ADNA: online, context-aware, intelligent framework for Anomaly Detection aNd Analysis in SCADA networks

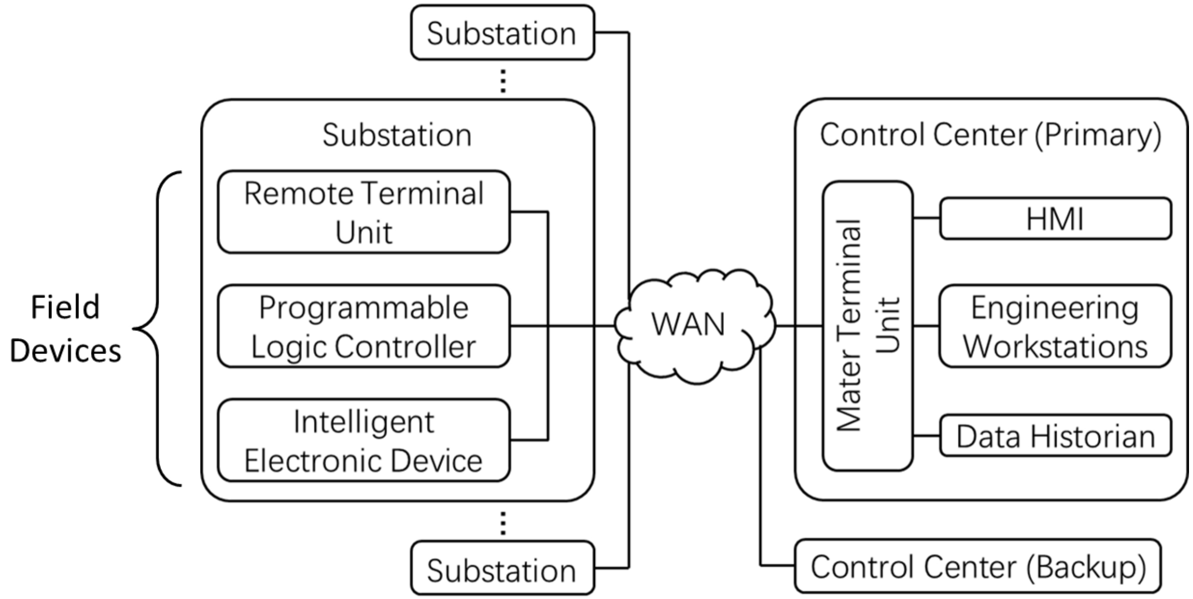
Researchers: **Wenyu Ren**, Klara Nahrstedt, Tim Yardley

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CREDC
CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Motivation

- Supervisory Control And Data Acquisition (SCADA)



- Problem with existing work
  - Fail to utilize all levels of network data in proper ways
  - Lack of further analysis of anomaly detected

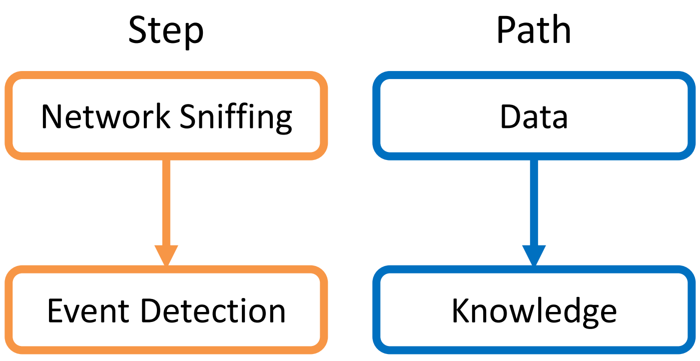
ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Motivation

- Data in SCADA networks generally can be divided into three levels:
  - Transport level: traffic flow statistics in transport layer
  - Operation level: operation statistics in industrial control protocols
  - Content level: measurement statistics from field devices
- Data in different levels have quite different characteristics
- Fail to utilize all levels of network data in proper ways
  - Most existing solutions only focus on one or two levels of data
  - Most existing solutions usually fail to utilize various data characteristics to select proper anomaly detection method for different levels

