Applications of Machine Learning to Cyber Security

Assefaw Gebremedhin and James Halvorsen

School of Electrical Engineering and Computer Science

Washington State University

CySER Summer Workshop

May 31, 2022

Two Parts

Overview of CySER	Assefaw Gebremedhin
Machine Learning in Cybersecurity	James Halvorsen

What is CySER?

- An Institute funded by the Department of Defense Air Force Command through the VICEROY initiative
  - VICEROY = Virtual Institutes for Cyber and Electromagnetic Spectrum Research and Employ
  - VICEROY Institutes are managed by the Griffiss Institute
- Directly responds to the VICEROY call
  - Training ROTC and DoD-aligned civilians in cybersecurity at the undergraduate and graduate level, with primary emphasis on undergraduate
- Builds a strong consortium in the Pacific Northwest for cybersecurity education and research
  - CySER brings together 5 institutions with complementary strengths and diversity of populations served
- Seeks to position WSU to attain Center of Academic Excellence in Cyber Operations (CAE-CO) designation
  - WSU will be starting a new BS in Cyber Operations program

CySER: Institutions and People

Washington State University (WSU)

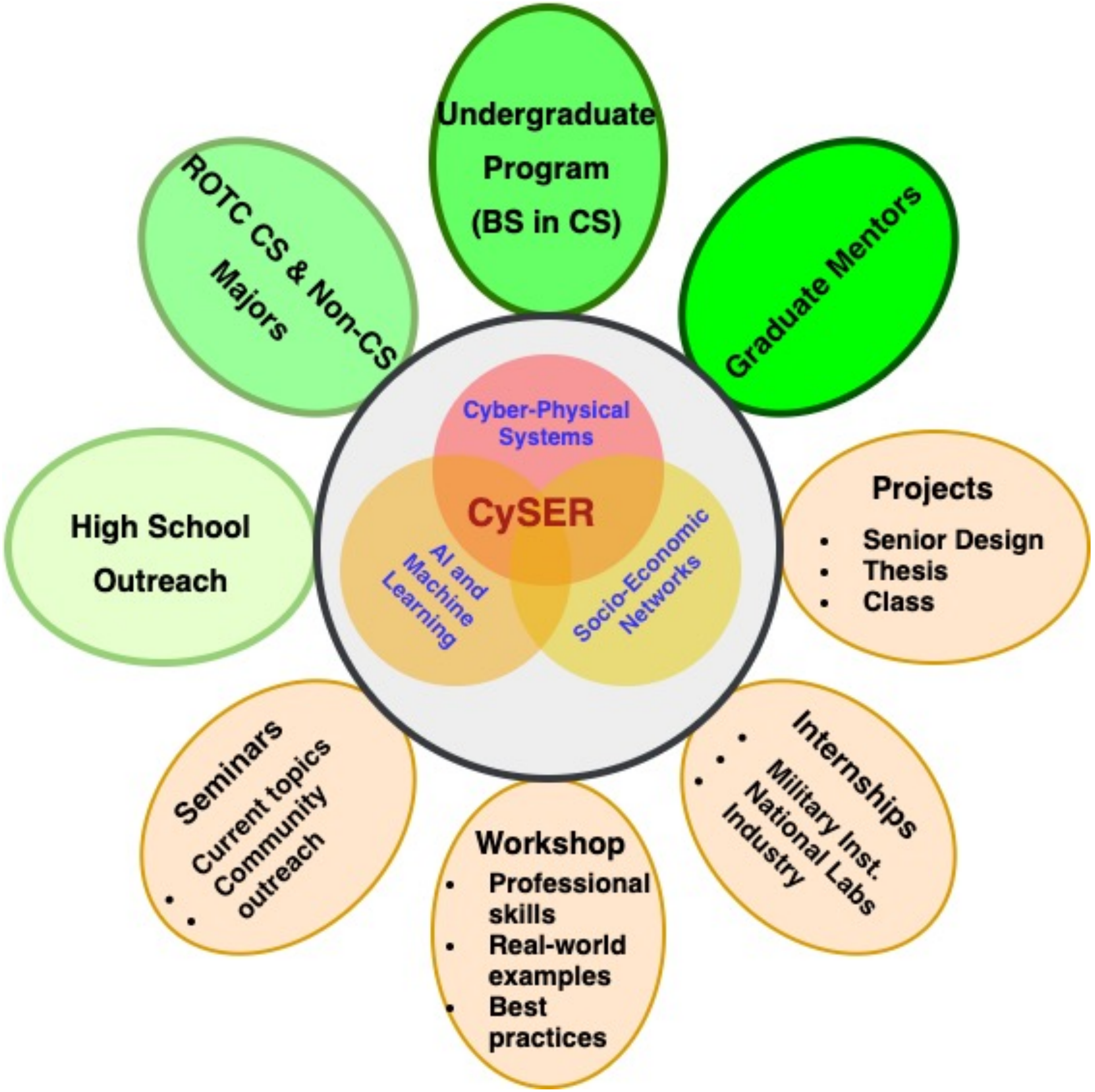
- Bernard Van Wie (VSCBE; Lead PI)
- Assefaw Gebremedhin (EECS; Co-PI; Research Lead)
- Noel Schulz (EECS; Co-PI; Industry Lead)
- Venera Arnaoudova (EECS; Co-PI; CS Curriculum)
- Olusola Adesope (Education; Evaluator)
- Partha Pande (EECS; SP)
- Haipeng Cai (EECS; SP)
- Robert Crossler (MISE; SP)
- Jana Doppa (EECS; SP)
- Arda Gozen (MME; SP)
- Larry Holder (EECS; SP)
- Chris Hundhausen (EECS; SP)
- John Miller (EECS; SP)
- Gabriel Nketah (Project Coordinator)

