Secure Data Logging and Processing with Blockchain and Machine Learning

Project Team

Secure Data Logging and Processing with Blockchain and Machine Learning

Leonel Lagos, PhD, PMP® (Principal Investigator)
Himanshu Upadhyay, PhD, PMP® (Co - Principal Investigator)
Team Members: Santosh Joshi , Pranav Gangawani , Adrian Muino
Applied Research Center
Florida International University (FIU)

Wenbing Zhao, PhD (Co - Principal Investigator)
Cleveland State University (CSU)

Project Description

- Secure Data Logging and Processing with Blockchain and Machine Learning research is focused on the development of platform to securely log and process sensor data in fossil power plant

- The platform integrates two emerging technologies -blockchain and machine learning, and incorporates several innovative mechanisms to ensure the integrity, reliability, and resiliency of power systems

- The goal is to protect the power plant from various cyberattacks such as false data injection and denial of service attacks using these technologies

Project Objectives

Various objectives of the research are as follows:

Objective 1: Secure authentication and identity verification of sensor nodes, actuators, and other equipment within a network

Objective 2: Develop a set of mechanisms that ensure only data sent by legitimate sensors are accepted and stored in the data repository

Objective 3: Develop data aggregation methodologies using machine learning / deep learning algorithms to minimize the noise / faulty data

Objective 4: Implement the blockchain technology to provide data security using secured IOTA framework & nodes

Project Tasks

Task 1 - Secure Authentication and Identity Verification of Sensor Nodes

Task 2 - Data Aggregator Machine Learning Platform

Task 3 - Secure Logging with Blockchain

Project Task and Subtasks

Task 1 - Secure Authentication and Identity Verification of Sensor Nodes

A critical foundation for secure process control and secure sensor data logging is to collect data from the right sensors and send commands to the correct actuators. This requires proper authentication and identification of sensor nodes, actuators, and other equipment within a network.

This tasks involves following three subtasks:

- Perform threat analysis
- Design authentication and identification mechanisms
- Design security key management mechanism

Project Subtasks

Subtask 1.1-Threat Analysis: Perform a thorough threat analysis via a comprehensive literature review. This will provide the foundation for the entire project. All other mechanisms and algorithms developed in this project will be focused on how to defend against these potential attacks

Subtask 1.2-Design Authentication and Identification Mechanism: Design and implement authentication and identification mechanisms for sensors and actuators, so that no attackers could impersonate a senor in the power plant

Subtask 1.3-Design Security Key Management Mechanism: The mechanism is based on a scheme proposed as part of a secure email system, where two nodes exchange their keys via physical contact. This scheme is extended by using a mobile device as a relay to facilitate the exchange of keys of nodes that are not directly connected to each other or are not mobile

Project Task and Subtasks

Task 2.0 - Data Aggregator Machine Learning Platform

Design and implementation of Data Aggregator Machine Learning platform for secure data logging and processing. This requires proper authentication and identification of sensor nodes, actuators, and other equipment within a network

This tasks involves following three subtasks:

- Traditional Model Development
- Deep Learning Model Development
- Machine Learning Model Deployment

Project Subtasks

Subtask 2.1 Traditional Model Development: Design and implement One-class SVM and K-means clustering algorithms for Anomaly detection. Store the models and predictions in the AI & Big Data Hub server for future processing with IOTA framework

Subtask 2.2 Deep Learning Model Development: Design and implement Deep Learning algorithms (LSTM, Auto Encoders and GAN) for Anomaly detection with Sensor data collected from sensor network. Store the models and predictions in the AI & Big Data Hub server for future processing with IOTA framework

Subtask 2.3 Machine Learning Model Deployment: Machine learning model will be deployed on edge gateway by using edge functions with acquired data. Anomaly detection will be performed with real time data by using deployed model on the edge and AI & Big data Hub server

Project Task and Subtasks

Task 3.0 - Secure Logging with Blockchain

The Blockchain technology will be used to securely log summative data as the technology ensures that the data placed on the Blockchain is immutable. Flat structure of the Blockchain facilitates easy inspection and validation by administrators and regulators

This tasks involves following three subtasks:

- Creating a Strong Linkage between Two Levels of Logging
- Energy-Aware Blockchain Solution (Hashgraph-based Blockchain)
- Visualization of Logged Data

