AI & ML in Cyber Security

Why Algorithms Are Dangerous

Kaspersky®
SECURITY
ANALYST
SUMMIT

Cancun, Mexico
March, 2018

Raffael Marty

A Brief Summary

- We don’t have artificial intelligence (yet)
- Algorithms are getting ‘smarter’, but experts are more important
- Stop throwing algorithms on the wall - they are not spaghetti
- Understand your data and your algorithms
- Invest in people who know security (and have experience)
- Build “export knowledge” absorbing systems
- Focus on advancing insights

The master of Kennin temple was Mokurai. He had a little protégé named Toyo who was only twelve years old. Toyo saw how students entered the masters room each day and received instructions and guidance in Zen. The young boy wished to do zazen (meditation) as well. Upon convincing Mokuri, he went in front of the master who gave him the following koan to ponder:

"You can hear the sound of two hands when they clap together," said Mokurai. "Now show me the sound of one hand."

Outline

1 ***Statistics, Machine Learning & AI***
Defining the Concepts

2 ***The Algorithmic Problem***
Understanding the Data and the Algorithms

3 ***An Example***
Let’s Get Practical

http://theconversation.com/your-questions-answered-on-artificial-intelligence-49645

“Everyone calls their stuff ‘machine learning’ or even better ‘artificial intelligence’ - It’s not cool to use statistics!”

“Companies are throwing algorithms on the wall to see what sticks - see security analytics market”

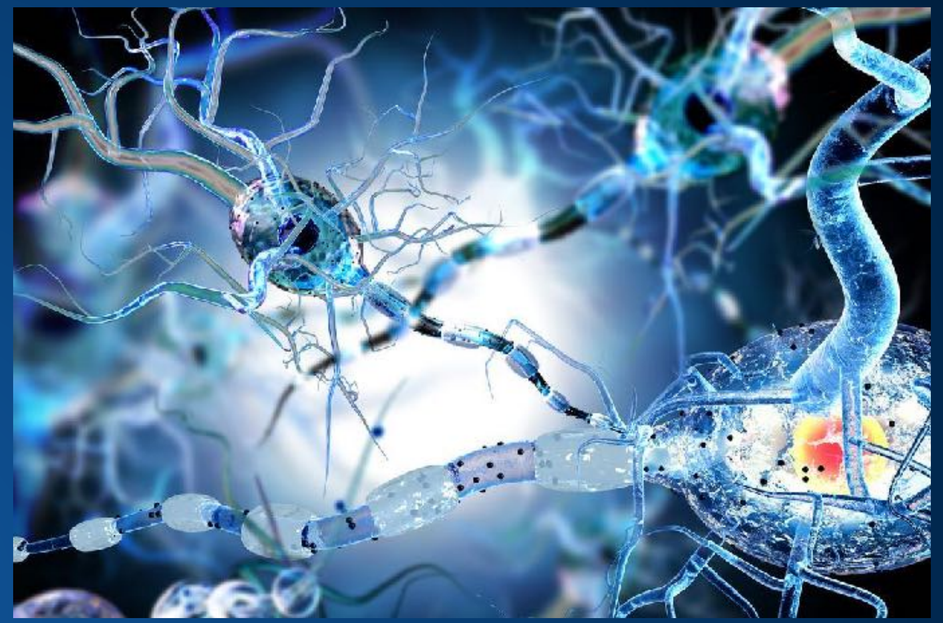
ML and AI – What Is It?

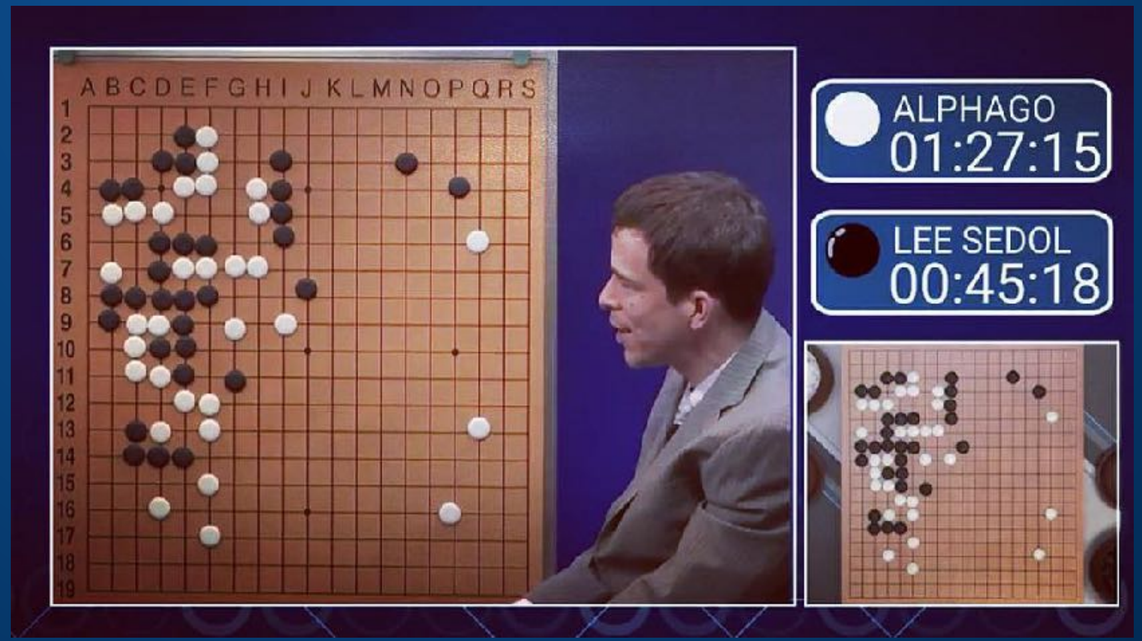
- **Machine Learning** – Algorithmic ways to “describe” data
  - Supervised - learning from training data
  - Unsupervised - optimization problem to solve (clustering, dim reduction)
- **Deep Learning** – a ‘newer’ machine learning algorithm
  - Eliminates the feature engineering step
  - Verifiability / explainability issues
- **Data Mining** – Methods to explore data – automatically and interactively
- **Artificial Intelligence** – “Just calling something AI doesn’t make it AI.”

*”A program that doesn't simply classify or compute model parameters, but comes up with **novel knowledge** that a security analyst finds insightful.”*