WSU/UI ROTC

- Lt. Col. Nicholas Jeffers
- Major Paul Hyde

- **Montana State University (MSU)**
  - Clemente Izurieta (MSU Site Lead)
  - Lt. Col. Lance Ratterman
- **University of Idaho (UI) – CAE-CD**
  - James Alves-Foss (UI Site Lead)
  - Terence Soule
- **Columbia Basin College – CAE-CD**
  - Mathew Boehnke (former CBC Site Lead)
  - Eric Robinson (current CBC Site Lead)
- **Central Washington University**
  - Deborah Wells (former CWU Site Lead)
  - Andrew Van Den Hoek (current Site Lead)
  - Lt. Col. Michael Morris (AFROTC)
  - Bob Lupton (ITAM Chairman)

CySER Program Elements



CySER certificate offerings

- **CySER CAE-CO Fundamentals**
  - BS in Computer Science
- **CySER Basics**
  - For non-CS majors (typically ROTC cadets)
  - Primarily affiliated with the MISE program in the college of business
- **CySER Advanced**
  - MS/PhD students in CS, CE, EE, MISE or similar field

CySER Research Topics and examples from this workshop’s sessions

CYBER-PHYSICAL SYSTEMS

- Natan Kipp's presentation on **industrial control systems**
- Noel and Tim Schulz's presentation **power grid security**
- Parta Pande’s presentation on **on-chip communication**
- Arda Gozen's presentation on **simulation of cyberattacks to biological systems**

NETWORKS & INFORMATION SECURITY

- Larry Holder's presentation on **graph mining for insider threat detection**
- Rob Croslier and Julia Stachofsky’s presentations on **behavioral threats and cyber warfare**

MACHINE LEARNING & AI

- **This presentation**
- Jana Doppa's presentation on **human-in-the loop anomaly detection**
- Khyati Panchal's presentation on **self-organizing maps for software vulnerabilities**

SOFTWARE SECURITY & QUALITY ASSURANCE

- Anthony Cochenour's presentation on **trusted software bills**
- Slater Weinstock's presentation on **securing supply chains**
- Clem Izurieta’s tutorial on **digital forensics**
- Deb Well’s presentation on **virtulaization**
- Haipeng Cai's presentation on **smart phone security**
- Jim Alves Foss’s tutorial on **reverse engineering**

CYBER EDUCATION

- Chris Hundhausen, James Crabb and Sola Adesope's sessions on **instructional design**
- Maj. Paul Hyde and Andrew Van Den Hoek’s session on **team building and leadership**
- Matt Boehnke and James Fulmer’s sessions on **scenarios in military applications and cyber competitions**

Part II

Overview of CySER

Assefaw Gebremedhin

Machine Learning in Cybersecurity

James Halvorsen

Outline of ML in Cybersecurity

Crash Course on Machine Learning (ML)

Issues Related to Data in Cyber Security ML Applications

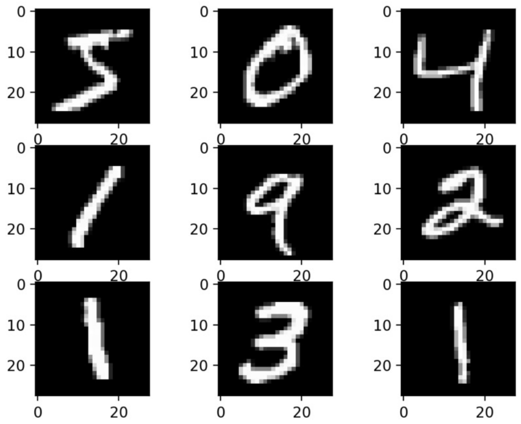
Applications of ML to Cyber Security

- Intrusion Detection
- Incident Response
- Automated Red Teaming

Is ML Worth it for Cyber Security?

A Crash Course on Machine Learning

- A very informal definition of Machine Learning:
    - “A collection of techniques that enable a machine to learn the definition of a function”
- Input for learned function: Feature Vector (typically)
- Output of learned function depends upon task
- Usually divided into three main approaches
    - Supervised
    - Unsupervised
    - Reinforcement Learning



Example of a machine learning task: Digitizing Text

Input: pixel array
Output: letter/number

Supervised Machine Learning

Scenario: We have a dataset with both inputs and outputs to a function

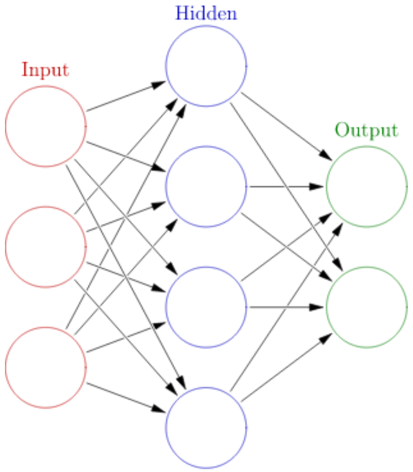
Want to learn a function that can correctly predict outputs, given the correct input

Will need to divide dataset into “training” and “testing” subsets to get this right.