Motivation

- Lack of further analysis of anomaly detected
  - The focus for most existing work is only turning data into knowledge by performing event detection on network traffic
  - Since the causes and consequences of the event are not identified, it is hard or impossible for the operator to quickly digest the event and react to it



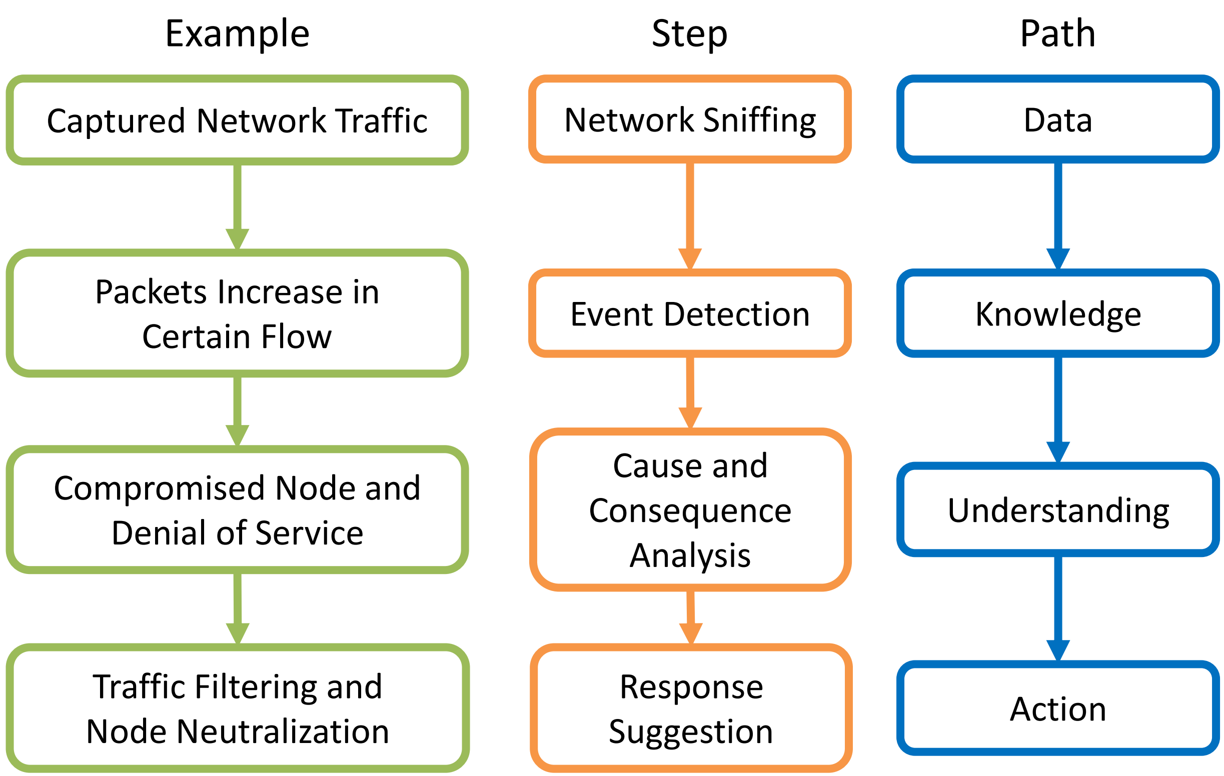
Our Approach

- Objective
  - An online, context-aware, intelligent framework for anomaly detection, cause and consequence analysis, and response suggestion for SCADA networks