What “AI” Does Today

- Kick a human's ass at Go
- Design more effective drugs
- Make Siri smarter



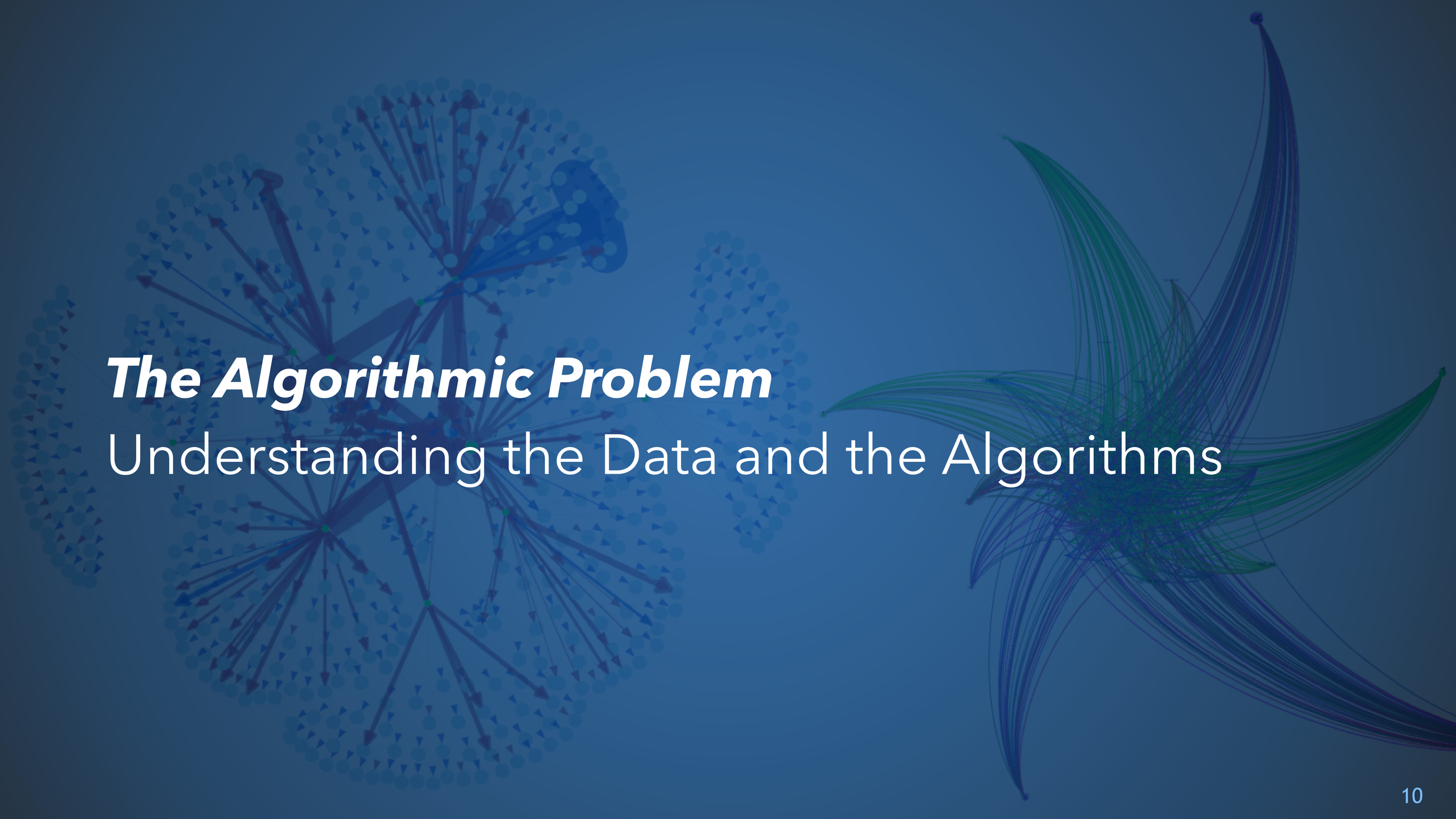


Machine Learning Uses in Security

- Supervised
  - **Malware classification** **(deep learning poster child)**
  - **Spam identification**
  - **MLSec project on firewall data**

- Unsupervised
  - **DNS analytics** (domain name classification, lookup frequencies, etc.)
  - **Threat Intelligence** feed curation (IOC prioritization, deduplication, …)
  - **Tier 1 analyst automation** (reducing 600M events to 100 incidents)*
  - User and Entity Behavior Analytics (UEBA)

* See Respond Software Inc.



The Algorithmic Problem

Understanding the Data and the Algorithms


WARNING
Algorithms Are Dangerous

Famous AI (Algorithm) Failures

December 2009 | Hewlett-Packard investigates instances of so-called "racist camera software" which had trouble recognizing dark-skinned people

March 2015 | A Carnegie Mellon University study determines that some personalized ads from sites such as Google and Facebook are gender-biased

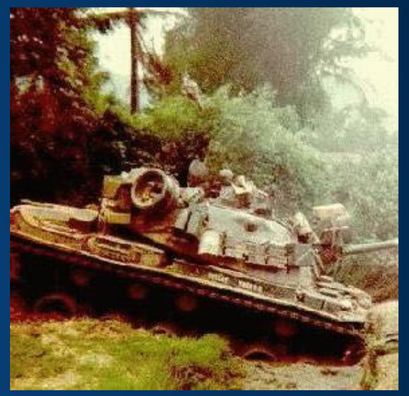
July 2015 | A Google algorithm mistakenly captions photos of black people as "Gorillas"

March 2016 | Microsoft shuts down AI chatbot Tay after it starts using racist language

May 2016 | ProPublica investigation finds that a computer program used to track future criminals in the US is racially biased

September 2016 | Machine-learning algorithms used to judge an international beauty contest displays bias against dark-skinned contestants

neil.fraser.name/writing/tank/



US government in October 2016 published a comprehensive report titled "Preparing for the future of artificial intelligence"

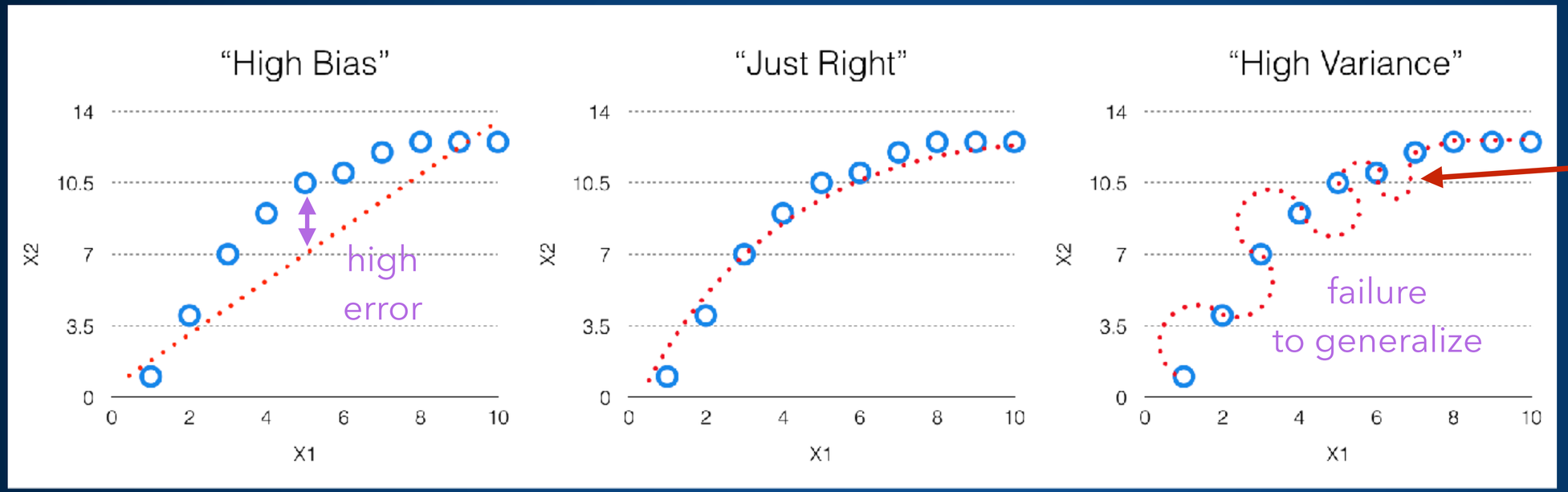
What Makes Algorithms Dangerous?