Task is generally one of:

Classification: Output is a class label (e.g. “stop sign”)

Regression: Output is a real value (e.g. 3.3598)



Example: Neural Networks

- First layer takes input from dataset
- Middle layers apply weights + activation functions
- Final layer uses activation function to produce output

Unsupervised Machine Learning

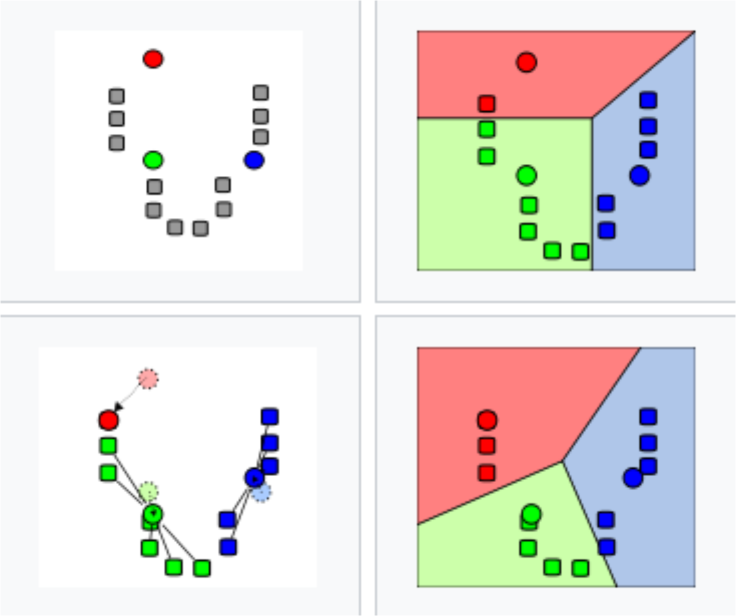
Scenario: We have a dataset with no output labels

Want to find a way to categorize values from patterns within the dataset

May require some assumptions (i.e. how many classes are there)

Related: Semi-Supervised Learning.

- Dataset has labels for some values, but not all
- Can use those labels to guide otherwise unsupervised learning

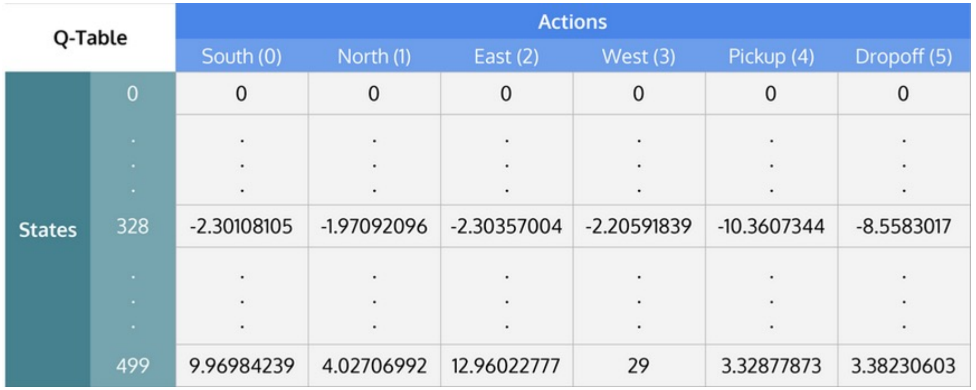


Example: K-Means Clustering

- Start with n randomly selected centroids
- Move values towards nearest centroid
- Continue until nothing changes class

Reinforcement Learning

- **Scenario:** We have an environment for an intelligent agent to act upon
- Want to learn a policy that defines how the agent should act
- Each action changes state within the environment and offers a reward
- Learning is focused on maximizing a cumulative reward
- **Example – Q Learning:**
  - Have no model of the environment
  - Create a table called Q with values for each state/action
  - Initialize values as 0, update as we receive reward
  - Agent will randomly either **explore** or **exploit** (take action with highest Q value in state)



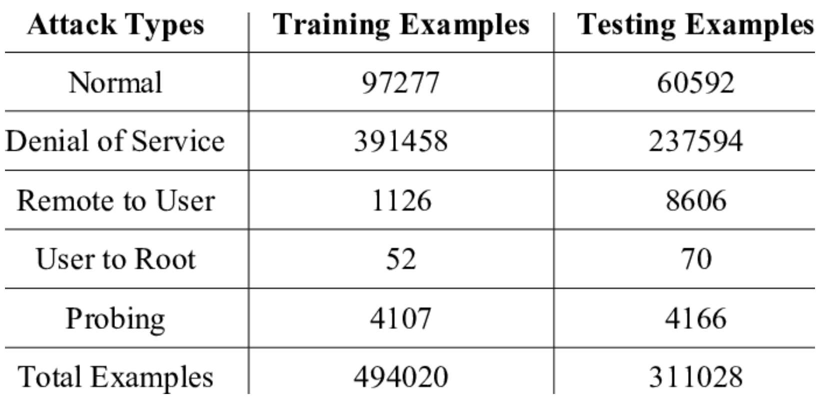
Q-Table		Actions					
		South (0)	North (1)	East (2)	West (3)	Pickup (4)	Dropoff (5)
States	0	0	0	0	0	0	0
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	328	-2.30108105	-1.97092096	-2.30357004	-2.20591839	-10.3607344	-8.5583017
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	499	9.96984239	4.02706992	12.96022777	29	3.32877873	3.38230603

Some Questions to Think About

What are some challenges that might come about trying to use an ML algorithm for cyber security problems?

Could there be security concerns for ML problems, even when ML is not used for a cyber security application?

Data: A Challenge for Cyber Security ML Applications

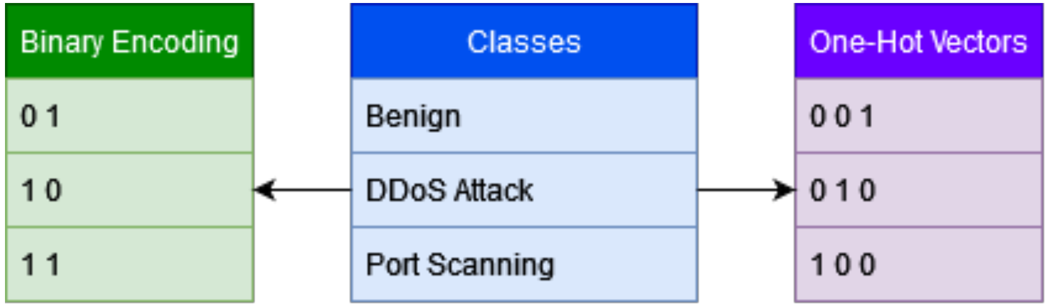


Attack Types	Training Examples	Testing Examples
Normal	97277	60592
Denial of Service	391458	237594
Remote to User	1126	8606
User to Root	52	70
Probing	4107	4166
Total Examples	494020	311028

KDD99: A security dataset with an *okay* balance of classes. But can you spot the major problem?