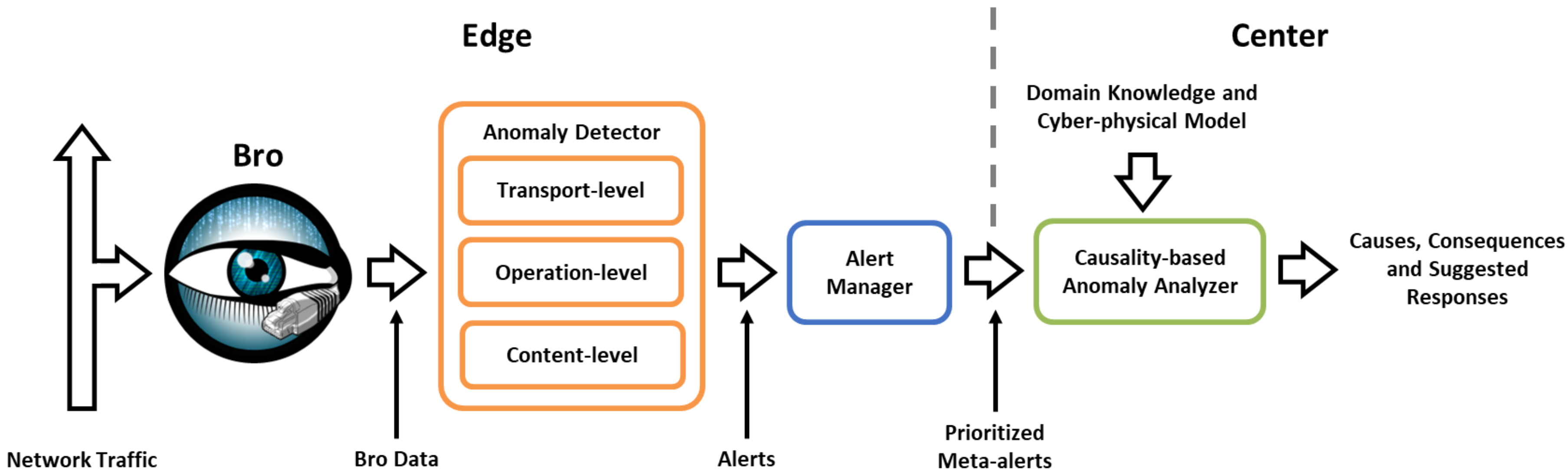
- Design decision
  - Build a multi-level anomaly and utilize proper anomaly detection methods to different levels of data
  - Incorporate the capability of not only detecting anomalies, but also analyzing causes and consequences of anomalies as well as suggesting feasible responses to our framework

