METAPHORTRESS

2021 NETL FE R&D Annual Project Review Meeting

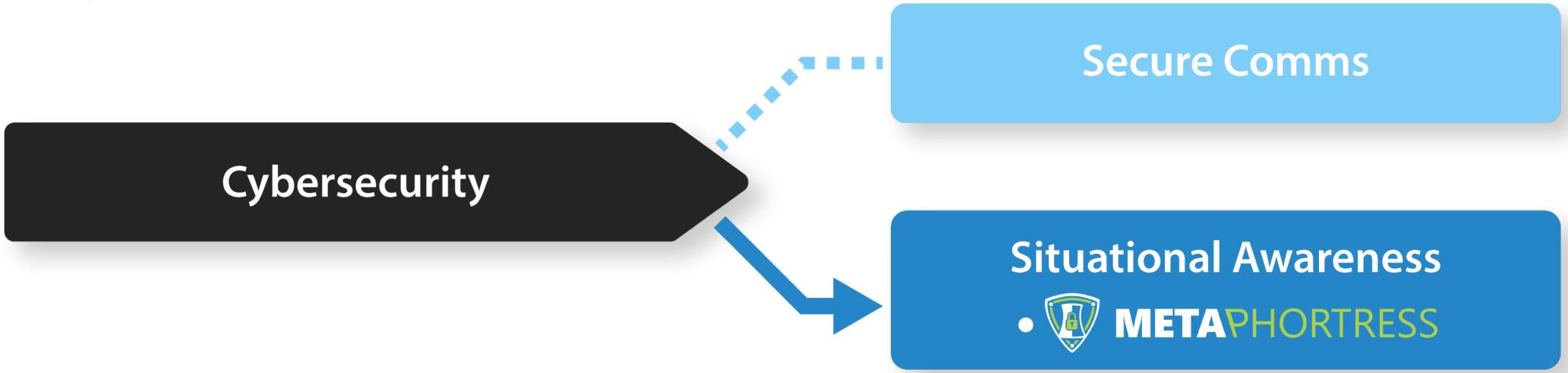
Artificial Intelligence

MetaPhortress Project Status

13 May 2021

Agenda

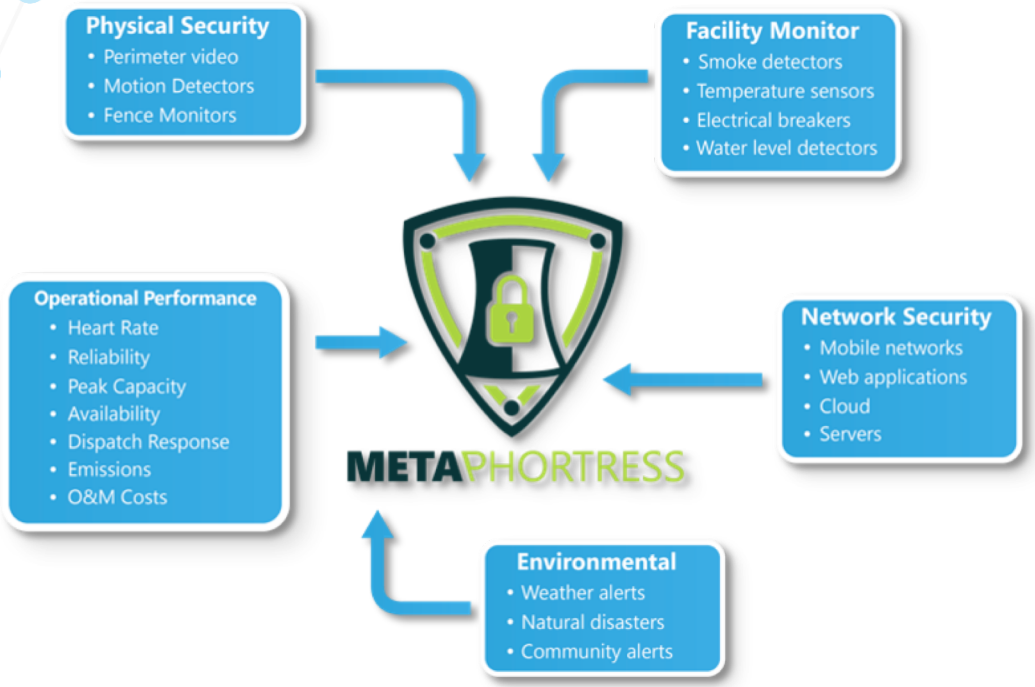
- Project Description & Objectives
  - System Concept and Features
  - Lessons Learned
- Project Update
  - Analytics Research
  - Situation Awareness Research
  - User Interface Design
- Conclusions