- High quality security data is relatively sparse
- Labeling data for supervised approaches takes time
- Simulating attacks takes time and expertise
- Organizations that have undergone cyber attacks are reluctant to share data
- Existing datasets become outdated quickly
- Not all data collected through cyber monitoring sources may be relevant for our tasks

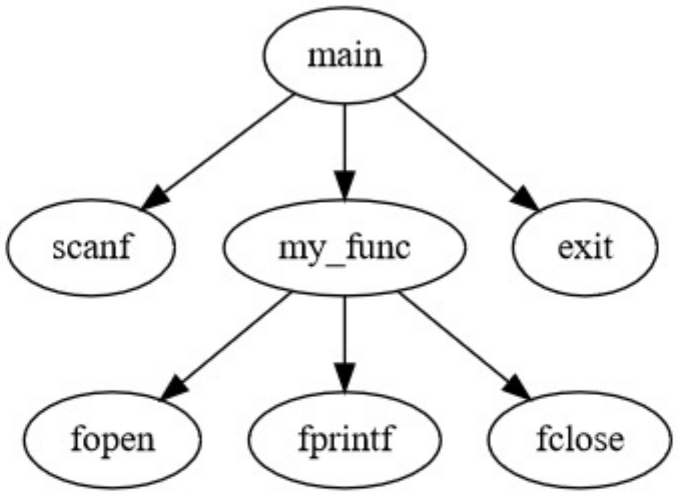
More Issues Related to Data (Preprocessing)



Binary Encoding	Classes	One-Hot Vectors
0 1	Benign	0 0 1
1 0	DDoS Attack	0 1 0
1 1	Port Scanning	1 0 0

- Most security data is either:
  - Highly categorical (e.g. filenames, IP addresses, alert IDs)
  - Nominal/Integers (e.g. number of packets)
- ML algorithms typically designed to work with continuous features (floats)
- Straight mapping of category labels to floats is bad
  - What does an off by 1 error mean here?
- Some ways of representing categorical data
  - One-Hot vectors
  - Binary/Integer encoding

Issues Related to Structure of Data



A function call graph. Can we find a straightforward transformation from this into a feature vector?

Traditional ML algorithms expect data to be a matrix (set of equal-size feature vectors)

Security data may have irregular structure

Function call graphs

Time-series data

Events in security data may not be independent

2+ events may be fine individually, but not collectively

May need specialized algorithms

Graph Neural Networks

Structured Prediction

Overview of Machine Learning Applications for Cyber Security

Intrusion Detection

- Given: A set of monitoring data from a machine/network
- Task: Determine if any behavior is associated with a threat

Incident Response

- Given: A set of actions from a known adversary
- Task: Decide upon an appropriate response

Automated Red Teaming

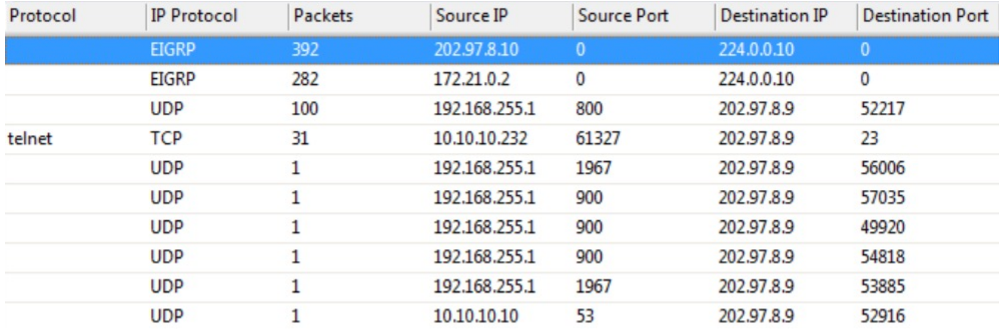
- Given: Initial environment and a goal state
- Task: Find a series of actions to carry out a simulated attack

Intrusion Detection: Basic Concepts

- Intrusion Detection Systems (IDS) monitor activity on a network or system, and report on threats
- Several types of implementation details
  - Signature-Based vs Anomaly-Based
  - Host-Based vs Network-Based
  - Rule-Based (expert system) vs Statistical Approach
- Generally confined to raising alerts
- Where response also occurs, this is an intrusion prevention system