- Algorithms make **assumptions** about the **data**
  - Assume ‘clean’ data (src/dst confusion, user feedback, etc.)
  - Often assume a certain type of data and their distribution
  - Don’t deal with outliers
  - Machine learning assumes enough, representative data
  - Needs contextual features (e.g., not just IP addresses)
  - Assume all input features are ‘normalized’ the same way

- Algorithms are **too easy** to use these days (tensorflow, torch, ML on AWS, etc.)
  - The process is more important than the algorithm (e.g., feature engineering, supervision, drop outs, parameter choices, etc.)

- Algorithms do not take **domain knowledge** into account
  - Defining meaningful and representative distance functions, for example
  - e.g., each L4 protocol exhibits different behavior. Train it separately.
  - e.g., interpretation is often unvalidated - beware of overfitting and biased models.
  - Ports look like numerical features, they are not, same is true for IPs, processIDs, HTTP return codes, etc.

Models Are Just Approximations



How do you know in what case you operate?

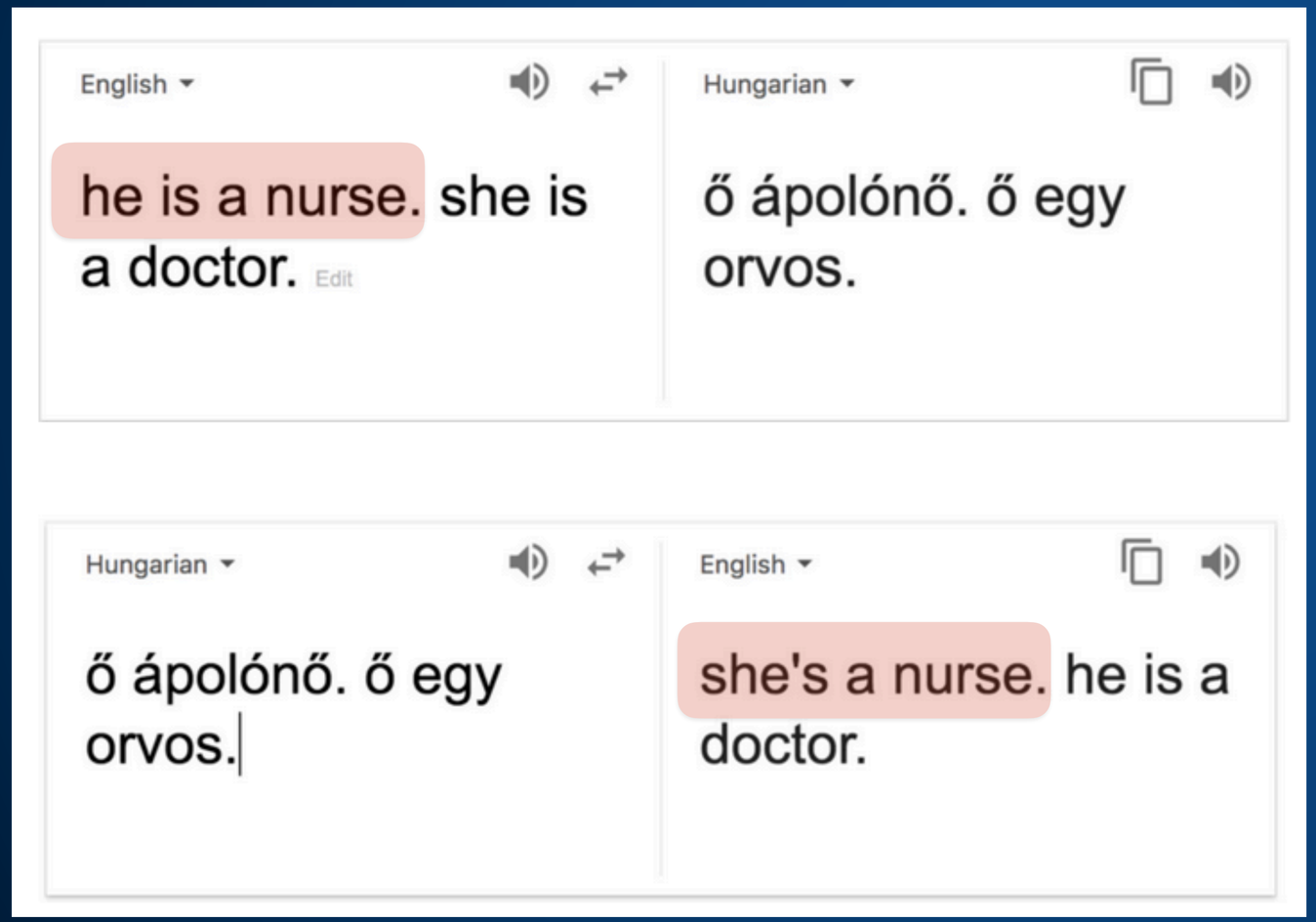
- ML explainability problem
- Compute error margins

High Bias - increasing the number of input features.

High Variance reduce the number of input features, increasing the number of training examples

Cognitive Biases

- How biased is your data set? How do you know?



- Only a single customer's data
- Learning from an 'infected' data set
- Collection errors
- Missing data (e.g., due to misconfiguration)
- What's the context the data operates in?
  - FTP although generally considered old and insecure, isn't always problematic
- Don't trust your IDS (e.g. "UDP bomb")

Don’t Use Machine Learning If …

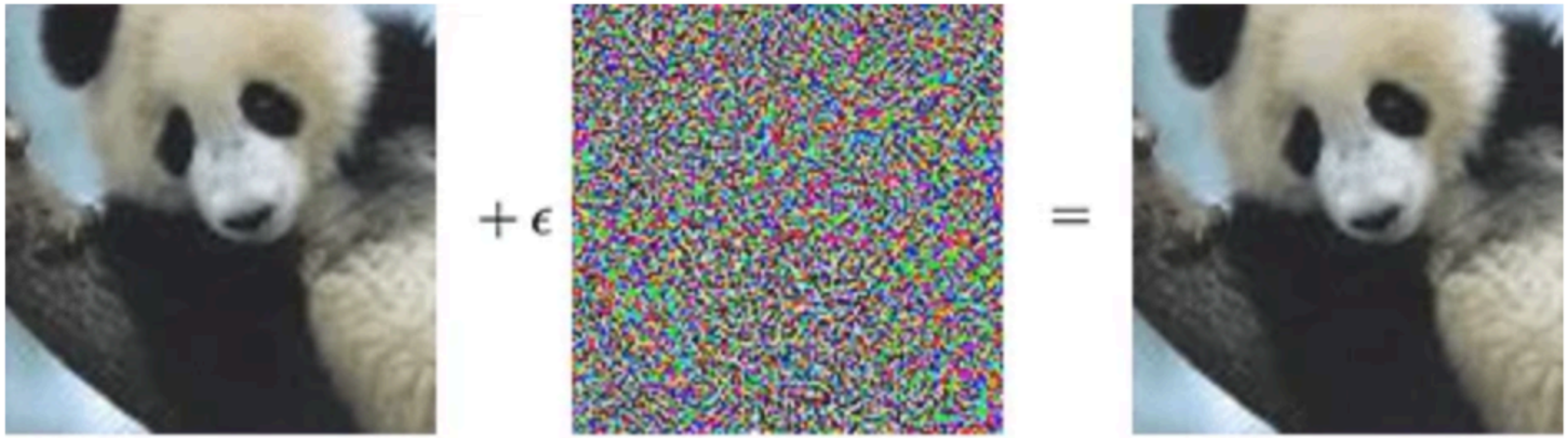
- Not enough or no quality **labeled data**
  - Don’t use for network traffic analysis - you don’t have labeled data - really, you don’t!
- No well trained **domain experts** and **data scientists** to oversee the implementation
  - Not enough domain expertise to engineer good **features**
- Need to understand what ML actually learned (**explainability**)