Project Description & Objectives

Project Description and Objectives

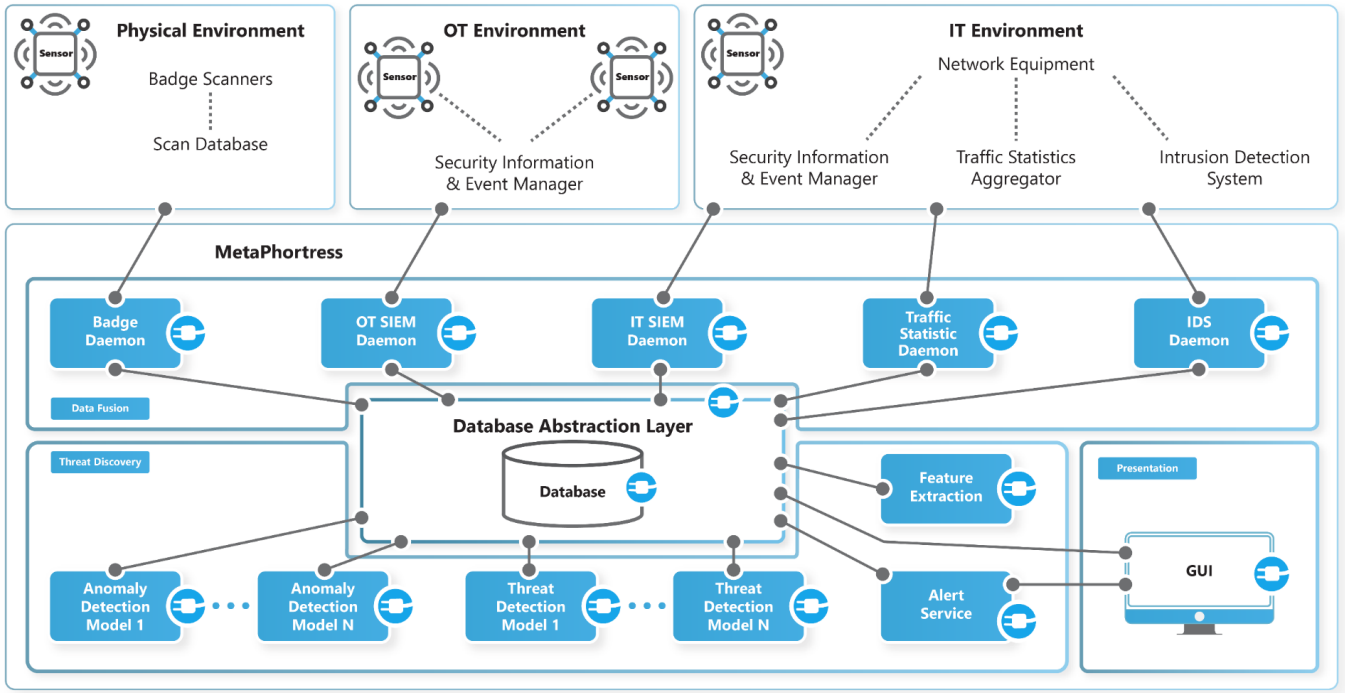
DOE Office of Fossil Energy 2018-2022 Strategic Vision, Objective 2.2:
Advance technologies to improve the efficiency, reliability, emissions, and performance of existing fossil-based power generation



- To avoid service interruptions, fossil fuel power plants need effective situation awareness to detect and mitigate cyber threats.
- MetaPhortress is an automated cyber situation awareness tool that will enhance the resilience, safety, and reliability of these facilities.
- This question drives us: How do we provide accurate, timely, and actionable cyber situation awareness and threat detection to power plants?

System Concept

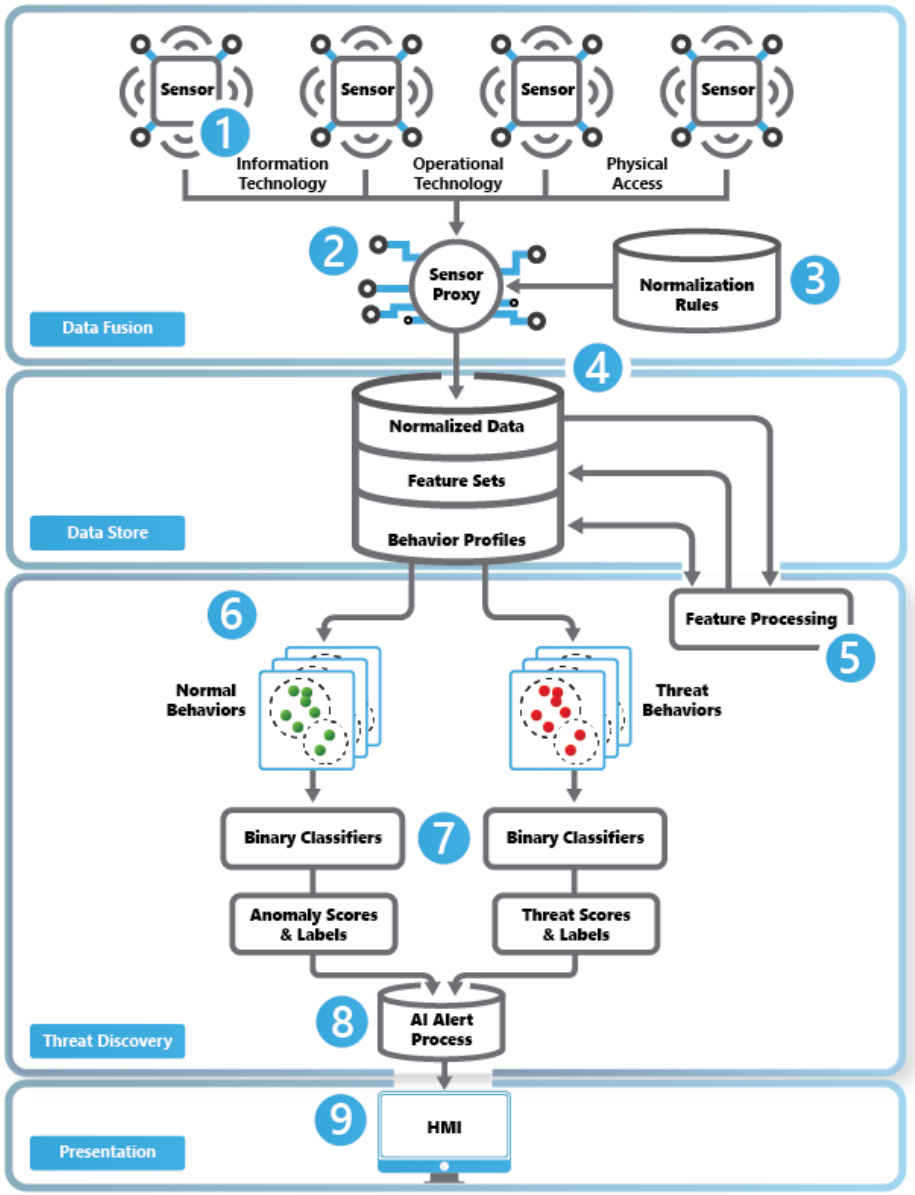
- MetaPhortress adapts our patented cyber feature-extraction and behavior analysis platform to provide comprehensive, simultaneous coverage of power plant operational technology (OT)/ICS, information technology networks (IT), and physical access control systems (PACS).
- Performs data fusion upon networked sensor outputs in all three domains to characterize nominal operational modes
- Uses machine learning and data analytics techniques extract features, detect deviations from nominal modes, determine which anomalous conditions correspond to malicious behavior, and alert system operators to potential cyber incidents.