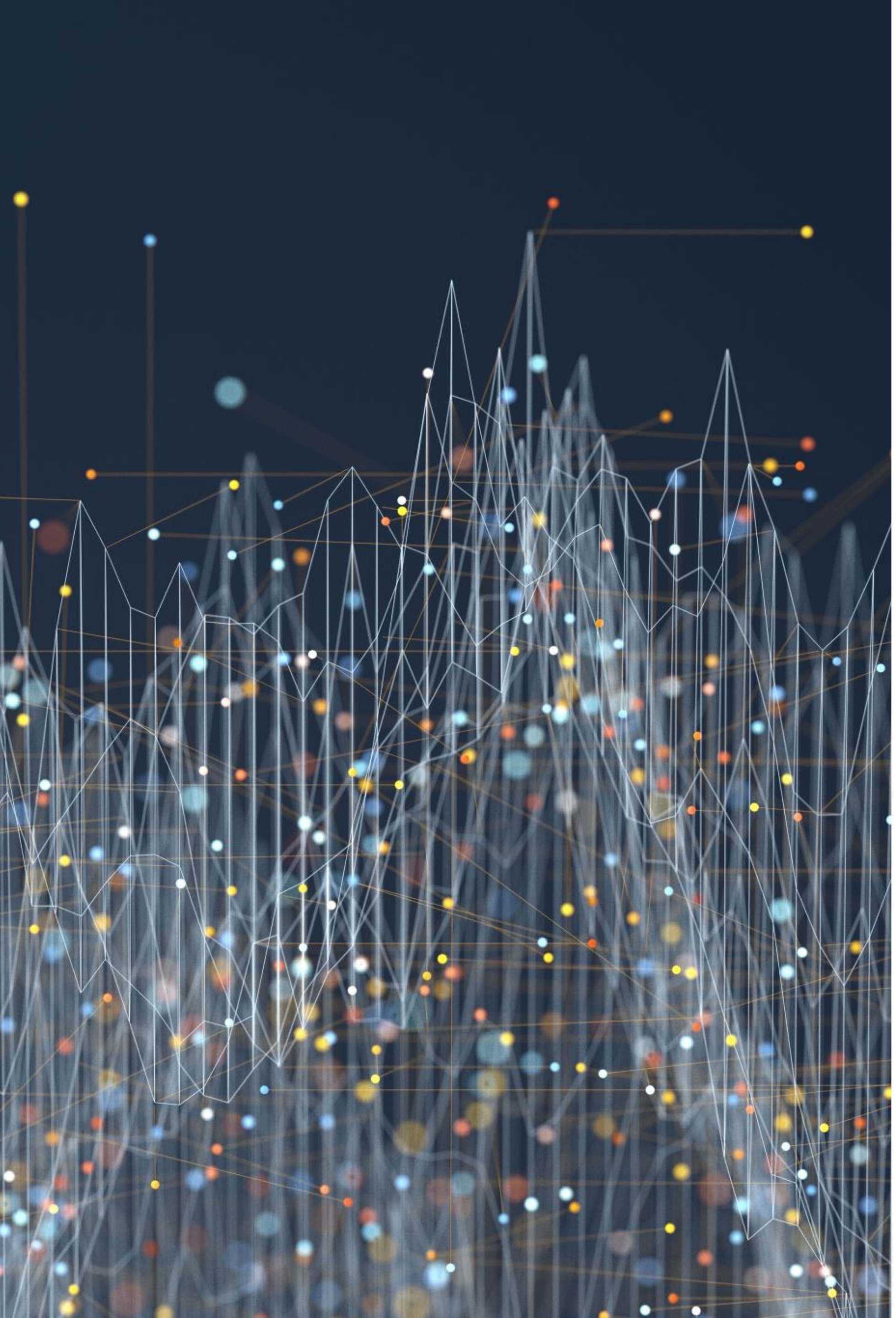
Data Collection in IDS

- Type of Data to collect depends upon scope of IDS
- Host-Based IDS (HIDS) may have more varied data sources
  - Incoming and outgoing packets
  - Filesystem changes
  - Process changes
- Network-Based IDS (NIDS) focused on traffic across a network
  - Netflow (summary of packets sent/received)
  - Packet captures (emphasis on content)



Protocol	IP Protocol	Packets	Source IP	Source Port	Destination IP	Destination Port
	EIGRP	392	202.97.8.10	0	224.0.0.10	0
	EIGRP	282	172.21.0.2	0	224.0.0.10	0
	UDP	100	192.168.255.1	800	202.97.8.9	52217
telnet	TCP	31	10.10.10.232	61327	202.97.8.9	23
	UDP	1	192.168.255.1	1967	202.97.8.9	56006
	UDP	1	192.168.255.1	900	202.97.8.9	57035
	UDP	1	192.168.255.1	900	202.97.8.9	49920
	UDP	1	192.168.255.1	900	202.97.8.9	54818
	UDP	1	192.168.255.1	1967	202.97.8.9	53885
	UDP	1	10.10.10.10	53	202.97.8.9	52916

Example of netflow data. The content of packets is not known, but we know the **who**, **when**, and **how**



Machine Learning in IDS

- ML Approaches to Intrusion Detection focused primarily on Anomaly-Based detection
- Want to learn function mapping [Events] -> [Benign|Anomaly]
- Dedicated Anomaly-Detection Algorithms
  - Isolation Forest
  - One-Class SVM (creates a hypersphere around data points)
- Supervised Approaches
  - Need mixed, labeled data
  - Numerous high quality algorithms
    - Deep Neural Networks
    - Random Forests
- Regardless of approach, IDS needs to know what is normal for *your* network.