Also remember

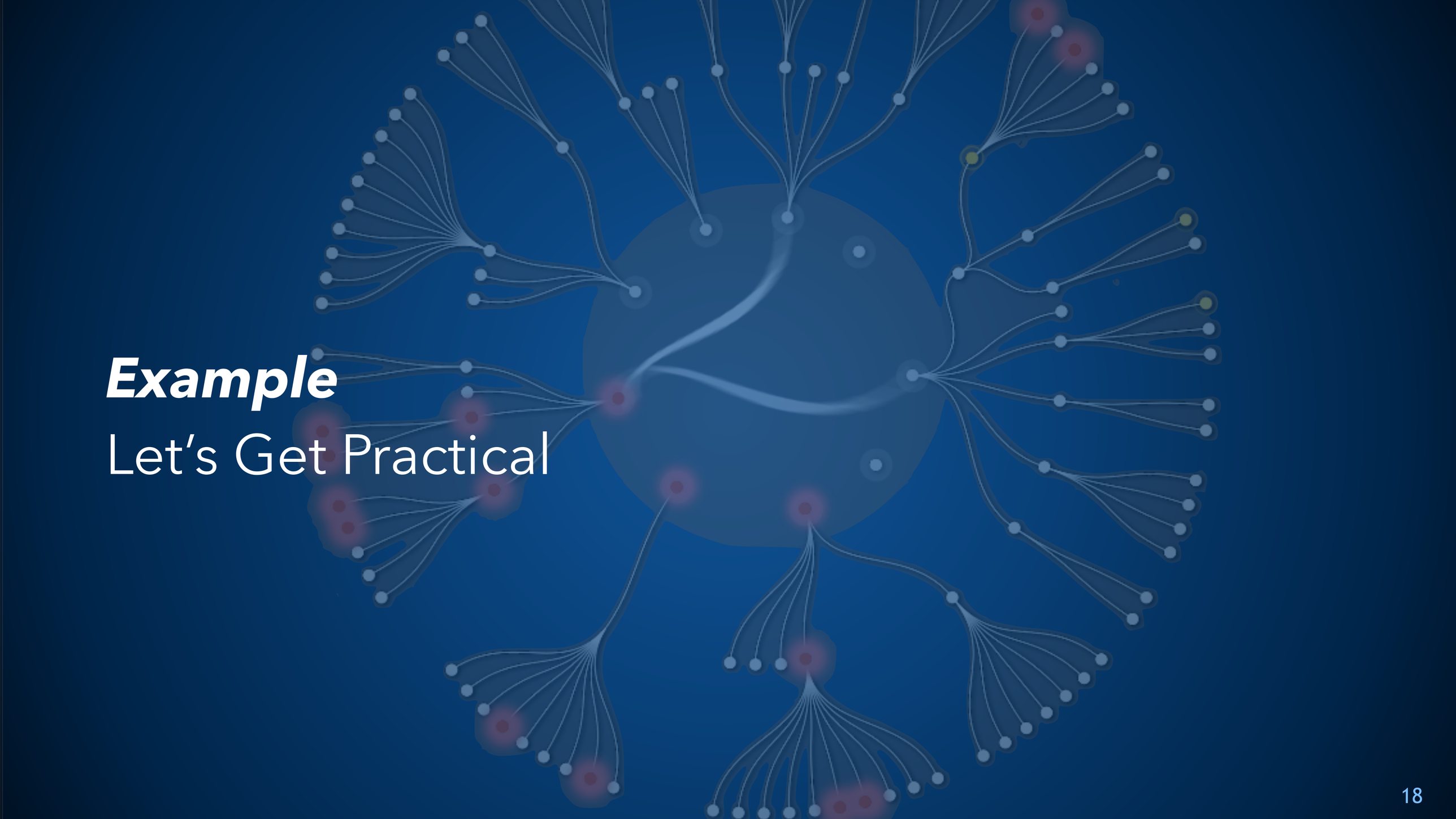
- Data cleanliness issues (timestamps, normalization across fields, etc.)
- Operational challenges (scalability and adaptability) of implementing machine learning models in practice

Adversarial Machine Learning

- An example of an attack on deep learning



Above: Image Credit: Ian Goodfellow

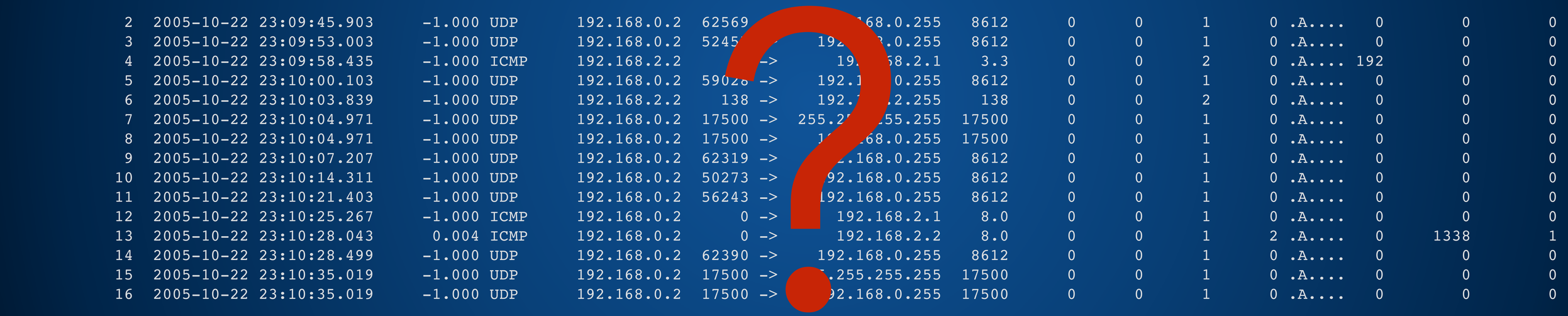


Example

Let’s Get Practical

Network Traffic - Finding Anomalies / Attacks

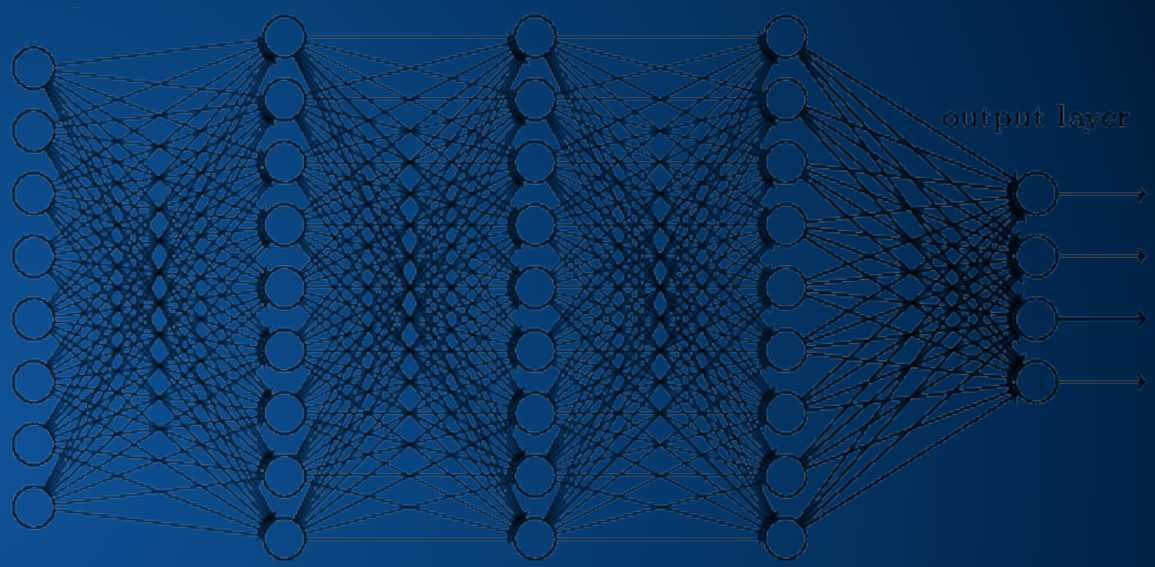
- Given: Netflow
- Task: Find anomalies / attacks