System Features

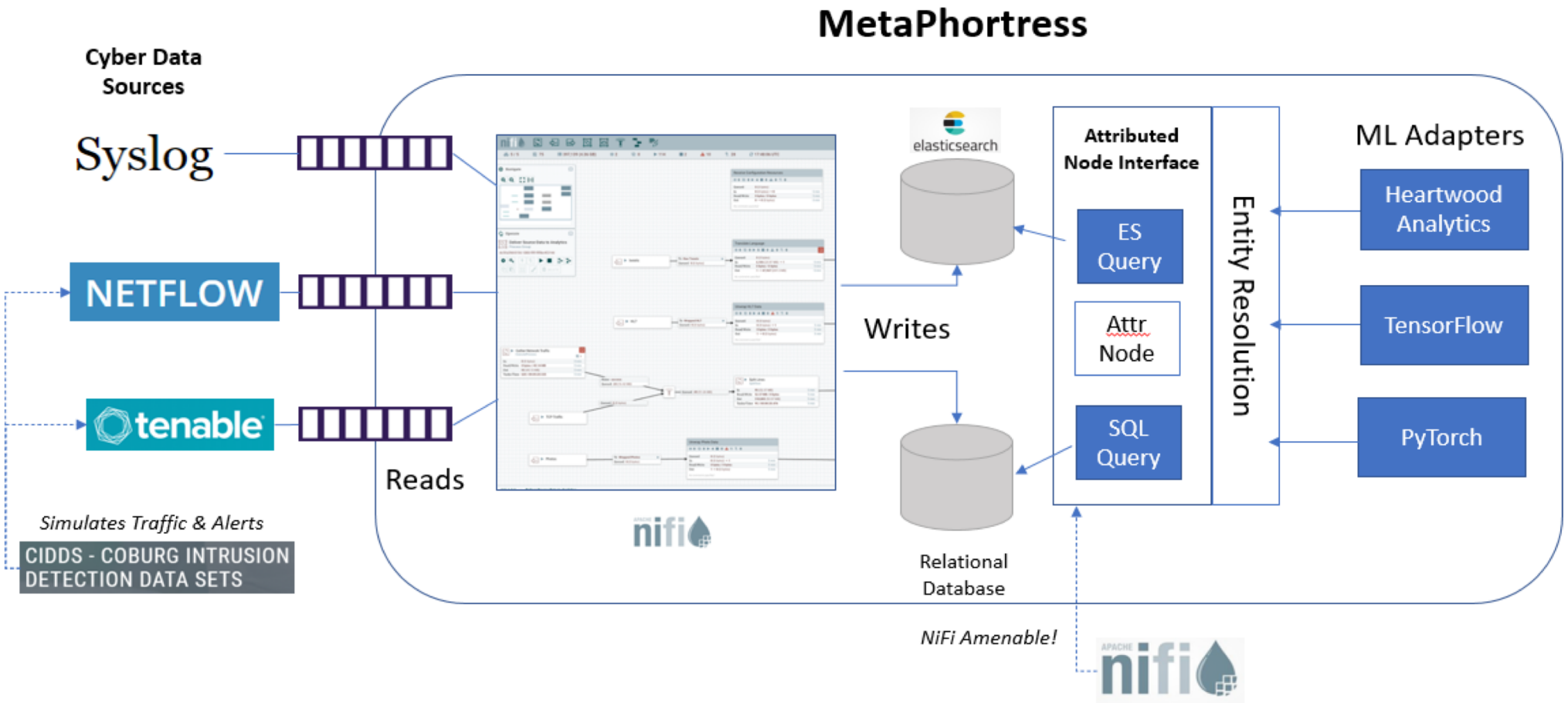
- Converged, simultaneous sensor data analysis of OT, IT, and PACS to discover cyber threats and resolve them against the time and system domains
- Aggregated behavior analysis to discover malicious entities that attempt multiple vectors across power plant attack surfaces
- Temporally aggregated analysis to detect attacks that unfold over varied timescales
- Rapid, clear, actionable presentation of threat alerts to power plant operators
- Improved defense of critical energy infrastructure to known and emerging cyber threats

- Collaboration Partners
  - **CUBRC** – data fusion and machine learning expertise
  - **TDi Technologies** – power generation domain knowledge, software integration requirements, energy sector software vendor



Technology Stack

NIST guidance for cyber protection of power generation facilities recommends converged threat analysis of the OT/ICS, IT, and PACS domains. Individual, siloed analysis of those data areas is common; MetaPhortress, instead, automates this combined analysis with data fusion over all three areas.