Difficulties with Machine Learning in IDS

Data Problems

- Availability of labeled data
- Need for effective pre-processing

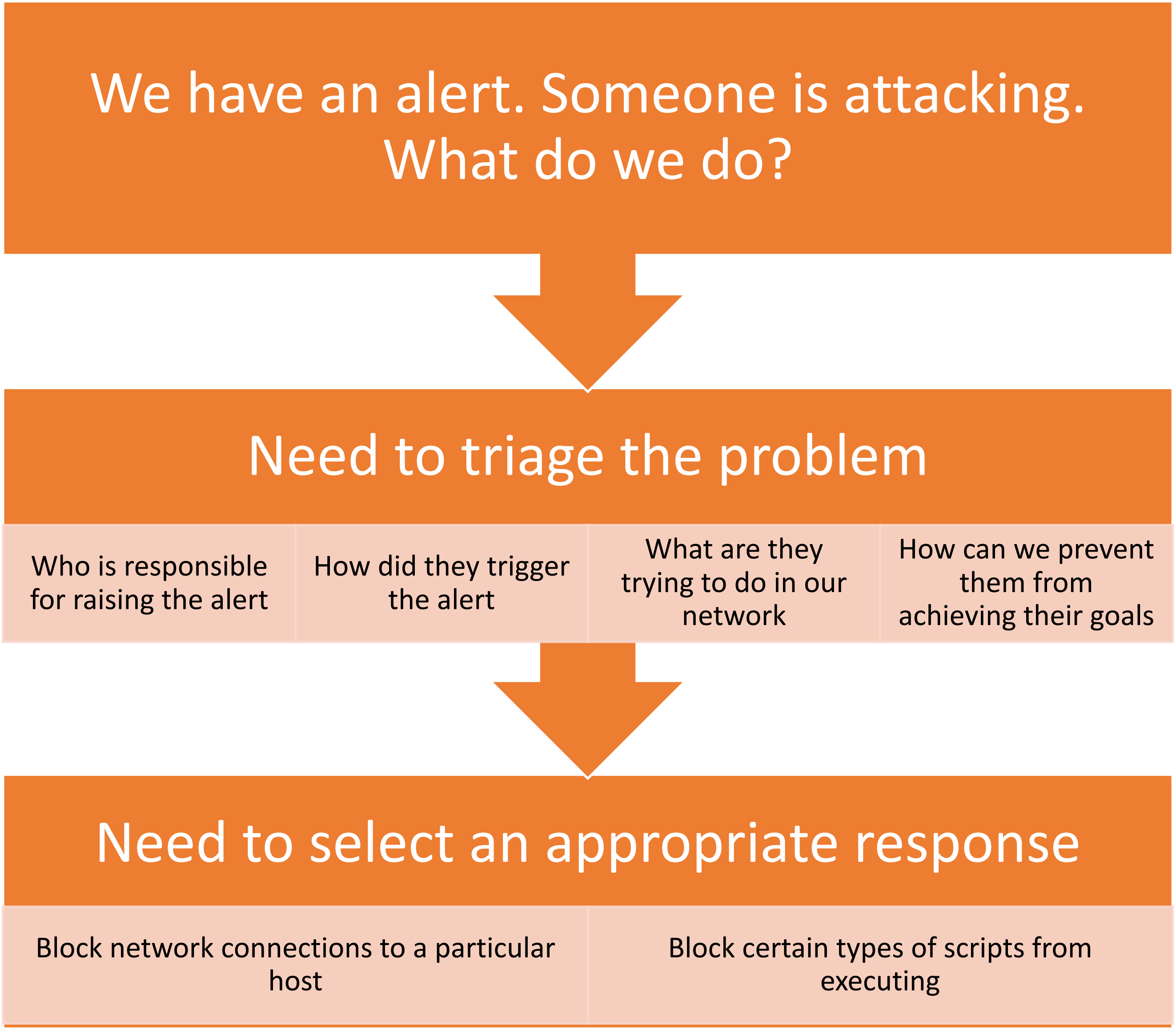
False Positives

- Can lead to incorrect responses (denies availability).
- May cause alerts to be ignored, IDS usage abandoned.

Zero Day Attacks

- Significant problems for anything Signature-Based.
- Anomaly-Based approaches still imperfect.
- If attacker has classifier model, can create attacks that evade detection.

Incident Response: Basic Concepts



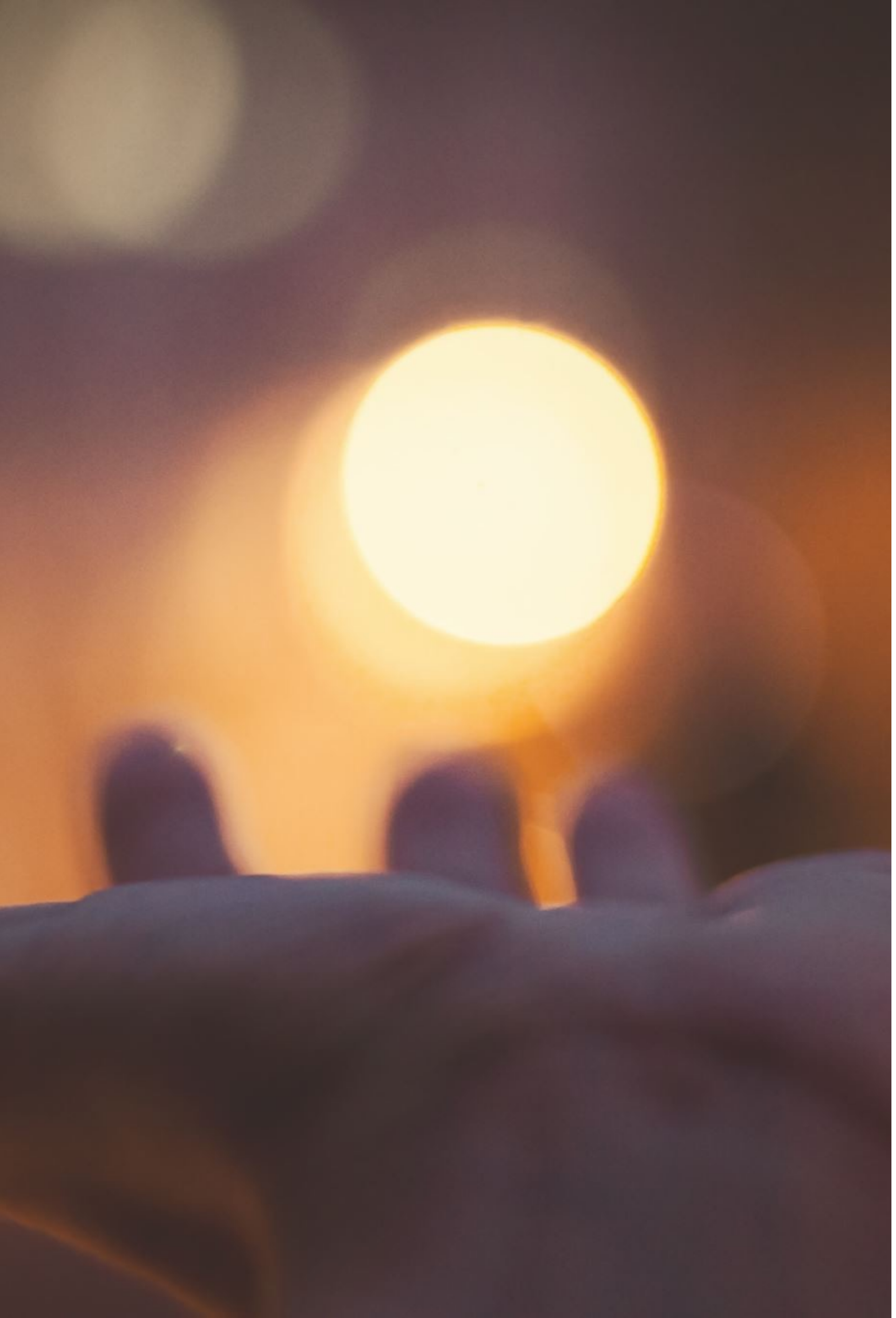
We have an alert. Someone is attacking. What do we do?

