ODSC EAST 2018, Boston
Exploiting Multiclass Probabilities for Solving Network Security Anomalies using SL/USL.

Ashrith Barthur, PhD
Security Research
@cyberbaggage

H2O.ai
Machine Intelligence

Sequence of Events (SoE)

- **What is a Sequence of Events?**
  - **A set of events, that usually includes sub-events that help you achieve a goal.**

SoE - In Depth

- **An individual event is usually a set of sub-events that we/machines do to achieve a state.**
  - **E.g.** Entering username and password and hit enter - login event.
- **An event by itself does not say much.**
  - **E.g.** Did you login to Google? Facebook?
- **So an event needs a context.**
  - **E.g.** Enter www.google.com - page load event.
  - Enter username and password - login event.

SoE - Importance

- **If you are predicting loan default / fraud then a sequence of events are not that important.**
- **But when you are classifying a potential attack /malicious behaviour, sequence of events is important.**

SoE - Importance

- Is this not just about building related features?
- Not so.
- This is actually chaining data from different sources and making them a sequence, by actual data joins, or algorithmically.

Why Do We Need a Sequence of Events While Identifying Potential Attack?

- Answer lies in how attacks occur, Anatomy.

Classification of Attacks

- **Short Term Goals**
  - DDoS - for different layers
  - Physical Attacks
- **Long Term Goals**
  - Network/Service Reconnaissance
  - Enterprise Service attacks - attack on infrastructure
  - Phishing, Spear Phishing (more focussed)
  - Social Engineering - Out-of-loop
-

Anatomy of An Attack - Short Term

- Identify Target
- Identify Service of Attack
- Overwhelm the service
- **Post-Attack Analysis**
  - Attack mechanism is simple.
  - Variations occur in source of attack, protocols levels.
  - Relatively short lived.
  - Damage quantifiable.

Anatomy of An Attack - Long Term

- Identify Target
- Reconnaissance
- Identify Infrastructure Vulnerability / Or means of phishing
- Network Foothold
- Lateral movement and service compromises
- Data Exfiltration/ Network Squatting, or passive sniffing.

Anatomy of An Attack - Long Term (cont)

- **Post-Attack Analysis (*Usually an Illusion*)**
  - Attack might still continue
  - Variations can occur based on services, new vulnerabilities, new softwares, unused access, network segments without VLANs, un-closed, outdated wall sockets, etc.
  - Usually very long term
  - Damage assessment is not usually accurate.

How are these two attack variants used?

Usage

- Used Together, if needed.
- Short Term Attacks are used as:
  - A means of Reconnaissance
  - A method of shielding another attack, or breaking down some basic protection before an attack is launched.
  - It is also used to shield any detection of data exfiltration

Usage

- **As you can clearly see a potential attack is set of connected events.**
- **Identifying only one event might not yield much information.**
  - E.g. An access to the database in itself is hardly a potential attack identifier.
  - Accessing the database outside work-hours too is hardly an identifier as people all around the world might be working on the same database.

Current Day Solutions.

1. Solutions do exist that correlate events
2. But are limited
3. They are purely rule-based, and mostly stateless.
4. Hardly capable of smartly identifying events related across time. - **A must for identifying long term attacks.**

CSec Solution Evolution

Rule-based Model

Feature-based Model

Pure Data Driven Model

CSec Solution Evolution

Feature-based Model

CSec Solution Evolution

Feature-based Model

- Using a feature based model we look for anomalies / potential attacks by:
  - First marking the kind of traffic it is.
  - And the likelihood of it being malicious
- These anomalies are further verified by having a human analyse the outcome of the model.

Features - (Used in Feature-based Model)

1. Features are meta data (Extracted from the data)
2. They help algorithms capture information from the data.
3. Feature engineering is a form of language translation: Between raw data and the algorithm.
4. Build much better features for your supervised models.