What We've Learned

- The MetaPhortress development team continues to meet with energy sector stakeholders in industry who provide valuable insights that guide needs assessment, requirements analysis, and system design.
- MetaPhortress team efforts have:
  - Researched and developed detailed requirements
  - Developed detailed design elements
  - Developed prototype MetaPhortress software
  - Tested MetaPhortress prototype
  - Conducted integration testing periods
- By executing these efforts, and working with our stakeholders, we realize that what we initially saw as an analytics challenge is actually also a human factors challenge – how do we convert machine learning outputs into clear and effective situation awareness cues that will help plant operators act on potential cyber threats?

Project Update

MetaPhortress Analytics

- Cross-domain analysis
  - Provide comprehensive, simultaneous coverage of power plant operational OT/ICS, IT, and PACS through:
    - Cyber feature-extraction and behavior analysis
      - Automated roll-up of feature sets into behaviors, pulling signal from noise
      - Models trained on behaviors
    - Feature sets from all domains in common time windows will be analyzed for signature and anomaly-based threats
      - Moving time windows with varying amounts of data from the domains within window
      - Within-window aggregation (e.g. simple average, exponential-weight moving average, etc.)
      - Imputation/extrapolation of data where none present

Techniques	Implementation
• Extract features • Detect deviations from nominal modes • Determine which anomalous conditions correspond to malicious behavior • Alert system operators to potential cyber incidents	• Combine OT sensors, generalized & rule-based, to apply to multiple plants • Training of machine learning models for IT and OT on various time scales using labeled IT and OT datasets, then running trained model on incoming data to produce classification output

Datasets

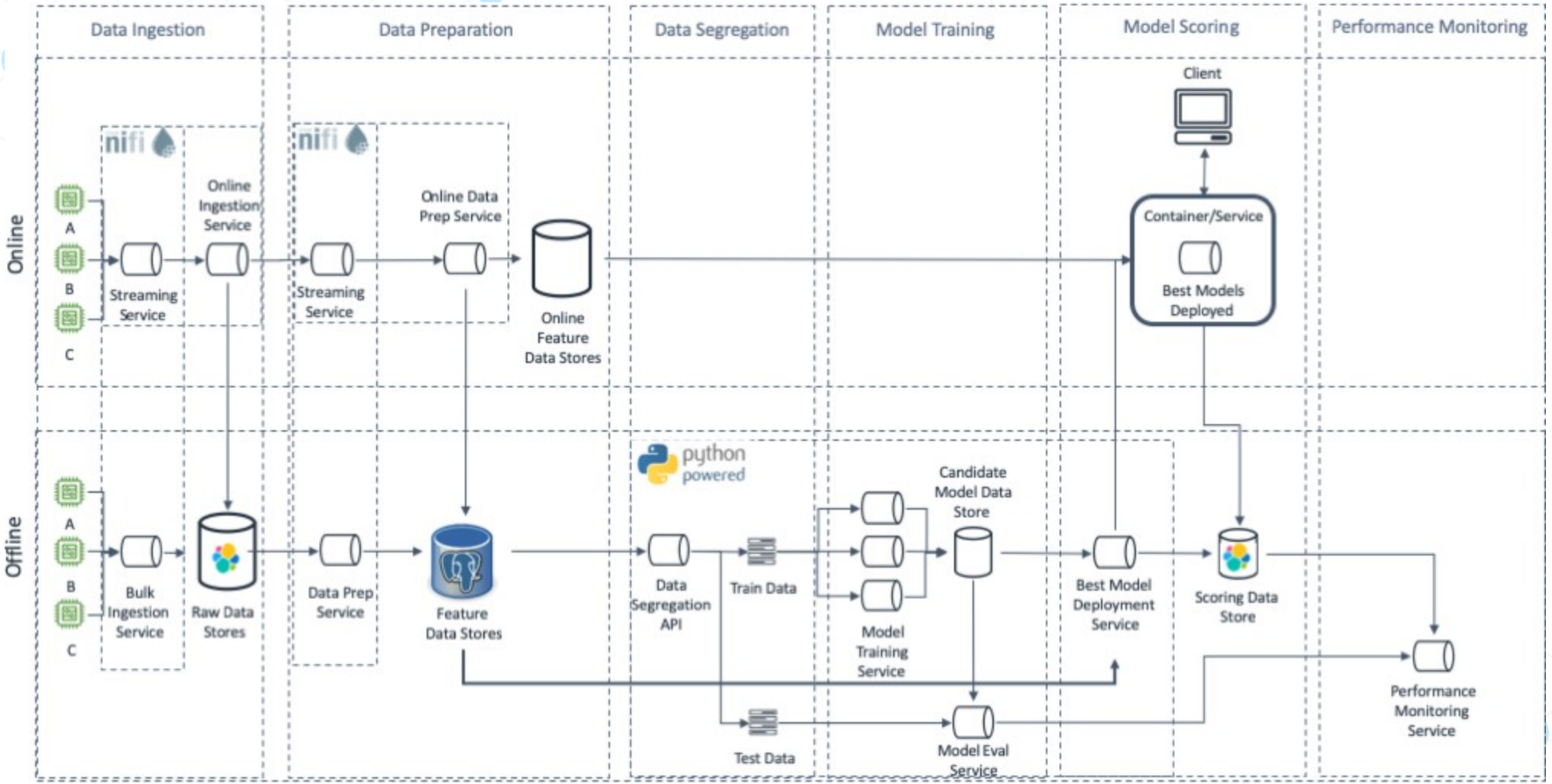
- Obtained uncorrelated OT and IT data sets for training and testing
  - Few public datasets of long time duration exist
  - Align time stamps of data sets to create a notional state of the power plant
  - Working to obtain correlated datasets
  - Datasets used exhibit divergence in OT and IT during tests