Our Approach

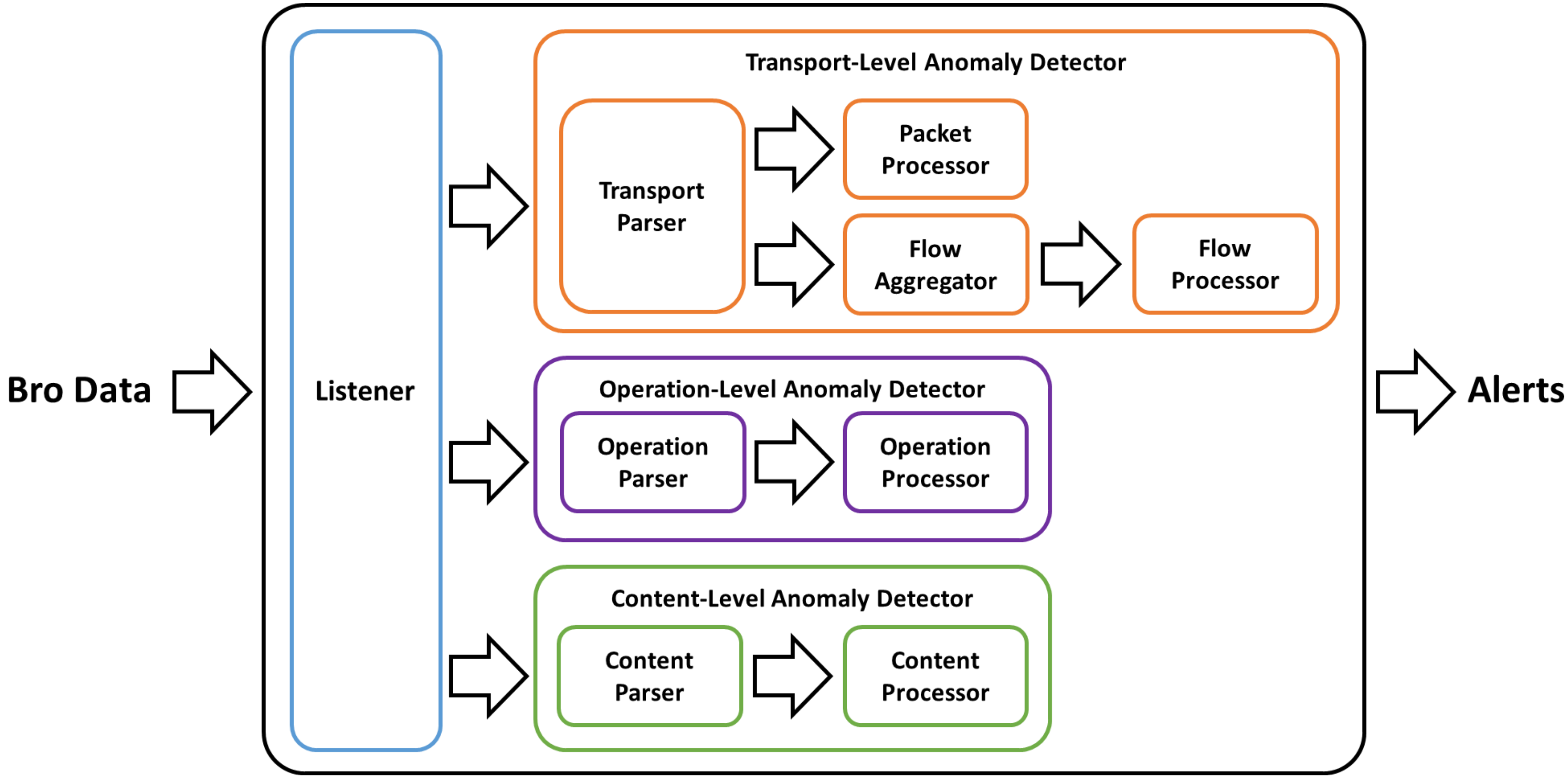
- DOS Attack example



Framework Architecture



Anomaly Detector



ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC
CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Anomaly Detector – Confidence Score of Alert

- Definition
  - Confidence that the corresponding alert is an anomaly.
- Calculation

$$Confidence\ Score = Model\ Accuracy\ \times Anomaly\ Score$$

Confidence Score: $\in [0, 1]$

Model Accuracy: How accurate is our model in describing normal behavior

- ↑ Use a modified sigmoid function of observed sample number to estimate

Anomaly Score: How far does the current value deviate from the normal value

- ↑ Different levels have different ways to calculate

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN CREDC CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Anomaly Detector – Transport Level

- Packet processor (runs every packet)
  - Index fields: originator, responder, transport protocol, port number
  - Data fields: interarrival time (IAT), packet size
  - Method: 1D-DenStream (utilizes a simplified 1D version of the clustering method DenStream[1])
- Flow processor (runs every period T_{flow})
  - Index fields: originator, responder, transport protocol, port number
  - Data fields: packet count
  - Method: mean and standard deviation (utilizes Chebyshev's Inequality to calculate anomaly score[2])

[1] Cao, F., Estert, M., Qian, W., & Zhou, A. (2006, April). Density-based clustering over an evolving data stream with noise. In *Proceedings of the 2006 SIAM international conference on data mining* (pp. 328-339). Society for Industrial and Applied Mathematics.
[2] Ren, W., Granda, S., Yardley, T., Lui, K. S., & Nahrstedt, K. (2016, November). OLAF: Operation-level traffic analyzer framework for Smart Grid. In Smart Grid Communications (SmartGridComm), 2016 IEEE International Conference on (pp. 551-556). IEEE.

