, May 2012.

[42] Benoît Mandelbrot and J. R. Wallis. Robustness of the rescaled range $R/S$ in the measurement of noncyclic long run statisticaldependence. 5:967–988, 1969.

[43] Alexandra Chronopoulou and Frederi G. Viens. Hurst index estimation for self-similar processes with long-memory. In *Recent development in stochastic dynamics and stochastic analysis. Dedicated to Zhi-Yuan Zhang on the occasion of his 75th birthday.*, pages 91–117. Hackensack, NJ: World Scientific, 2010.

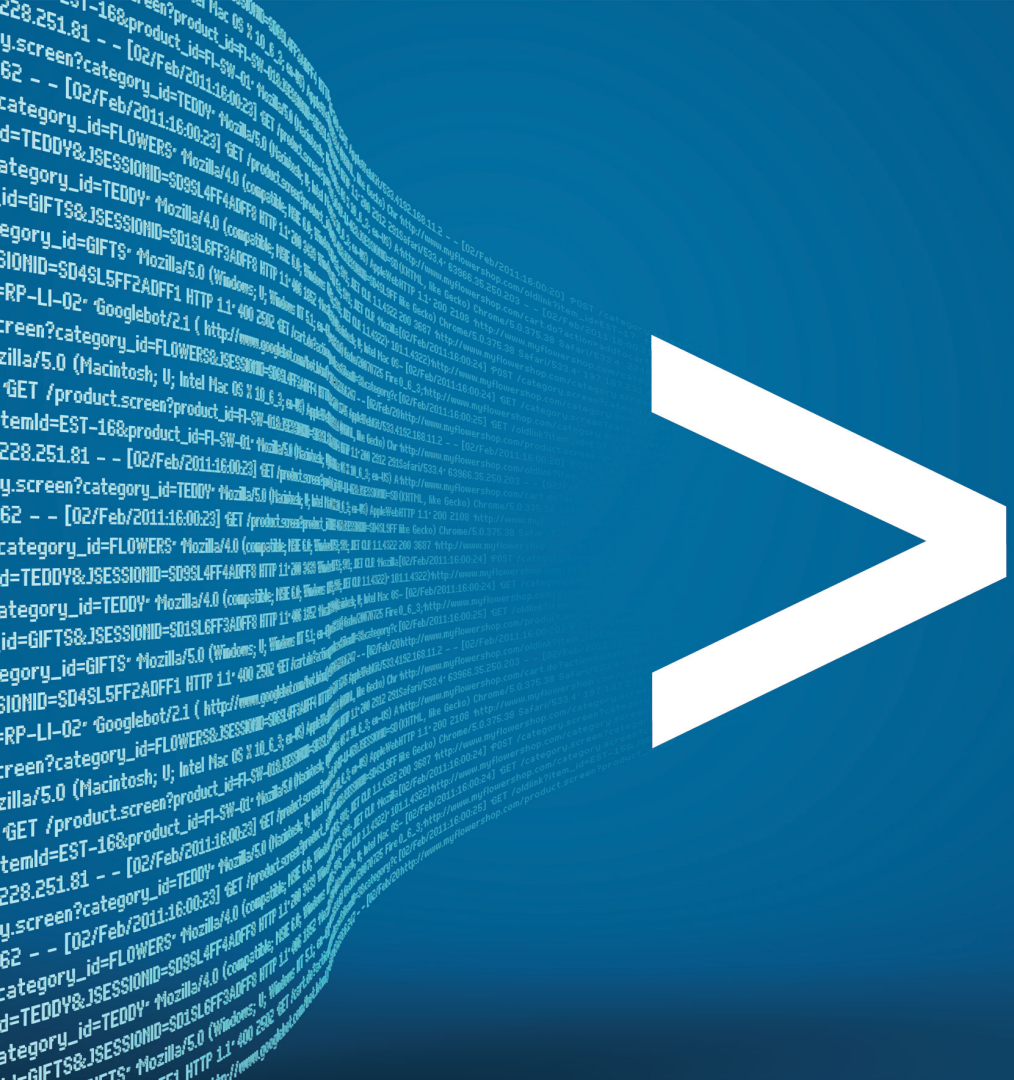
[44] Mila Parkour. Contagio malware dump. contagiodump.blogspot.com/2013/04/collection-of-pcap-files-from-malware.html, 2015.