- ORNL power system attack dataset
  - CSV format; labeled
  - Four intelligent electronic devices (IED)
  - Features include: Phase current and voltage phase angle/magnitude, Pos-Neg-Zero current and voltage angle/magnitude, frequency and appearance impedance for relays

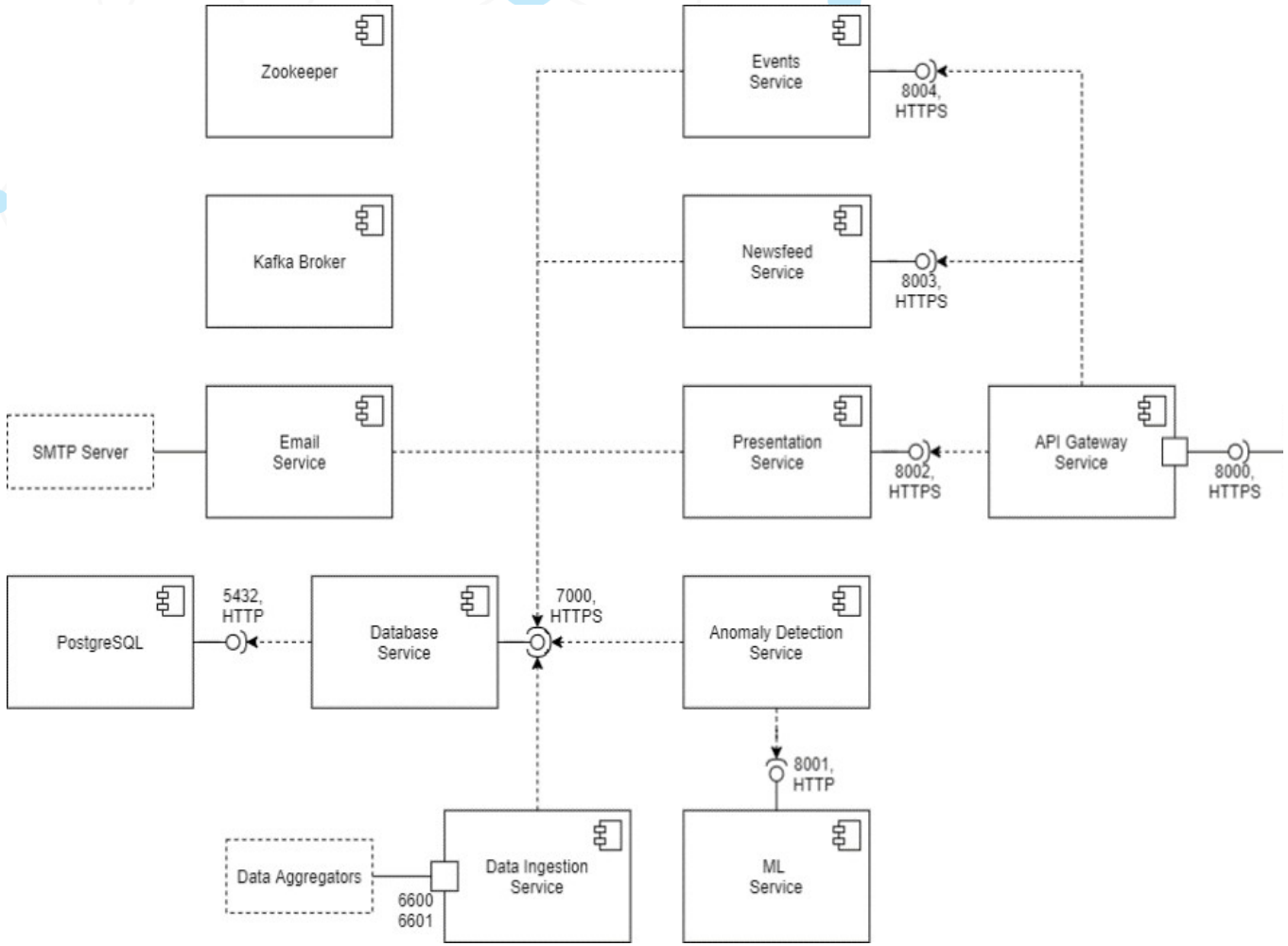
- U. New Brunswick CICIDS-2017 dataset
  - PCAP formatted network data; labels added from metadata
  - Features include: total packets, protocol, time length, src IP, src port, dst IP, dst port, total bytes, avg packet size, smallest packet, and largest packet
  - Injected Modbus PCAP from CSET'16 into dataset

- Evaluating BATADAL (Battle of the Attack Detection Algorithms)
  - Water system ICS

Data Processing Pipeline



MetaPhortress Analytics



- Developed API that serves multiple predictions from the OT and IT domains
  - Specific endpoints per domain, per grouping, and per time domain as requested
  - Predictions are returned in a response body with two scores, raw and human readable
- Developed data model for the IT and OT domains, validating on sensor ID
  - Incoming data is rejected if the sensor ID does not match the IDs in the data model