2	2005-10-22 23:09:45.903	-1.000	UDP	192.168.0.2	62569	->	168.0.255	8612	0	0	1	0	.A....	0	0	0
3	2005-10-22 23:09:53.003	-1.000	UDP	192.168.0.2	5245	->	192 .0.255	8612	0	0	1	0	.A....	0	0	0
4	2005-10-22 23:09:58.435	-1.000	ICMP	192.168.2.2		->	19 68.2.1	3.3	0	0	2	0	.A....	192	0	0
5	2005-10-22 23:10:00.103	-1.000	UDP	192.168.0.2	59028	->	192.1 0.255	8612	0	0	1	0	.A....	0	0	0
6	2005-10-22 23:10:03.839	-1.000	UDP	192.168.2.2	138	->	192. 2.255	138	0	0	2	0	.A....	0	0	0
7	2005-10-22 23:10:04.971	-1.000	UDP	192.168.0.2	17500	->	255.25 55.255	17500	0	0	1	0	.A....	0	0	0
8	2005-10-22 23:10:04.971	-1.000	UDP	192.168.0.2	17500	->	1 68.0.255	17500	0	0	1	0	.A....	0	0	0
9	2005-10-22 23:10:07.207	-1.000	UDP	192.168.0.2	62319	->	2.168.0.255	8612	0	0	1	0	.A....	0	0	0
10	2005-10-22 23:10:14.311	-1.000	UDP	192.168.0.2	50273	->	92.168.0.255	8612	0	0	1	0	.A....	0	0	0
11	2005-10-22 23:10:21.403	-1.000	UDP	192.168.0.2	56243	->	192.168.0.255	8612	0	0	1	0	.A....	0	0	0
12	2005-10-22 23:10:25.267	-1.000	ICMP	192.168.0.2	0	->	192.168.2.1	8.0	0	0	1	0	.A....	0	0	0
13	2005-10-22 23:10:28.043	0.004	ICMP	192.168.0.2	0	->	192.168.2.2	8.0	0	0	1	2	.A....	0	1338	1
14	2005-10-22 23:10:28.499	-1.000	UDP	192.168.0.2	62390	->	192.168.0.255	8612	0	0	1	0	.A....	0	0	0
15	2005-10-22 23:10:35.019	-1.000	UDP	192.168.0.2	17500	->	.255.255.255	17500	0	0	1	0	.A....	0	0	0
16	2005-10-22 23:10:35.019	-1.000	UDP	192.168.0.2	17500	->	92.168.0.255	17500	0	0	1	0	.A....	0	0	0

Network Traffic - Deep Learning

- Solution: Deep Learning
  - No feature engineering - really?
  - Lots of data available
  - What are the labels?

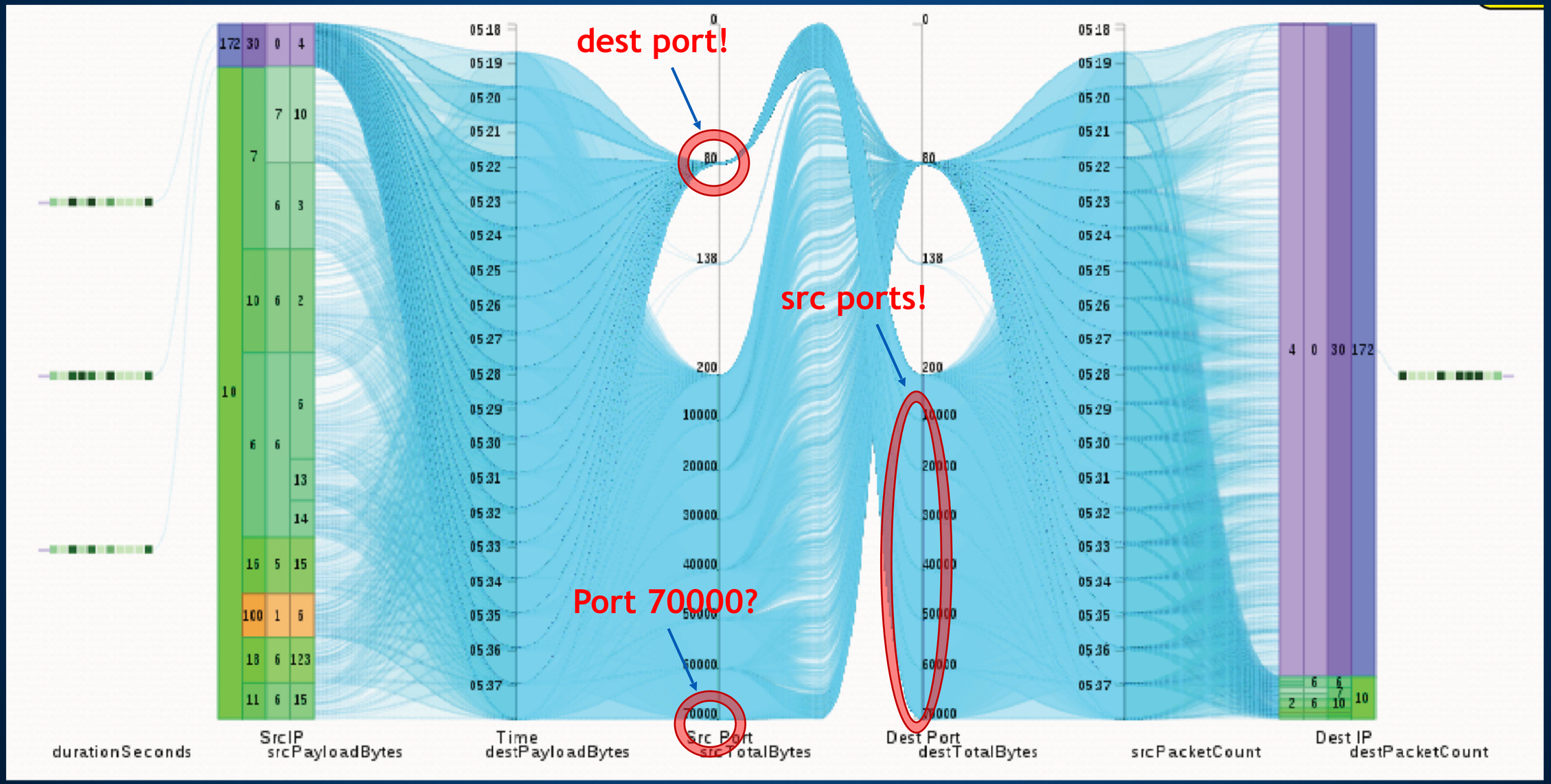


Most security problems can't be solved with Deep Learning

Analytics Challenges

- Data cleansing
- Wrong features -> wrong results
- Distance functions (for unsupervised approaches)
- Selecting the right algorithm (there is more than k-means!)
  - Algorithmic stability across iterations (iterative)?
- Parameter choices for algorithm