Need to triage the problem

- Who is responsible for raising the alert
- How did they trigger the alert
- What are they trying to do in our network
- How can we prevent them from achieving their goals

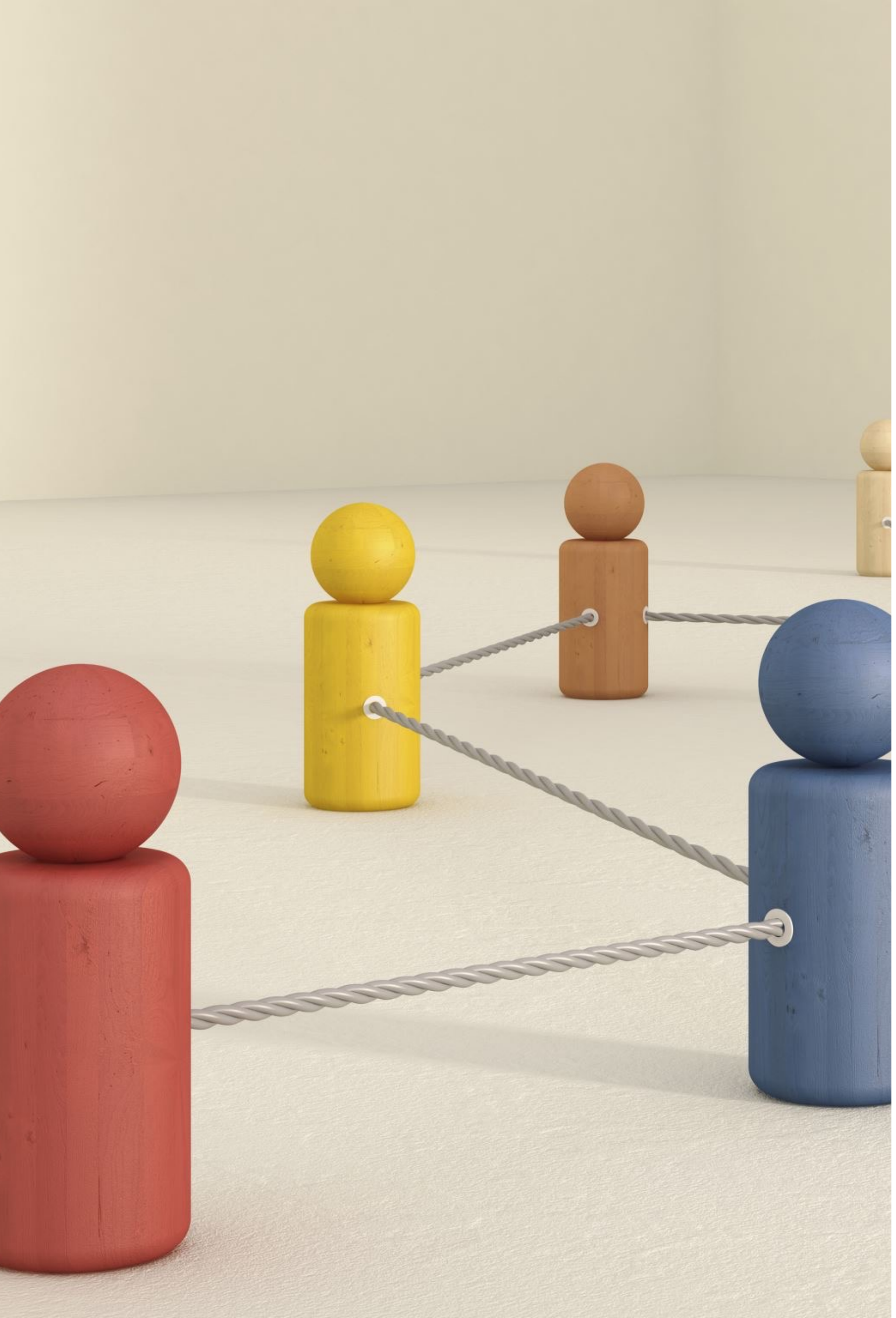
Need to select an appropriate response

- Block network connections to a particular host
- Block certain types of scripts from executing