MetaPhortress Analytics

- Developed anomaly detection ML models for the OT and IT domain
  - IT: hourly anomaly detection capability
    - Model covers domain as a whole (network traffic flow)
  - OT: hourly and daily detection capability
    - Built and deployed anomaly machine learning model for each of the different sensor groupings (set of OT relays) and domain as a whole
  - Models evaluated
    - Statistical Process Controls, Local Outlier Factor, One-Class SVM, Autoencoders, Elliptic Envelope, Local Outlier Factor, and Isolation Forests
    - Isolation Forests emerged as the preferred model due to its explainable properties, ease of implementation, and robustness
- Models provide predictions in raw and a human readable output
  - Raw score between 0 and 1 from the isolation forest algorithm
    - Path length averaged over a forest of random trees
    - Shorter paths and scores closer to 1 are more anomalous
    - Scores smaller than 0.5 are regarded as normal instances
  - Human readable output bins the raw score into three bins; not anomalous, caution, and anomalous

Model Evaluation

Dataset	Algorithm									
	SPC (%)		IF (%)		OcSVM (%)		AE (%)		LOF (%)	
	TP	TN	TP	TN	TP	TN	TP	TN	TP	TN
ORNL	58.31	85.93	78.42	74.68	88.66	86.03	66.10	63.07	84.68	86.75
BATADAL	42.01	98.77	75.80	78.55	85.84	82.89	68.95	69.50	83.56	84.52
Modbus	54.84	93.71	100.00	94.77	80.65	95.91	81.48	88.55	93.55	94.88
CIC	60.24	87.93	95.77	93.92	86.12	82.55	72.01	70.19	38.39	96.07

- Evaluated models on four datasets
- Data was shuffled and models trained on 75% of clean data
- Remaining 25% data used to determine the algorithms accuracy for non-outliers, or True Negatives (TN)
- The approach is then used to determine its accuracy in detecting outliers, or True Positives (TP), by using the labeled outlier data provided by the datasets
- The exception to this methodology occurs for the CICIDS dataset when using the one-class SVM.
  - Training time increases exponentially with the number of samples
  - Down selected training and test sets to 10% of available data

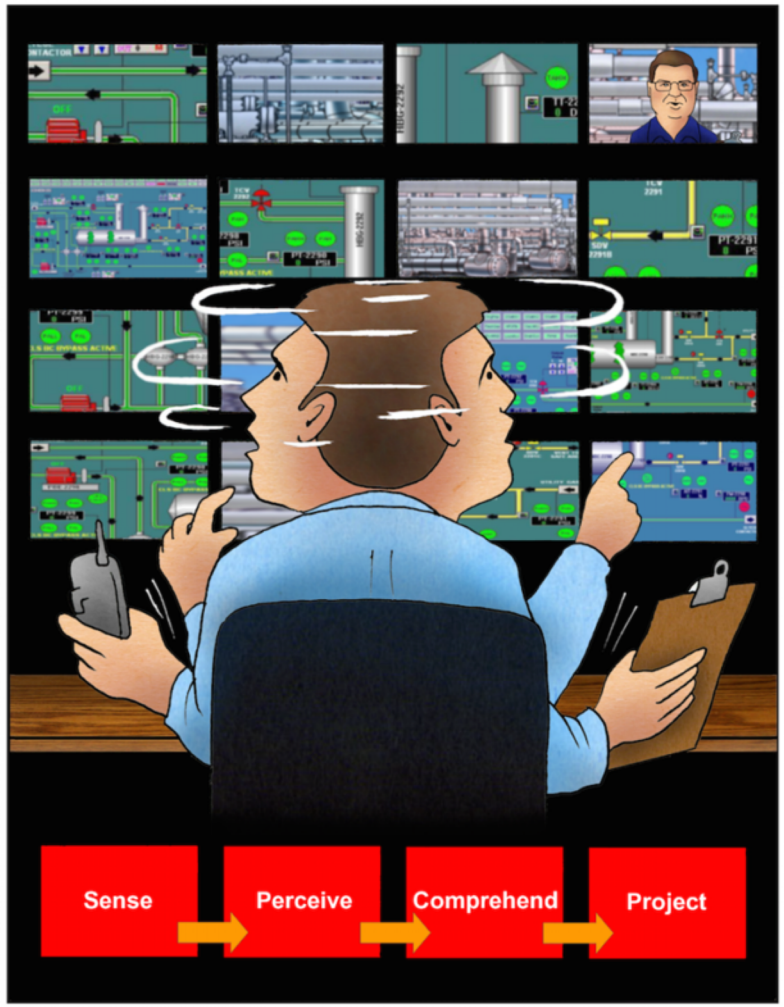
Situation Awareness (SA)

SA: Knowing what's going on, so you can make good decisions

- Experimental psychology construct, theory, and model
- Describes how different factors… affect a human's ability to acquire and interpret information for effective decision making (Endsley, 1995)
- SA Model is composed of three levels (Endsley, 1995, 2000):
  - SA_1: Perception of elements in the environment
  - SA_2: Comprehension of the current situation
  - SA_3: Projection of future status

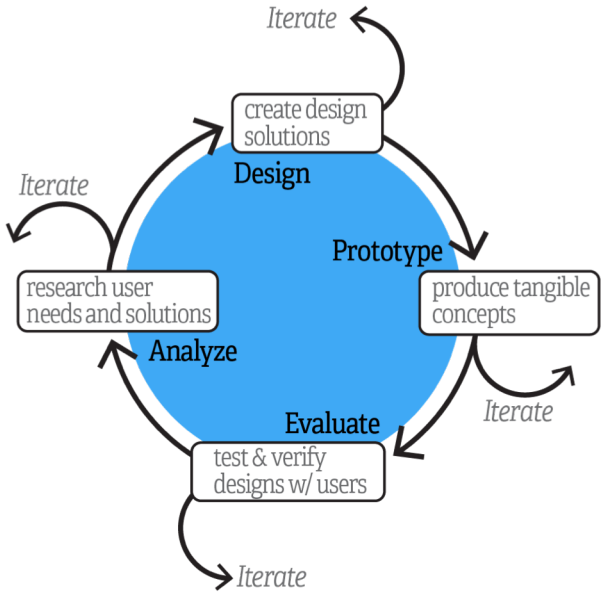
M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), pp 32-64, 1995.
M. R. Endsley, "Theoretical underpinnings of situation awareness: A critical review," In Situation Awareness Analysis and Measurement, M. R. Endsley and D. Garland, Eds., Mahwah: Lawrence Erlbaum, 2000, pp. 3-32.