# Q&A

splunk>


Thank You
splunk>



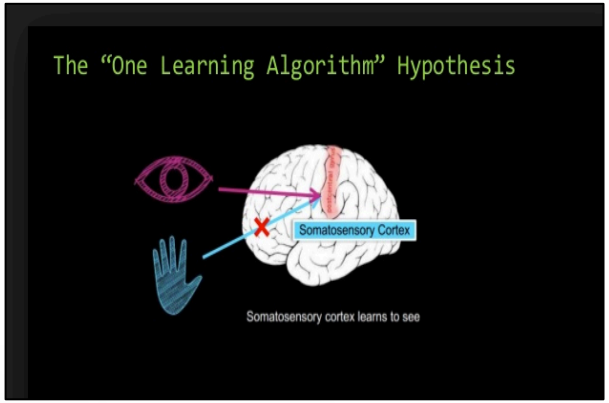
# Appendix

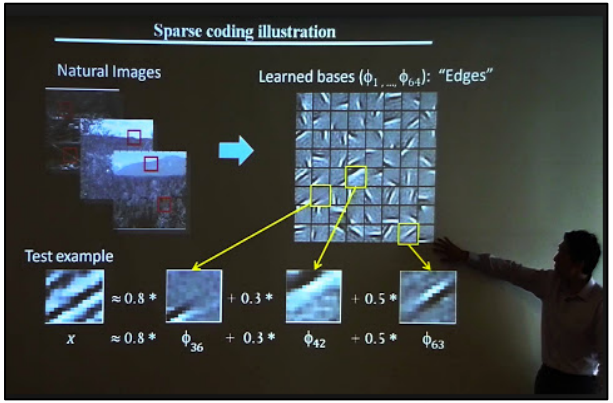
splunk>

# One Learning Algorithm Hypothesis

- *Modular Minds Hypothesis* — The mind is primarily composed of stable cortical circuits which encapsulate specific cognitive competences and exhibit a high degree of structural and informational modularity.

- *Single Algorithm Hypothesis* — There is one fundamental algorithm that underlies all or most cortical computations; it is implemented on a computationally homogeneous cortical substrate and runs simultaneously in multiple instances on different inputs.