Anomaly Detector – Transport Level

- Different methods are used for different data

Interarrival time (IAT), Packet size } Multimodal distribution → Clustering

Packet count → Unimodal distribution → μ, σ

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC
CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Anomaly Detector – Operation Level

- Operation processor
  - Objective: detect anomalies in operations of industrial control protocols (Modbus, DNP3)
  - Index fields: originator, responder, industrial control protocol, unit id, function
  - Data field: interarrival time (IAT)

Anomaly Type	Method
Invalid operation (invalid function code, wrong direction)	Check against rules
Abnormal operation (emerging/disappearing operation, abnormal IAT)	Use statistics: mean and standard deviation (IAT of the same operation is a unimodal distribution)

Anomaly Detector – Content Level

- Content processor
  - Objective: detect anomalies in measurement values which are included in responses to read requests
  - Index fields: holder, industrial control protocol, unit id, measurement type, measurement index
  - Data field: measurement value
  - Method: different methods for different measurement types

- DNP3 measurement type
  - Binary } most common
  - Analog } most common
  - Counter

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC
CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Anomaly Detector – Content Level

- Binary
  - Intuition: binary measurement usually has a normal value and an abnormal value
  - Method: count zeros and ones and try to identify the normal value
  - Anomaly Score (AS): 1 – Entropy(observed samples)

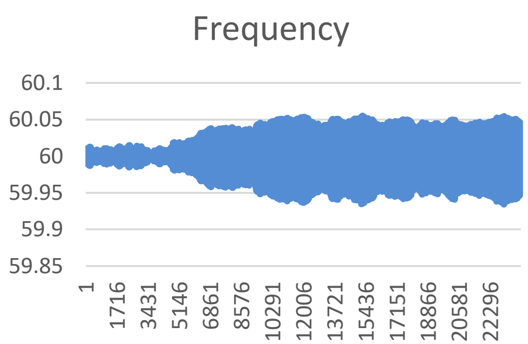
$$AS = \begin{cases} 1 & x = 0 \ or \ 1 \\ 1 + x\log_2 x + (1-x)\log_2(1-x) & 0 < x < 1 \end{cases}$$

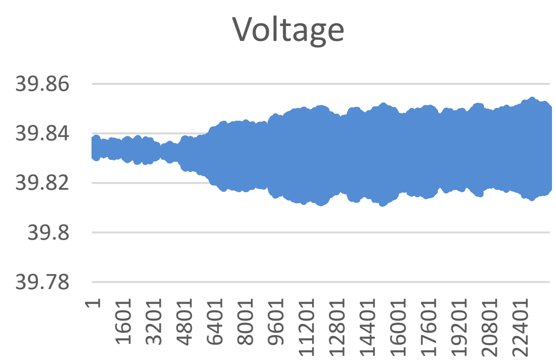
where $\mathrm{x} = \dfrac{number\ of\ ones\ observed}{number\ of\ samples\ observed}$

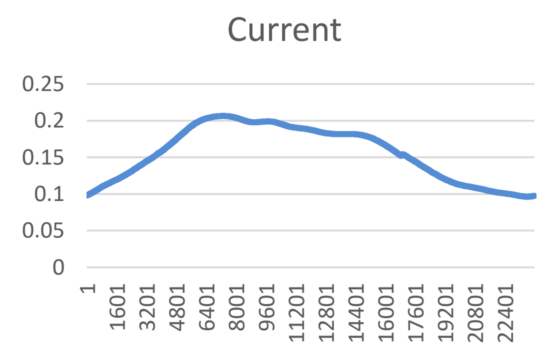
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC
CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Anomaly Detector – Content Level