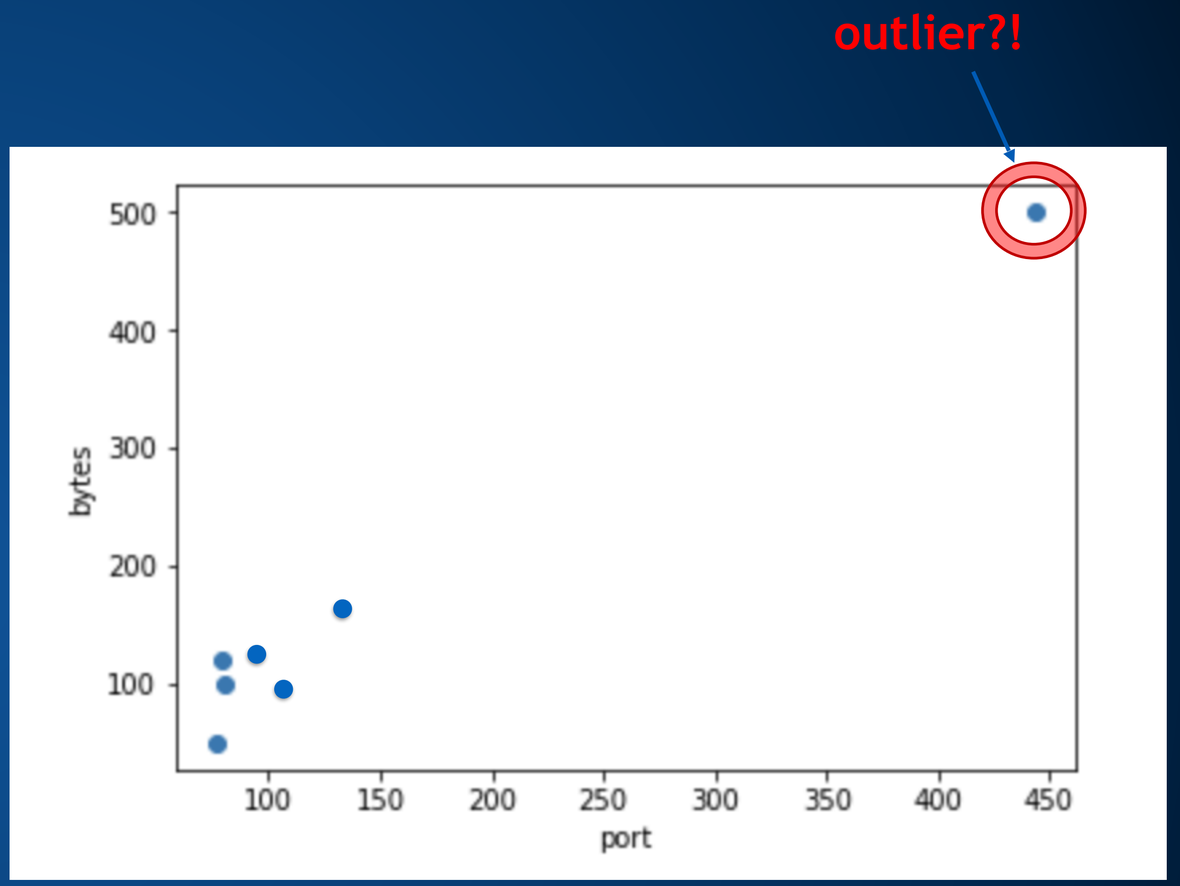
VAST Challenge 2013 Submission - Spot the Problems?



http://vis.pku.edu.cn/people/simingchen/docs/vastchallenge13-mc3.pdf

Distance Functions

- Need a domain-centric similarity function
  - URLs (simple levenshtein distance versus domain based?)
  - **Ports** (and IPs, ASNs) **are NOT numerical** features
  - Treat user names as categories, not strings



Distance Functions

2.3 Metric

One of the main assumptions made was that data instances having the same label will tend to be closer together than instances with different labels under some metric. Therefore, finding or constructing an appropriate metric is critical to the performance of the method.

The particular choice of metric is likely to be dictated by the domain. In detecting network intrusions, it seemed at first that some features of the data instances would be important (have greater weight) than others, and thus differences in the values of those features should have a greater contribution to the overall distance. Therefore, we experimented with several weighted metrics, with higher weights assigned to different subsets of features.

However, in the end we used a standard Euclidean metric, with equally weighted features. One reason for this was that while the weighted metric did show some increase in performance, it was not a significant amount. But more importantly, tuning the metric's parameters to achieve maximum performance for a particular domain, data distribution, and feature set would undermine the system's generality and would contribute to over fitting.

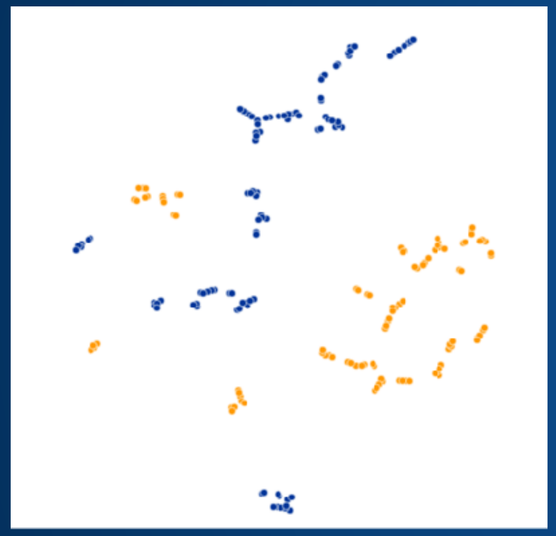
Intrusion Detection with Unlabeled Data Using Clustering

Leonid Portnoy, **Eleazar Eskin** and **Sal Stolfo**
Department of Computer Science
Columbia University
New York, NY 10027
{lp178,eeskin,sal}@cs.columbia.edu

Contact Author: Eleazar Eskin (eeskin@cs.columbia.edu)
Keywords: intrusion detection, anomaly detection, clustering, unlabeled data