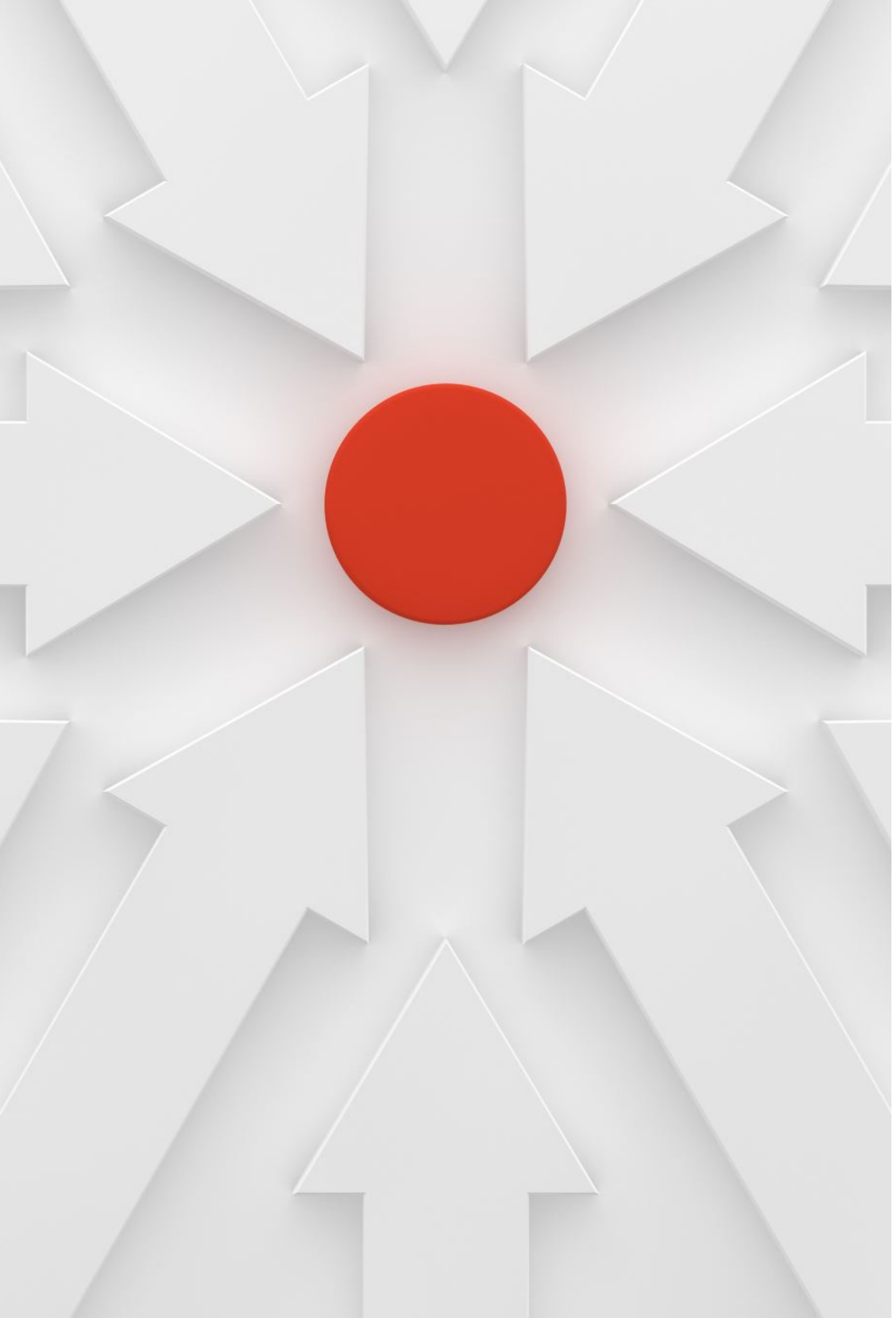
Machine Learning in Incident Response

- Not as well researched as intrusion detection but goes hand in hand with the problem
- Most research focused on helping humans to respond
  - Try to find other actions from the attacker
  - Help find relevant supporting information
  - Maybe suggest an action
- Hypothetical: Fully-Automated Incident Response
  - Assume we have an IDS that never gives a false positive
  - Train a classifier on a number of attack scenarios that provide context and correct response
  - Let the machine take that response without human input



Red Teaming: Basic Concepts

- Defensive capabilities of organization require testing to measure effectiveness
- Testing can be performed with a red/blue team setup
  - Red teams given a goal to access sensitive information, use whatever tactics that work
  - Blue teams should aim to prevent red teams from achieving their goals
  - Blue teams not necessarily aware that this is an exercise
- Generally requires human experts that understand offensive security, which can be costly
- More companies should be doing this – can we automate it?



Modeling the Problem of Automating Red Teaming

- Automating Red Teaming can be viewed as an instance of an AI planning problem
- Planning problem description
  - Given: Initial state, goal state, available actions
  - Task: Find sequence of actions to get to goal state from initial state
- Actions in planning problem will have certain properties
  - Preconditions (what must be true before action is taken)
  - Effects on environment (what is added/deleted)
  - Costs

Case Study: Privilege Escalation with Reinforcement Learning

- This slide provides a summary of a research paper pre-print[1]
- Environment: Simulated Windows machine with randomly selected combinations of vulnerabilities
- Actions: 38 carefully constructed actions
  - Some designed to exploit specific vulnerabilities, others more general
  - Examples: test credentials, overwrite a DLL
- Uses Actor Advantage Critic (A2C) method to learn policy
- Result: Agent was able to learn several methods of privilege escalation, some of which avoided AV detection
- Discussion
  - RL agents can produce flexible attacks and adapt to environment.
  - Possible in future to exploit unknown vulnerabilities

[1] Kujanpää, K., Victor, W., &amp; Ilin, A. (2021). Automating privilege escalation with deep reinforcement learning. Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security. https://doi.org/10.1145/3474369.3486877

Is Machine Learning worth it for Cyber Security?

Perfecting results can be a hard problem

- Depending on task, misclassification can be very bad
- ML models with 100% accuracy when training are “overfit”
- Also can’t have any model perfect for all use cases (no free lunch theorem)

Can reduce dependence on humans

- Can you response to an attack while you sleep?
- Or while you’re sick with COVID?
- Is it economically feasible for every company to have human security experts?

Possible that our adversaries may use ML

- Human Defender vs ML Attacker
- ML Defender vs ML Attacker
- Will humans still be winning this arms race decades later?

Some More Questions to Think About

What are some other ways that you think ML could be used to solve cyber security problems?

What would be the limitations of ML in those contexts?