- Analog
  - Most common analog measurements include frequency, voltage, current, power
  - They have quite different characteristics







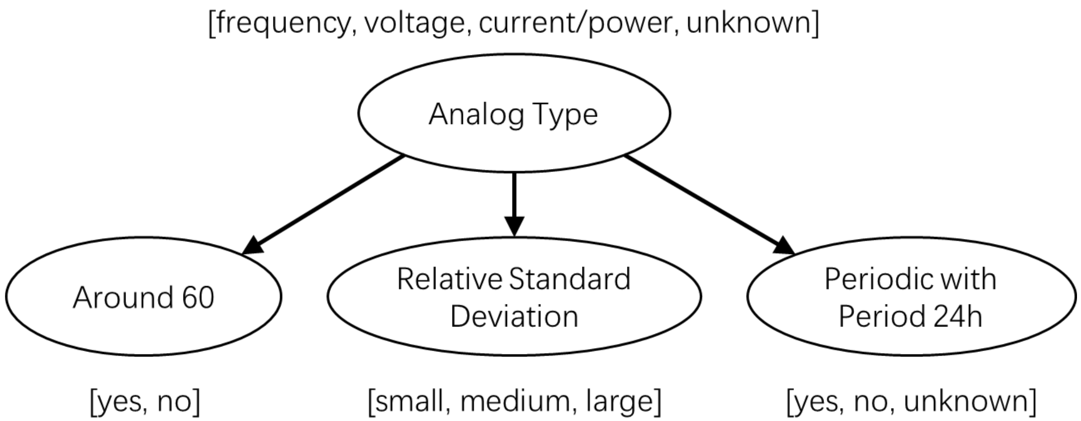
  - 2-step anomaly detection

1. Categorizes analog measurements into different analog types
2. Uses proper method for each type

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Anomaly Detector – Content Level

- Step 1: Bayesian-network-based analog type inference model



  – We denote y^k as the observation at k^{th} leaf node and x_i as the i^{th} analog type at the root node

$$P(x_i|y^1,y^2,y^3) = \alpha P(x_i)\prod_{k=1}^{3} P(y^k|x_i)$$

where $\alpha = \dfrac{1}{P(y^1,y^2,y^3)}$ and can be calculated using $\sum_i P(x_i|y^1,y^2,y^3) = 1$

Anomaly Detector – Content Level

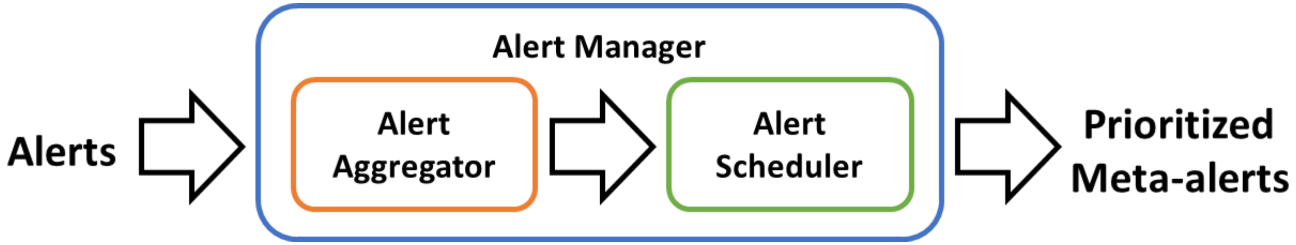
- Step 2: Different anomaly detection method for each analog type

Analog Type	Anomaly Detection Method
Frequency	Mean and standard deviation
Voltage	Mean and standard deviation
Current/Power	Time-slotted mean and standard deviation
Unknown	Mean, maximum, and minimum