Illustration of Parameter Choices and Their Failures

- t-SNE clustering of network traffic from two types of machines

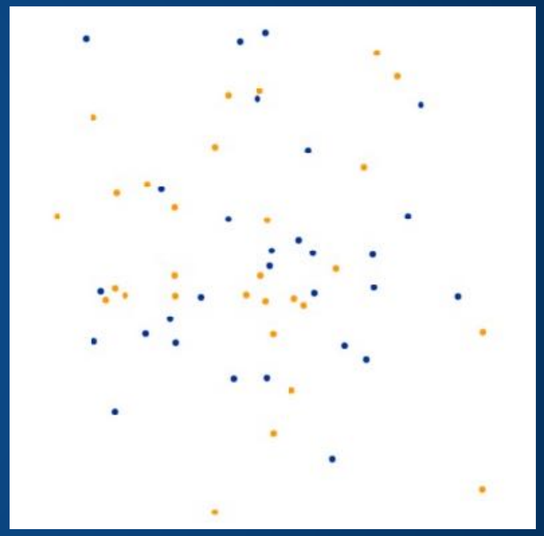


perplexity = 3
epsilon = 3

No clear separation

perplexity = 3
epsilon = 19

3 clusters instead of 2

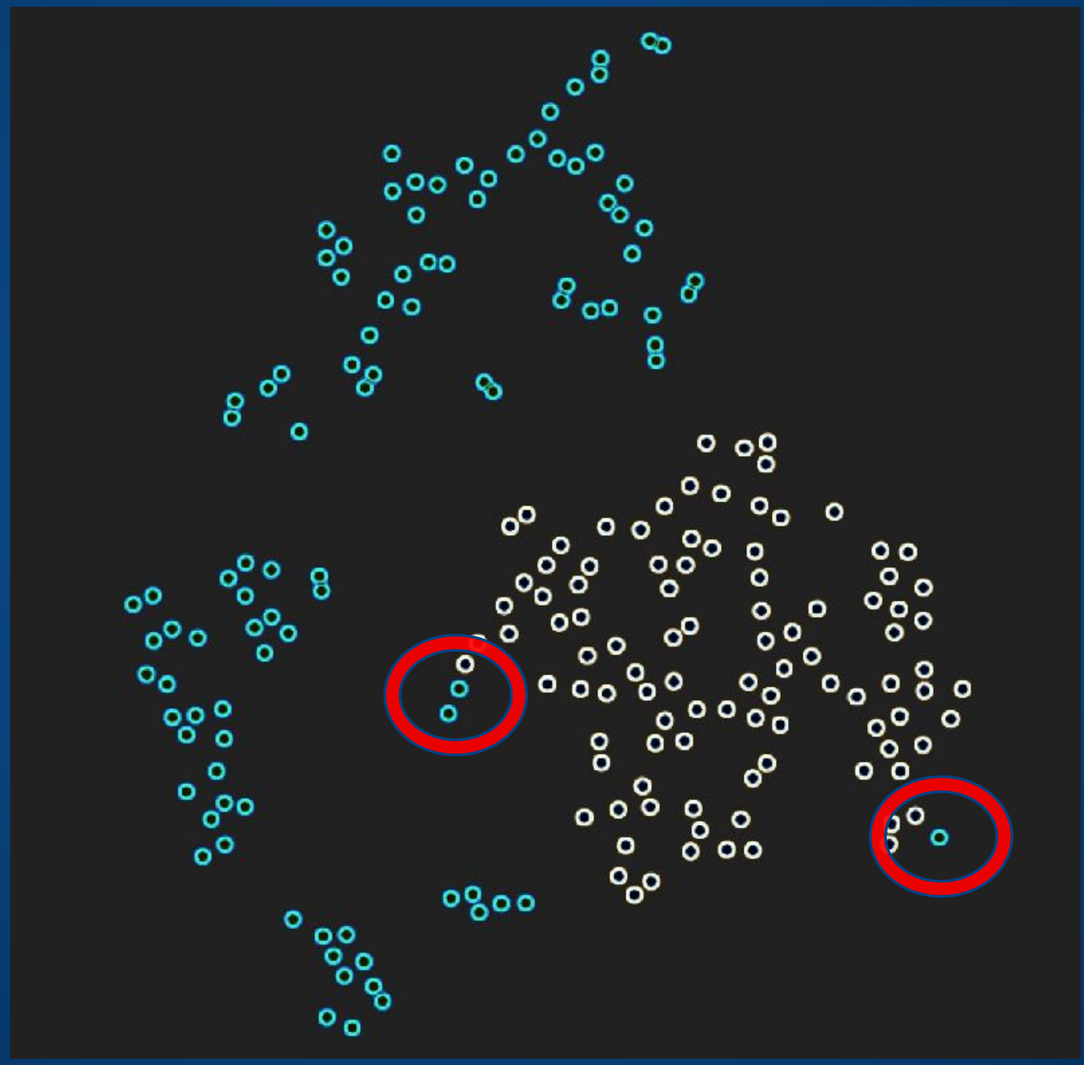


perplexity = 93
epsilon = 19

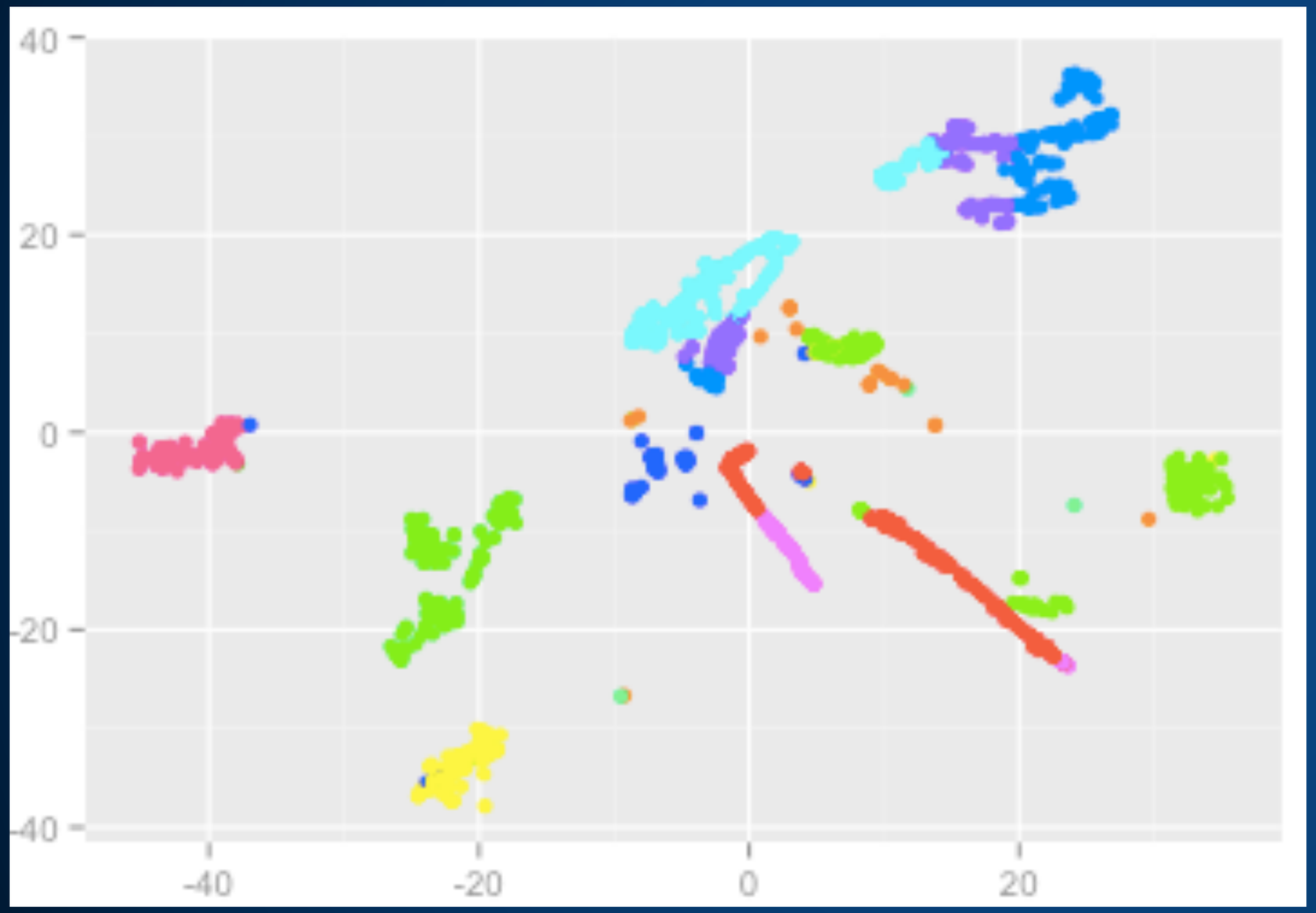
What a mess

Illustration of Parameter Choices and Their Failures

- Dangerous clusters



Network Traffic - Unsupervised Attempt



The graph shows an abstract space with colors being machine identified clusters.