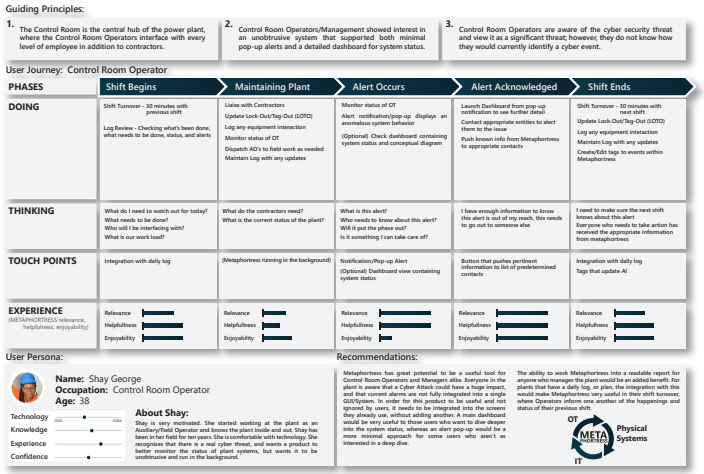
Situation Awareness and Sensemaking

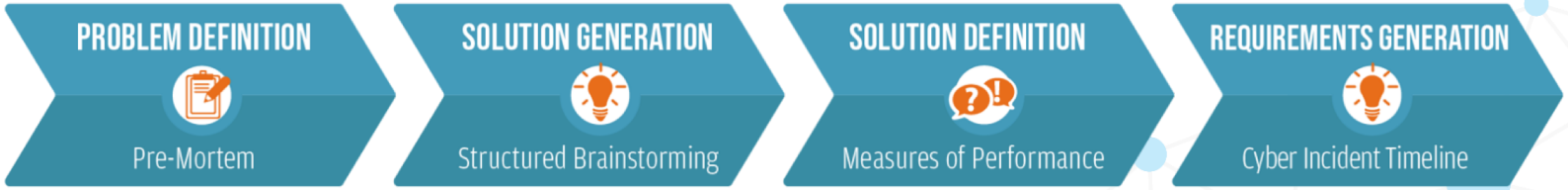


- Sensemaking is a part of situation awareness (SA).
- SA is “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future“ (Endsley, 1995).
- Sensemaking is both retrospective and prospective and is a process (rather than a state).