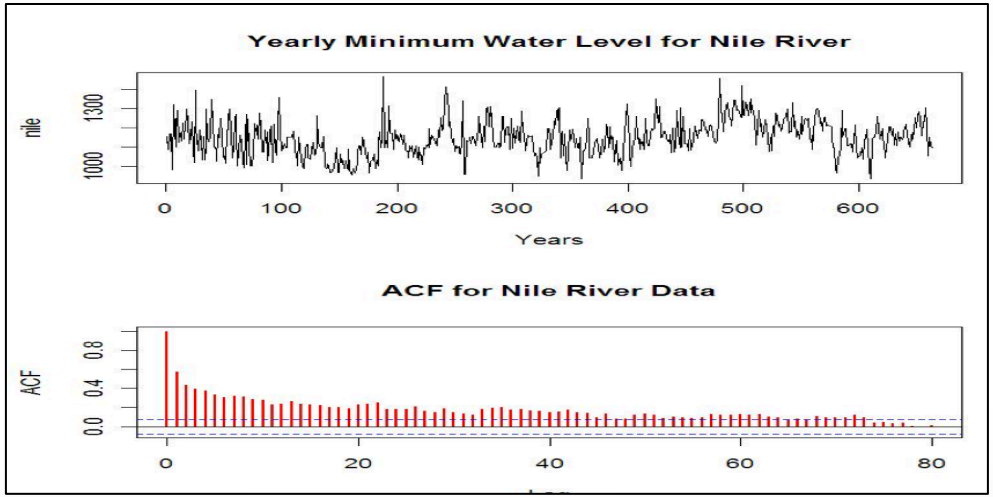



# Theoretical Background

- Doctoral Research
  - Iterated Processes: What happens when we replace time with a random process?
    - Set t "=" B(t) where B is a Brownian motion
  - Can Feynman Path Integral be defined Mathematically?
    - Feynman-Kac's Formula: Duality between PDE's and SDE's
    - Measures on the space of continuous functions
  - Fractional Brownian motion and processes with long memory
    - Random walks that are not Markov
    - Malliavin Calculus: (Used to prove Hörmander's Theorem)
      - Malliavin built a calculus out of Random processes replacing time by h in a Hilbert space

# Stochastic Processes with Long Memory

- Since ancient times the Nile River has been known for its long periods of dryness followed by long periods of floods
- The hydrologist Hurst was the first one to describe these characteristics when he was trying to solve the problem of flow regularization of the Nile River.



# Fractional Brownian Motion

**Definition 2.2.1** *Fractional Brownian motion*

*A Fractional Brownian motion with Hurst parameter* $H \in (0,1)$ *is a centered Gaussian process* $B = \{B(t) : t \geq 0\}$ *with covariance function given by* $\mathbf{E}[B(t)B(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$.

The first mathematical definition of Brownian motion was given by Bachelier in his 1900 Ph.D. thesis entitled "Theorie de la Speculation." Albert Einstein also worked on Brownian motion and its relation to the heat equation in the 1906 paper titled "On the theory of the Brownian movement" [14]. In this seminal paper Einstein derived the heat equation involving a so called Diffusion constant $\kappa$.

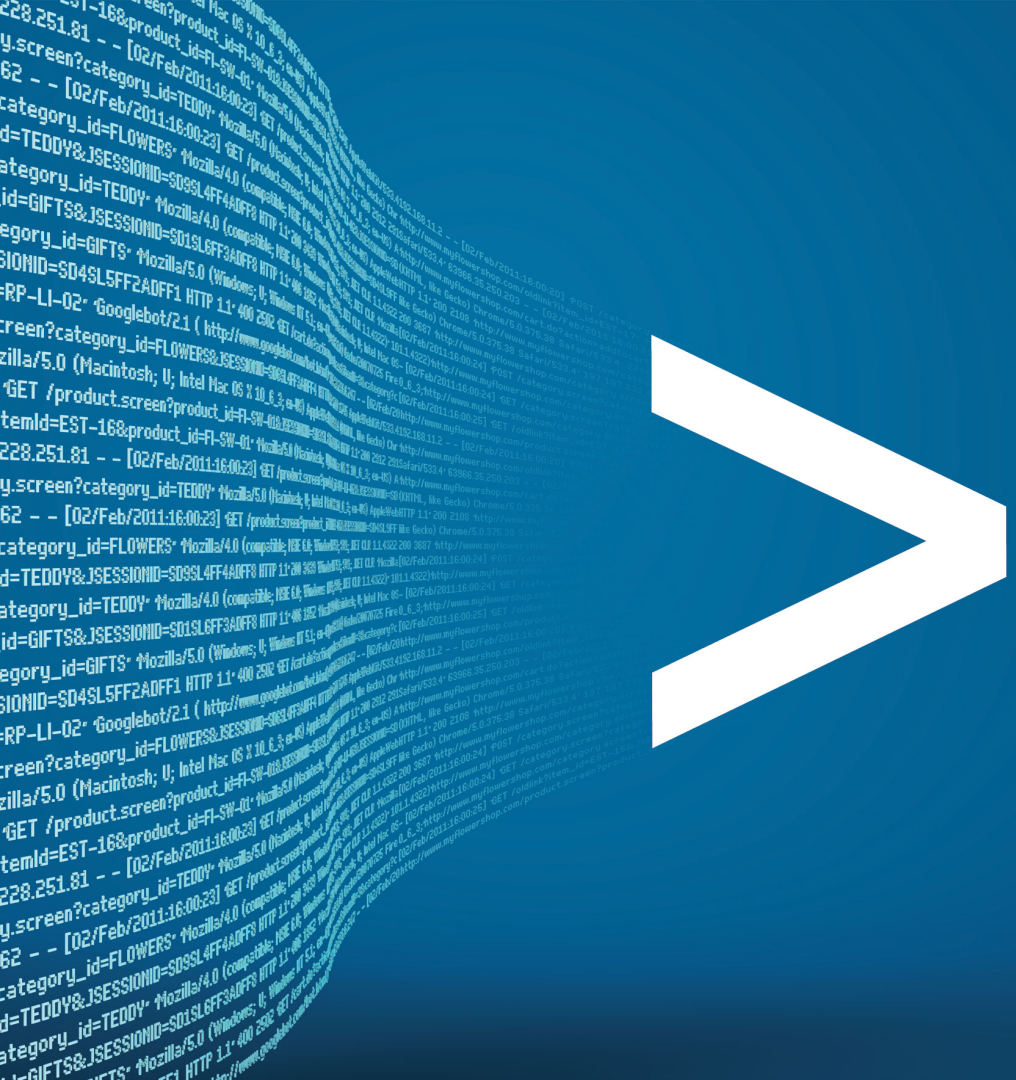
$$\frac{\partial \rho}{\partial t} = \kappa \Delta \rho \tag{2.1}$$