Preparation:

- ***Feature engineering***
- ***Distance functions*** *(what's similar?)*
- ***Algorithm parameters***

Hard Questions:

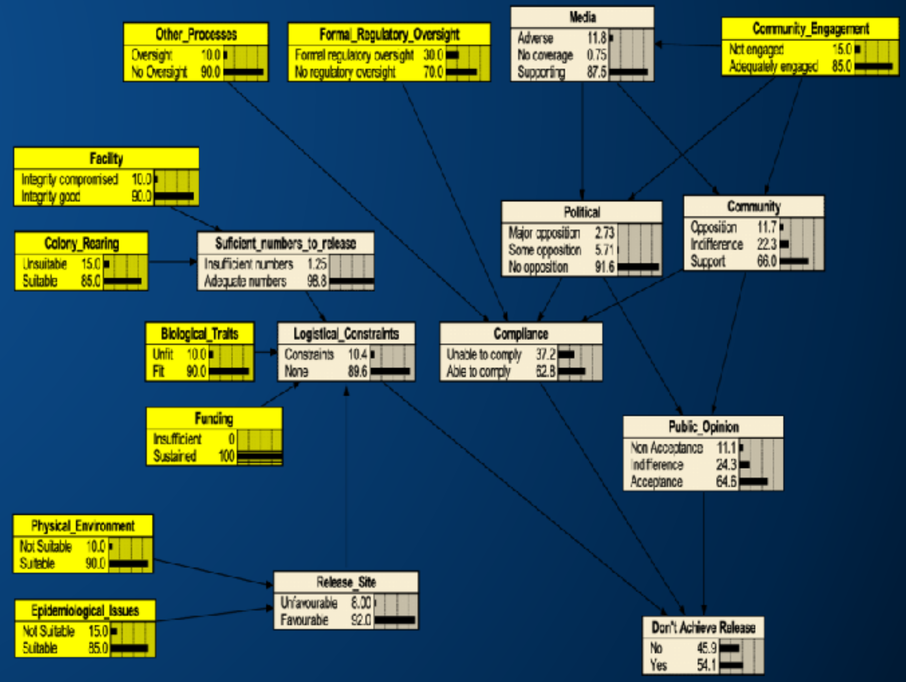
- ***What*** are these clusters?
- What are ***good*** clusters?
- What's ***anomalous***? What are the attacks?

The Real Problems

- Missing context
  - Asset inventory
  - User information
- Missing expert knowledge
  - Domain expertise

Possible Solutions

- Enable data exploration
- Improved Human Computer Interfaces & visualization
- (Bayesian) Belief Networks



In Summary

Summary

- Build solutions for actual problems with real data that produce actionable insight
- Encode expert knowledge - leverage experienced experts
- Use simple systems - how about letting users give input? Push problem to the edge
- Don’t start with the algorithms - EVER
- Start with the problem at hand and choose the right approach (hardly ever ML)
- From the problem gather the right data and context
- Use ML for problems where you have a large corpus of well labeled data
- Chose meaningful distance functions
- Verify your models - use visualization to help with that
- Measure through user feedback, what you have implemented makes sense and pleases users
- Allow for expert supervision - feedback loops
- Share your insights with your peers – security is not your competitive advantage

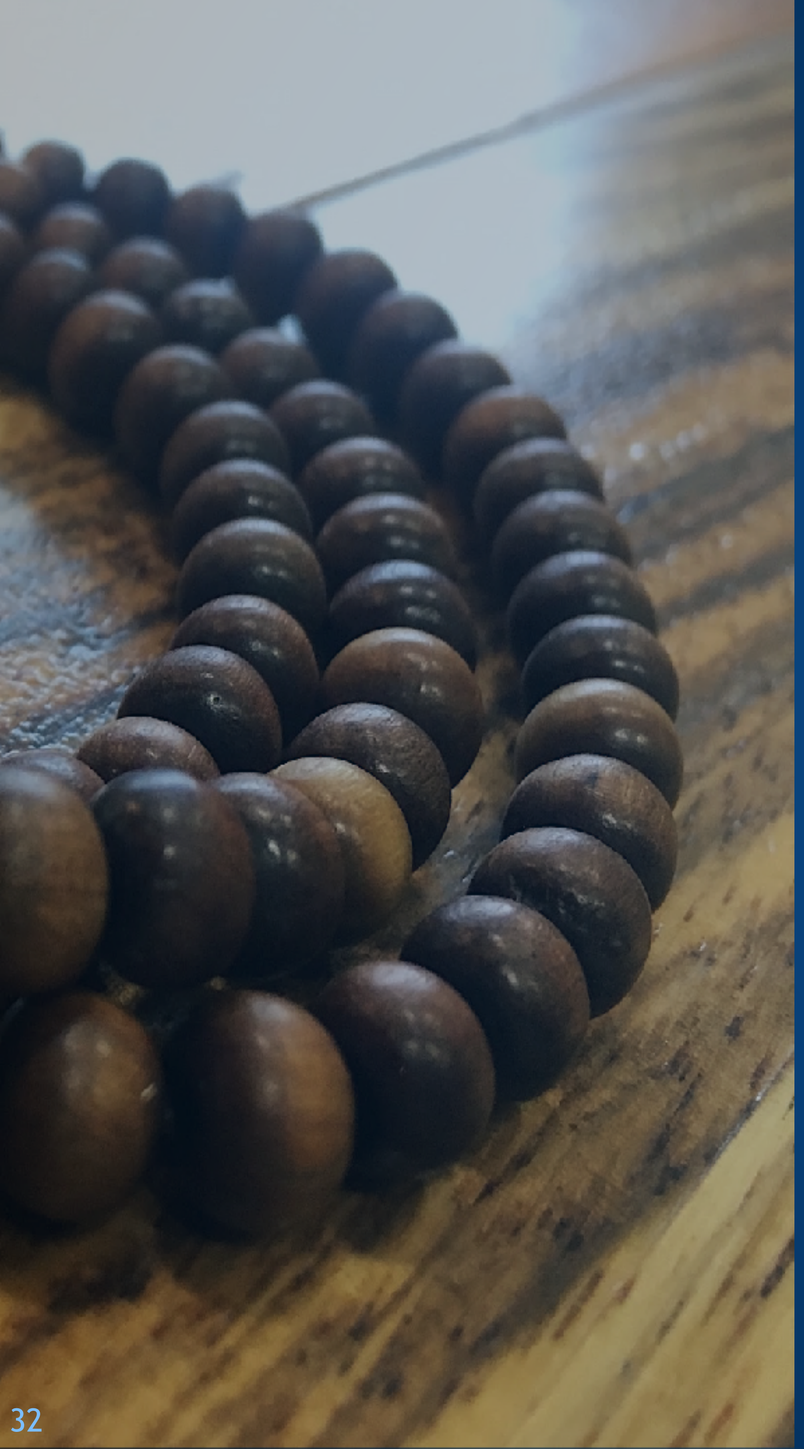
BlackHat Workshop

Applied Machine Learning
for
Identity and Access Management

ML | AI | IAM

August 4,5 & August 6,7 - Las Vegas, USA

http://secviz.org



"You can hear the sound of two hands when they clap together," said Mokurai. "Now show me the sound of one hand."

Questions?

http://slideshare.net/zrlram
@raffaelmarty