Alert Manager

- Alert field
  - Index fields (same as index fields of the corresponding processor)
  - Alert type
  - Timestamp
  - Confidence score
  - Statistical fields (current value, mean, standard deviation, etc.)
  - Abnormal data (original parsed data of the corresponding level)

- Alert manager structure



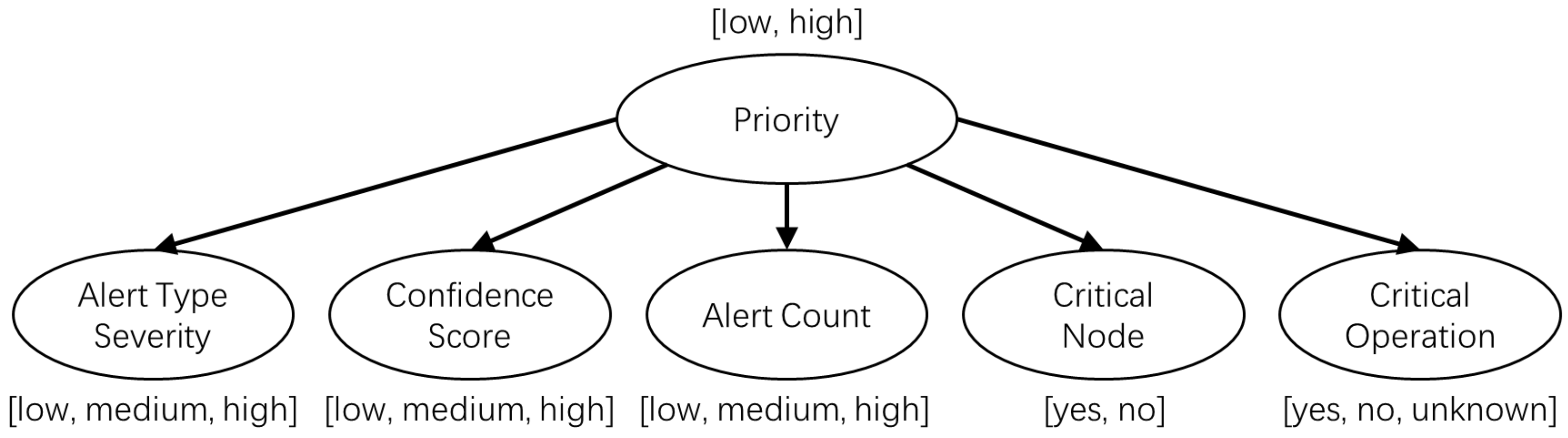
Alert Aggregator

- Objective
  - Aggregate alerts that have same type as well as index fields and have little difference in timestamp

- Meta-alert field
  - Index fields (shared by all of the aggregated alerts)
  - Alert type (shared by all of the aggregated alerts)
  - Timestamp (minimum, maximum)
  - Confidence score (maximum)
  - Count (number of aggregated alerts)
  - Statistical fields (statistical fields of the last alert aggregated)
  - Anomaly data (anomaly data of the last alert aggregated)

Alert Scheduler

- Objective
  - Calculate priority score for each meta-alert and decide when to report it to the control center
- Priority score
  - We denote y^k as the observation at k^{th} leaf node
  - Define $Priority\ Score = P(Priority = high \mid y^1, y^2, y^3, y^4, y^5)$



ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CREDC CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Alert Scheduler

- Meta-alert report frequency

	High-Priority Meta-alert	Low-Priority Meta-alert
Definition	$Priority\ Score \geq \theta$	$Priority\ Score < \theta$
Report when first created	Yes	No
Report frequency	T_1 if updated within T_1	$T_2(> T_1)$ if updated within T_2

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN CREDC CYBER RESILIENT ENERGY DELIVERY CONSORTIUM

Next Step

- Utilize alert correlation and attack plan recognition techniques to analyze the meta-alarms.
- Domain knowledge, causal relationships, and cyber-physical models of the system will be utilized to aid cause and consequence analysis of anomalies.