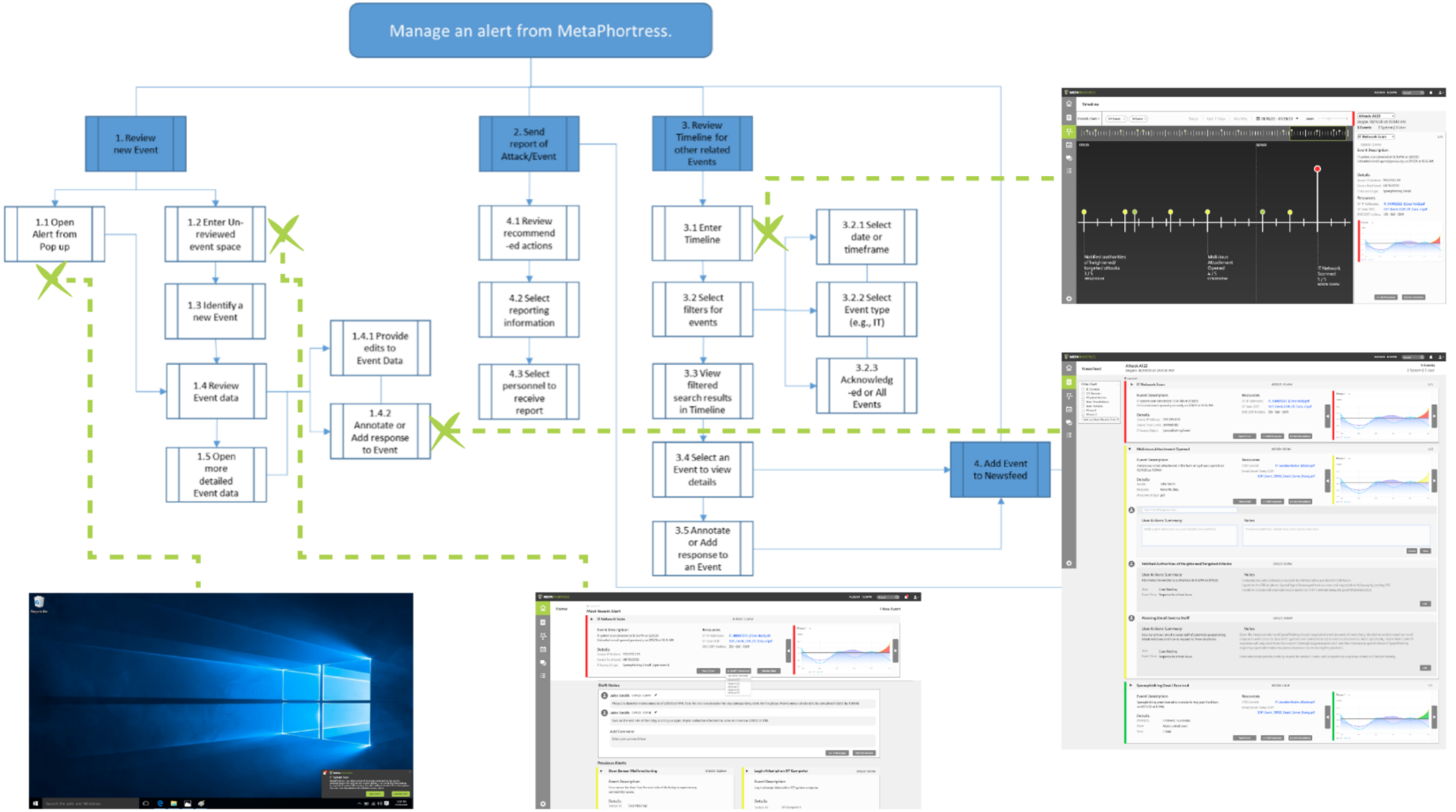
User Interaction – Feedback Cycle



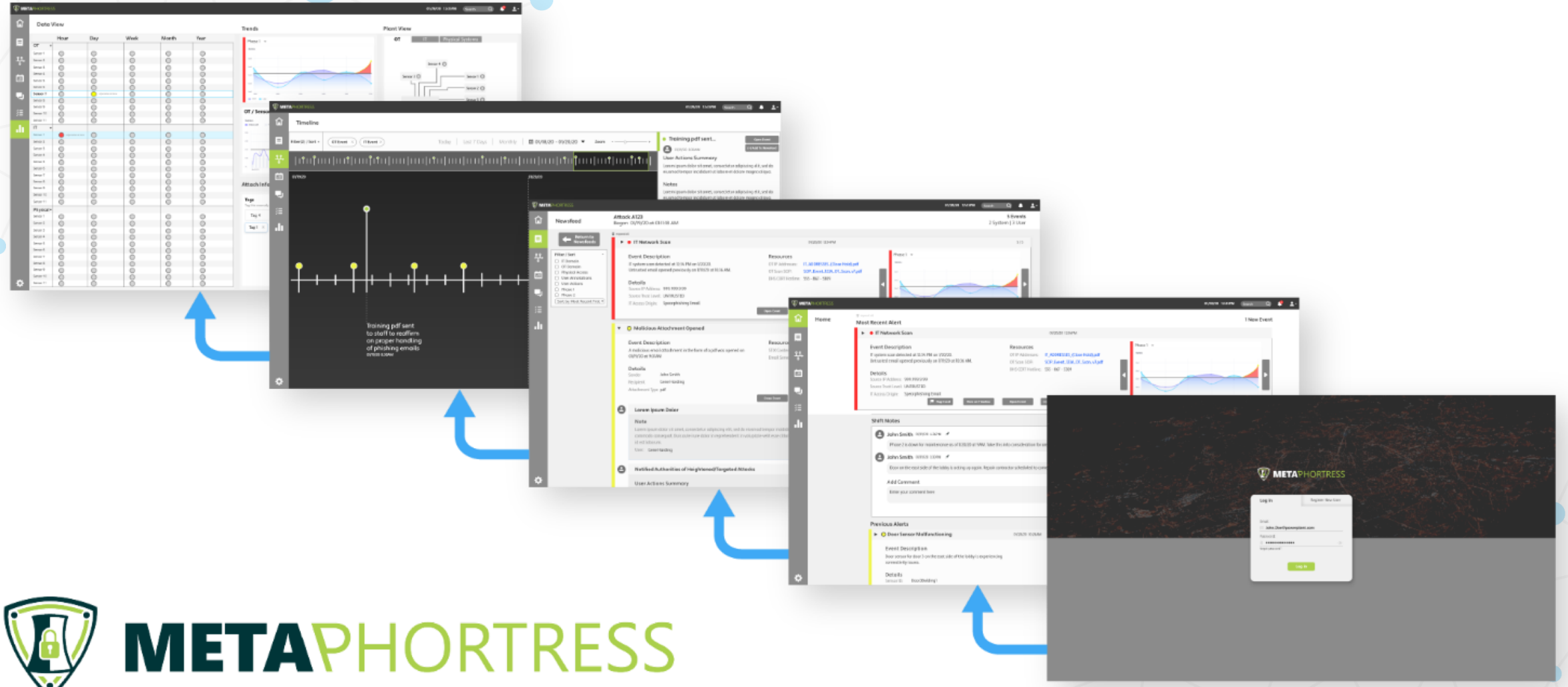




Mock-ups



Continued Refinement of Designs



METAPHORTRESS

Conclusions

Conclusions

- MetaPhortress will increase SA and cybersecurity at power generation plants by:
  - Fusing information from classically disparate domains (IT, OT, PACS)
  - Using Machine Learning to detect potential cyber threats
  - Provide operators with an intuitive interface that encourages sensemaking of voluminous and highly uncertain data
- Next steps:
  - Advance frontend with iterative user research and testing
  - Advance backend with iterative testing in representative power generation environments
    - Adding PACS domain processing
    - Automate testing of new algorithms
      - Automated optimization approaches to test different hyper parameter combinations
      - Identify the optimum match of ML model to network traffic type and environment
  - Develop CONOPS and system architecture to guide transition to market

Acknowledgment: "This material is based upon work supported by the Department of Energy Award Number DE-SC0018729."