The solution to this equation is

$$\rho(x,t) = \frac{1}{(4\pi\kappa t)^{\frac{3}{2}}} \exp\left(\frac{-|x|^2}{4\kappa t}\right)$$

From this formula Einstein expressed the position of a Brownian particle $B(t)$ at time $t$ using a probabilistic notation or more specifically

$$\mathbf{P}\left(B(t) \in [a,b]\right) = \int_a^b \rho(x,t)\mathrm{d}u \tag{2.2}$$



# Splunk for Security

splunk>

# Thousands of Customers & Analyst Validation

ExxonMobil
Raytheon
Bank of America
cisco
First Data
JOHNS HOPKINS MEDICINE
UnionBank
sears
MOTOROLA
Interac
China unicom中国联通
BEA東亞銀行
HUGHES
BARCLAYS
verizon
SAMSUNG
SAFEWAY
sky
intuit
LinkedIn
MLB.com
PSA PEUGEOT CITROËN
The University of Akron
Deloitte.
THE UNIVERSITY OF TEXAS AT AUSTIN
CEDARCRESTONE
comcast
BECHTEL
GENERAL DYNAMICS
ABN·AMRO Private Banking

**Gartner MQ for SIEM 2014**

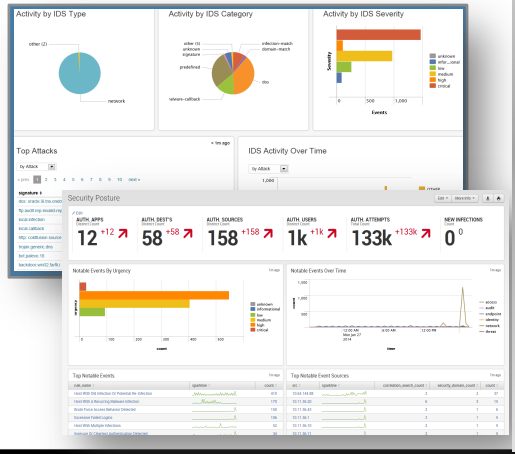
# Analytics-Driven Security Use Cases


splunk>

Splunk software complements, replaces and goes beyond traditional SIEMs.