Source of Data

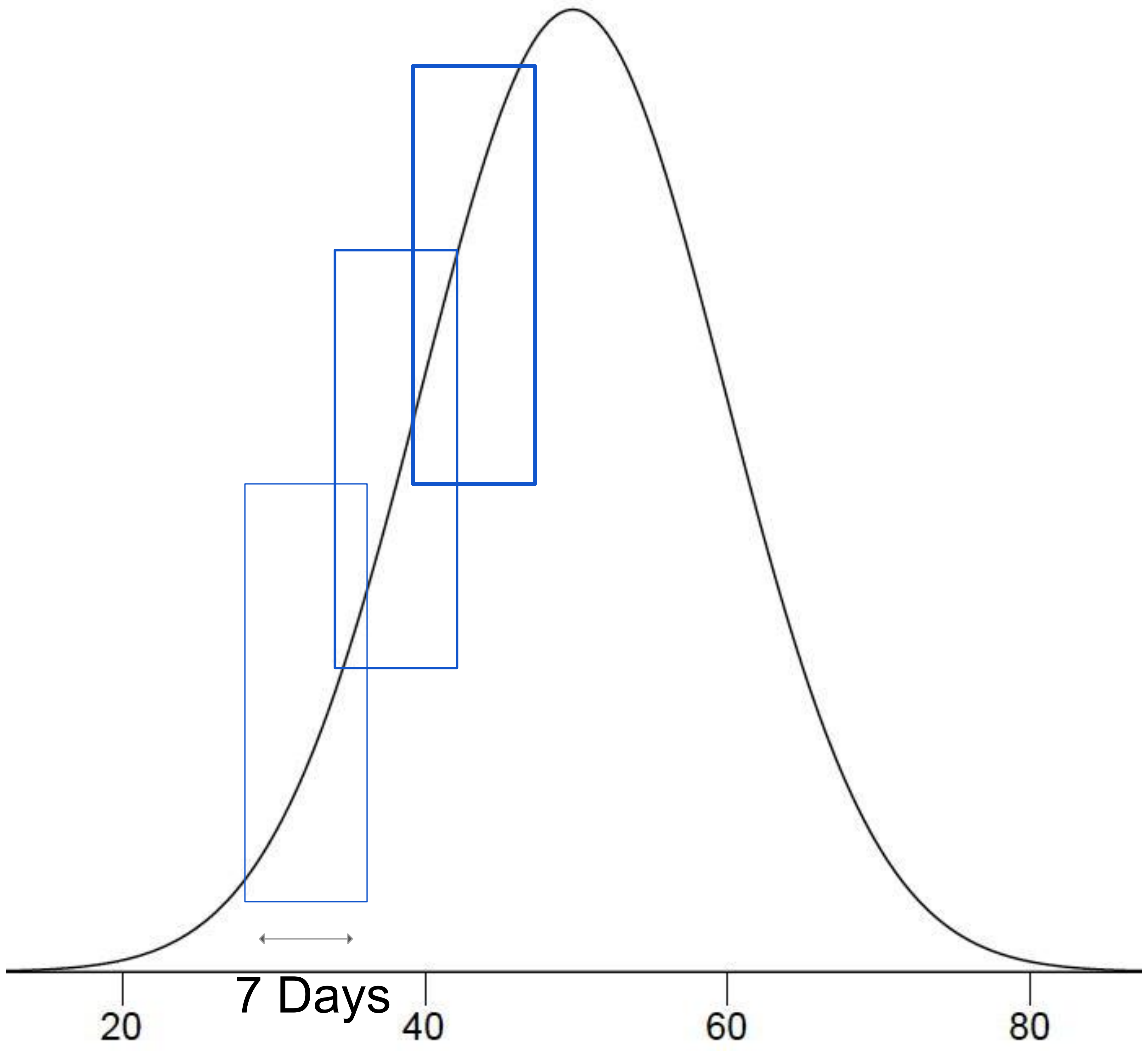
1. Past Attack
2. Past Traffic
3. Current Traffic
4. Application Logs
5. System logs
6. PCAP files - raw network capture files.
7. ASA, IDS, etc.

Features - Example

1. Average length of connection (too small, too large)
2. Average number of DNS requests (within network/outside network)
3. Average number of new domains
4. Change in MTU ratio vs. Windows/Mac/*Nix machine churn.
5. Packet Utilization - segmentation
6. Window Size
7. Arrival Jitter Variance

Features - Example

average tcp connect length by protocol



H2O.ai
Machine Intelligence

Features: Advantages

1. Designed Features Highlight Transactional Behaviour
2. Features Continuously Track Network's Transactional Behaviour
3. Rules Variables can only Identify Threshold Changes

Feature-based Model: Advantages

1. Uses AI - artificial intelligence
2. AI with features uses a consistent and objective approach
3. Quick classification
4. Multiclass - quickly identifies types of traffic - event.
5. Low false positive rate - tweaked based on risk appetite.