# Splunk Security Intelligence Platform

## Splunk App for Enterprise Security




## Caspida

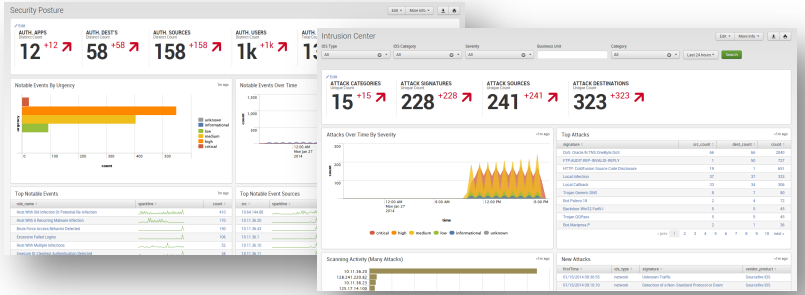

## 240+ security apps

- Palo Alto Networks
- Blue Coat Proxy SG
- Cisco Security Suite
- OSSEC
- F5 Security
- FireEye
- NetFlow Logic
- Active Directory
- Juniper
- Sourcefire

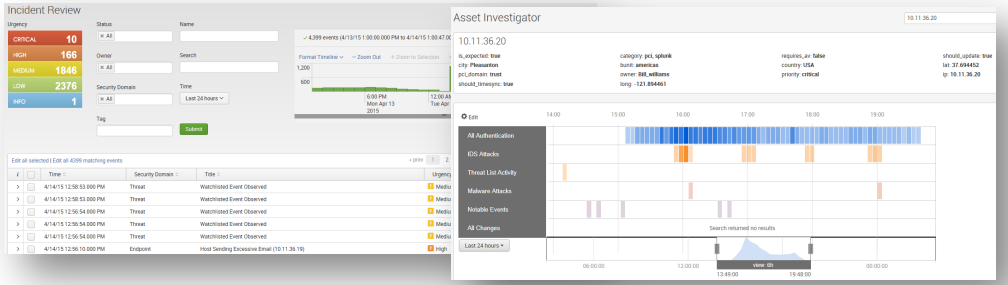
splunk>

# Splunk App for Enterprise Security

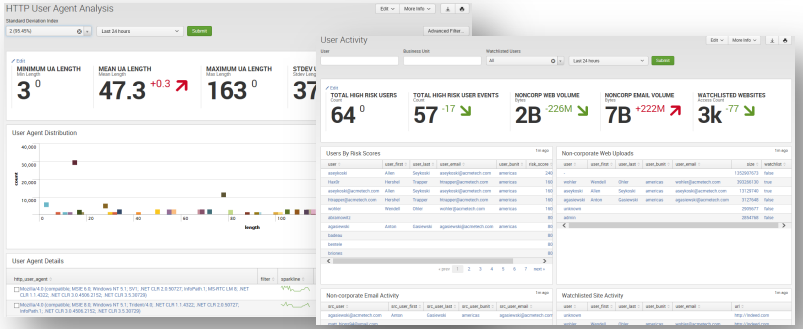
## Pre-built searches, alerts, reports, dashboards, incident workflow, and threat intelligence feeds



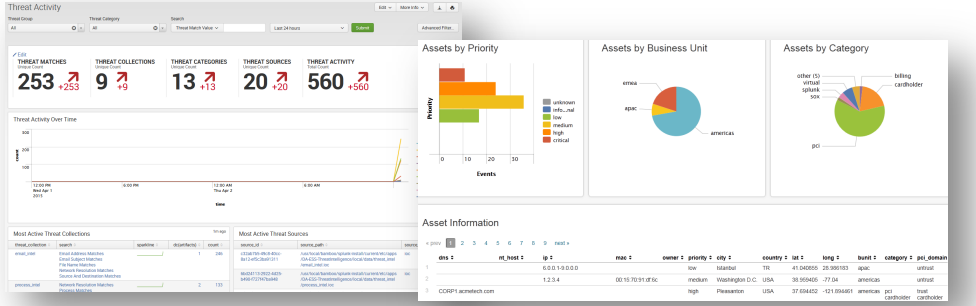

### Alerts & Dashboards & Reports




### Incident Investigations & Management




### Statistical Outliers & Risk Scoring & User Activity




### Threat Intel & Asset & Identity Integration

# Splunk Is Used Across IT and the Business

***Strong ROI & facilitates cross-department collaboration***