Limitation of the Model

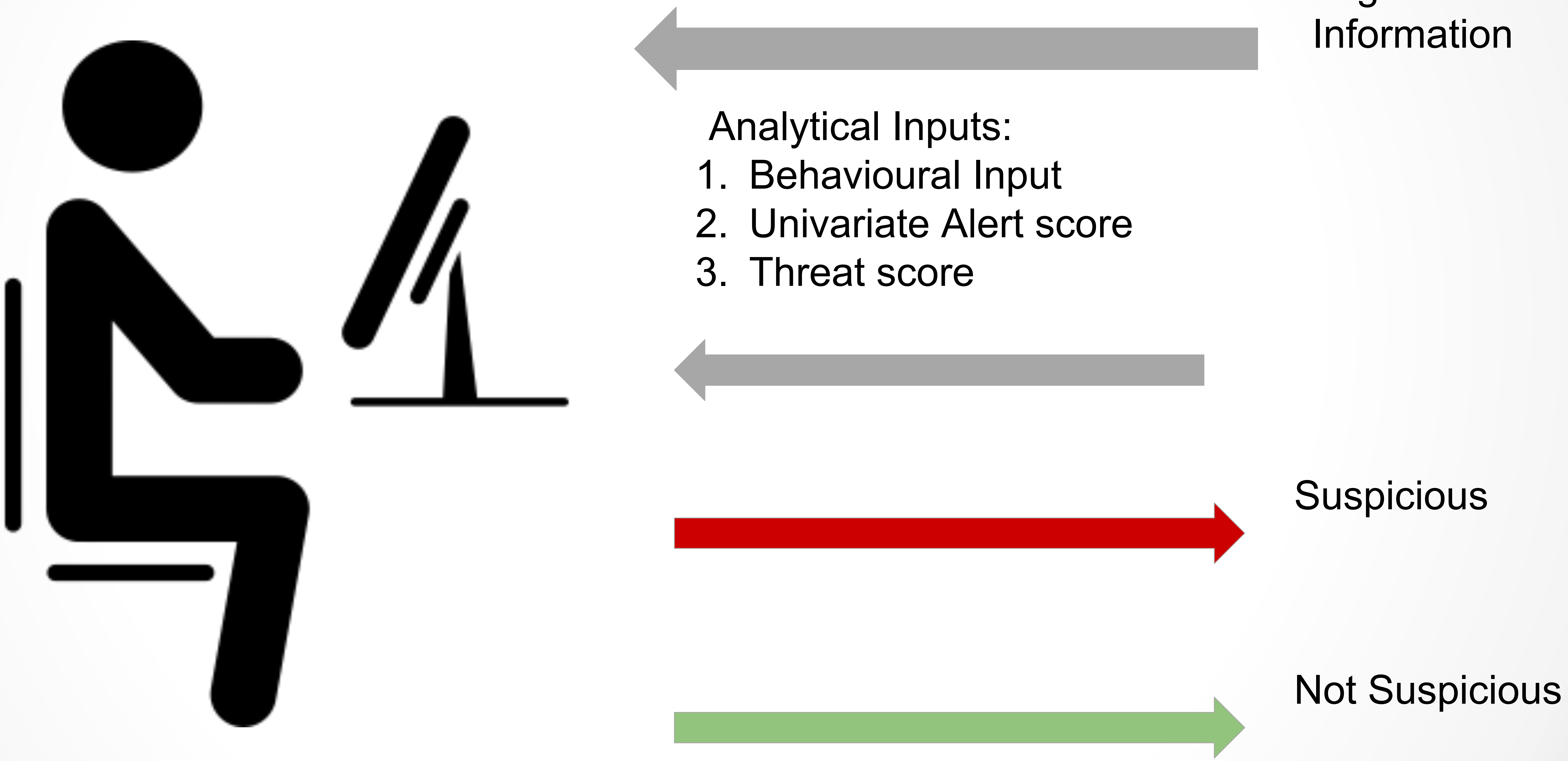
1. A single traffic classification
2. A single likelihood for the specific type of traffic.
3. It still needs to be verified by a security analyst
    a. An analyst needs to go through large amounts of data for identification

Identification and Labeling

Two different methods

1. **Completely Manual**
2. **Assisted by Clustering**

Manual Labeling



Logs
Information

Analytical Inputs:

1. Behavioural Input
2. Univariate Alert score
3. Threat score

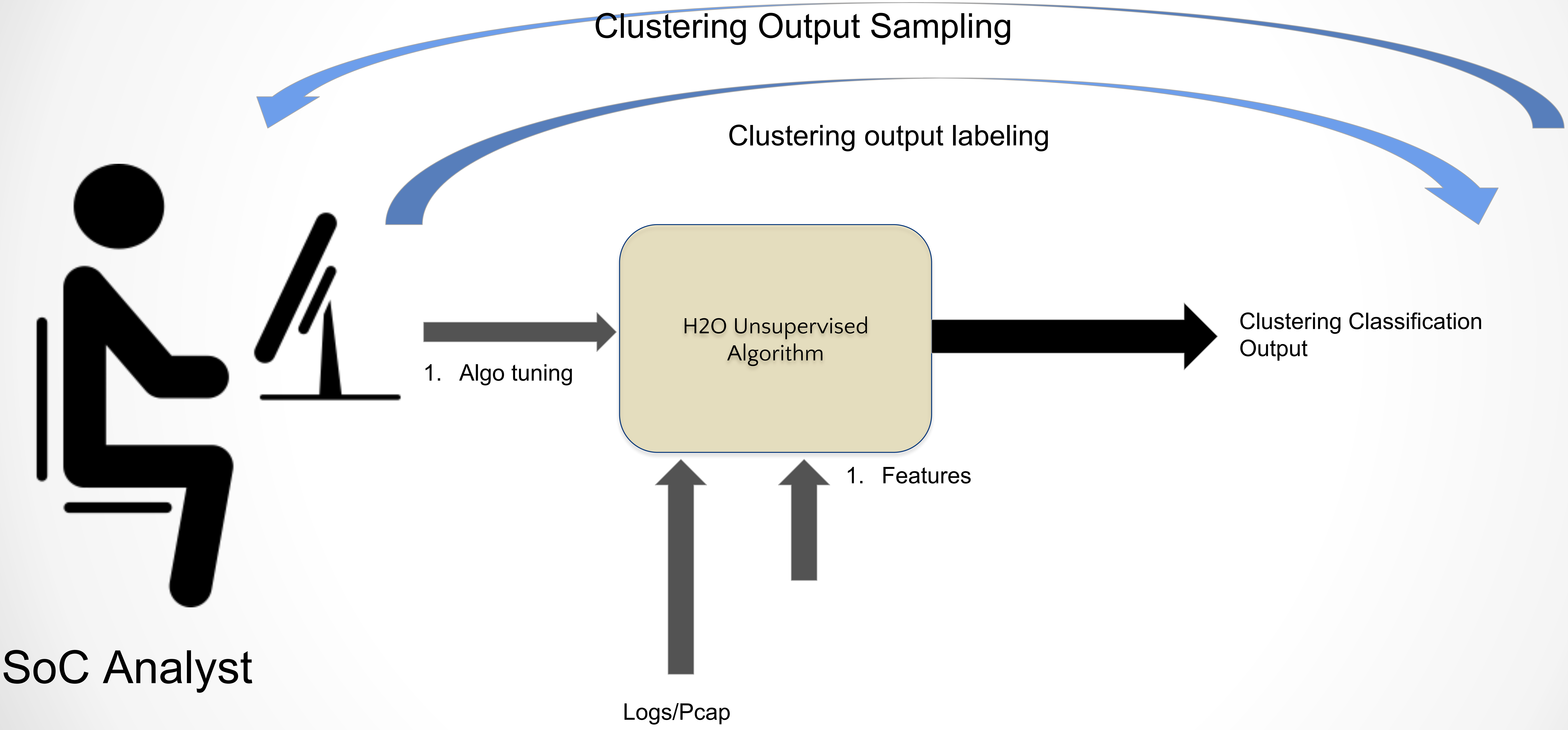
Suspicious

Not Suspicious

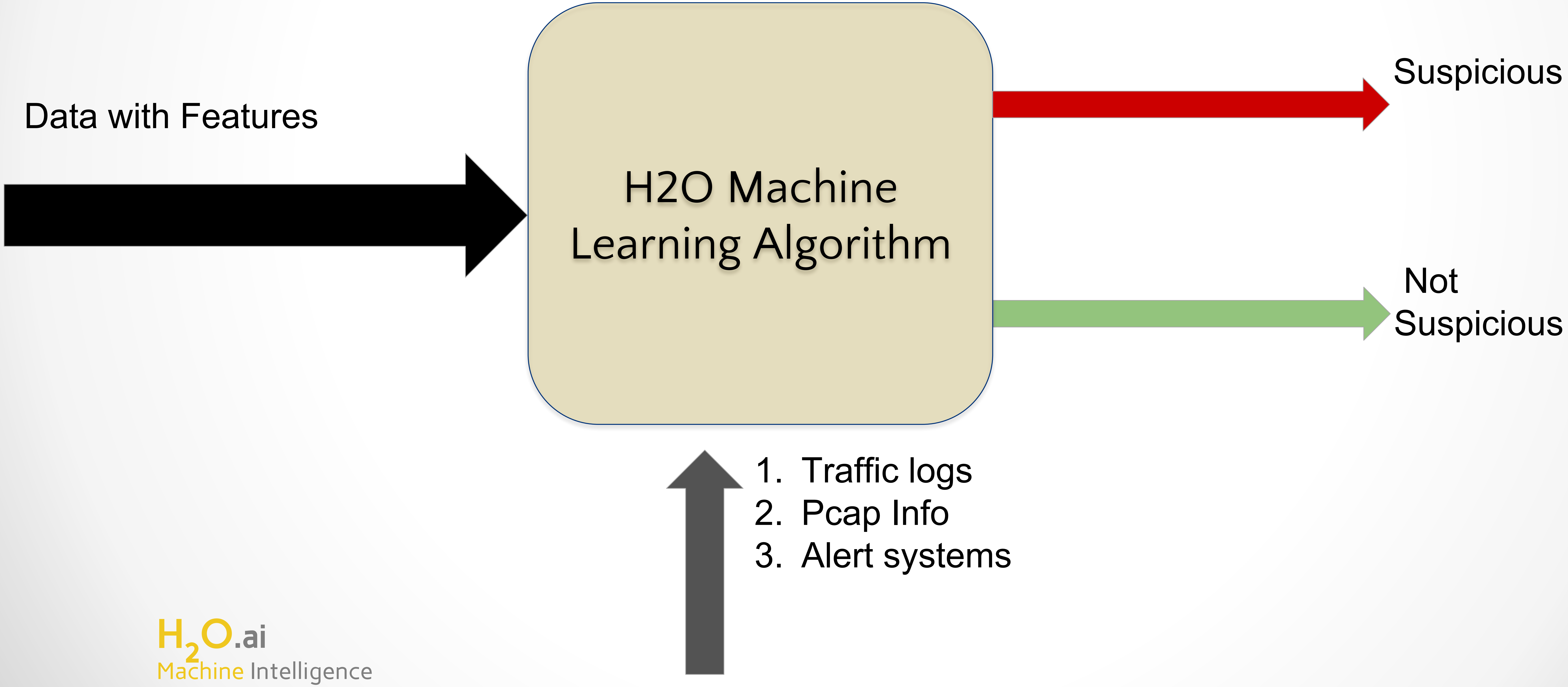
Assisted Labeling

- The approach of Manually Labeling is slow.
- Therefore, we involve an assisted Labeling approach.

Assisted Labeling


Clustering Output Sampling
Clustering output labeling
H2O Unsupervised Algorithm
1. Algo tuning
Clustering Classification Output
1. Features
SoC Analyst
Logs/Pcap

Model Deployment



Data with Features

H2O Machine Learning Algorithm

Suspicious

Not Suspicious

1. Traffic logs
2. Pcap Info
3. Alert systems

Limitation of This Approach

1. Slow
2. Loss of Classification information

Loss of Classification of Information

Output Class	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**
Class 1	0.7	0.2	0.05	0.04	0.0	0.0
Class 1	0.7	0.2	0.05	0.04	0.0	0.0
...	...	...	...	...	...	...
Class 1	0.55	0.0	0.0	0.0	0.0	0.45
...	...	...	...	...	...	...

Loss of Classification of Information

- In a multiclass ML problem we get probability scores for all possible candidates
- But we disregard all scores except the highest score.
- Benign events and potential attacks get class-probabilities in a multi-classification.
- Events that are benign, in a given class e.g. *Class 1,* tend to have similar scores.
- Events that are potential attacks in a certain class e.g. *Class 1*, tend to have different scores when compared to benign events.

Model Improvement

- We exploited this information from the multi-classification.
- The classes in multi-classification are the **sequence of events**.
- We passed the probability scores thru an autoencoder.
- By exploiting the multi-classification probability values we calculated reconstruction errors.
- Using reconstruction errors we were able to classify traffic that seemed anomalous - potential attack, and benign.

Model Improvement - Advantages

- FAST!
- Results reinforced with bit more information.
- Reinforced events are the sequence of events.
- Analyst looks at a smaller set of data and can quickly identify potential attacks.

Thank You

Questions?