Project Subtasks

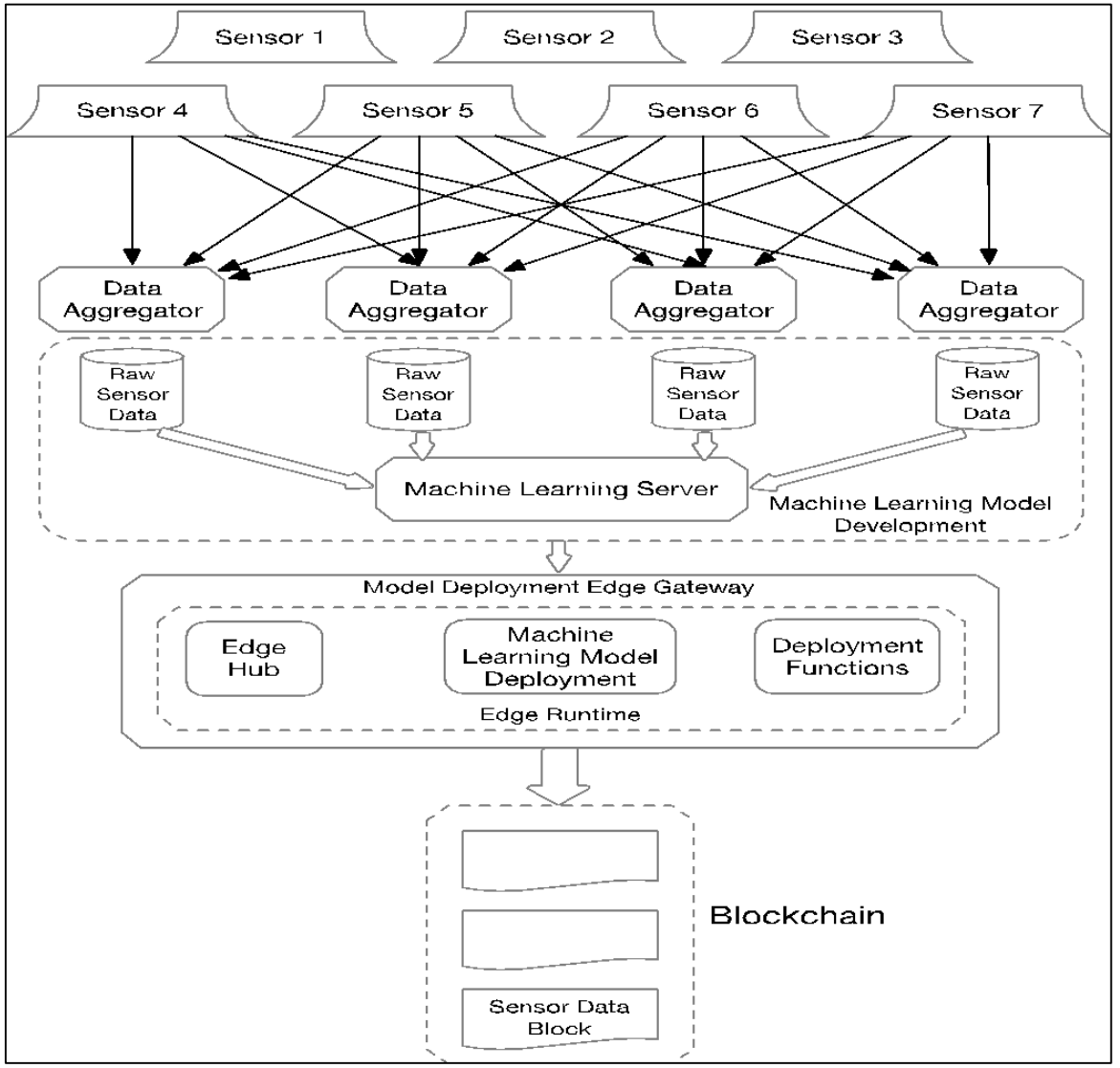
Subtask 3.1: Creating a Strong Linkage between Two Levels of Logging: Develop a mechanism to enable hierarchical logging where Blockchain is used to store summative sensor data, and raw data are stored locally at replicated data stores

Subtask 3.2: Energy-Aware Blockchain Solution: Study and try-out several candidate Blockchain solutions, select the most appropriate one for this project, and make the necessary modifications so that sensor data can be placed in the blockchain

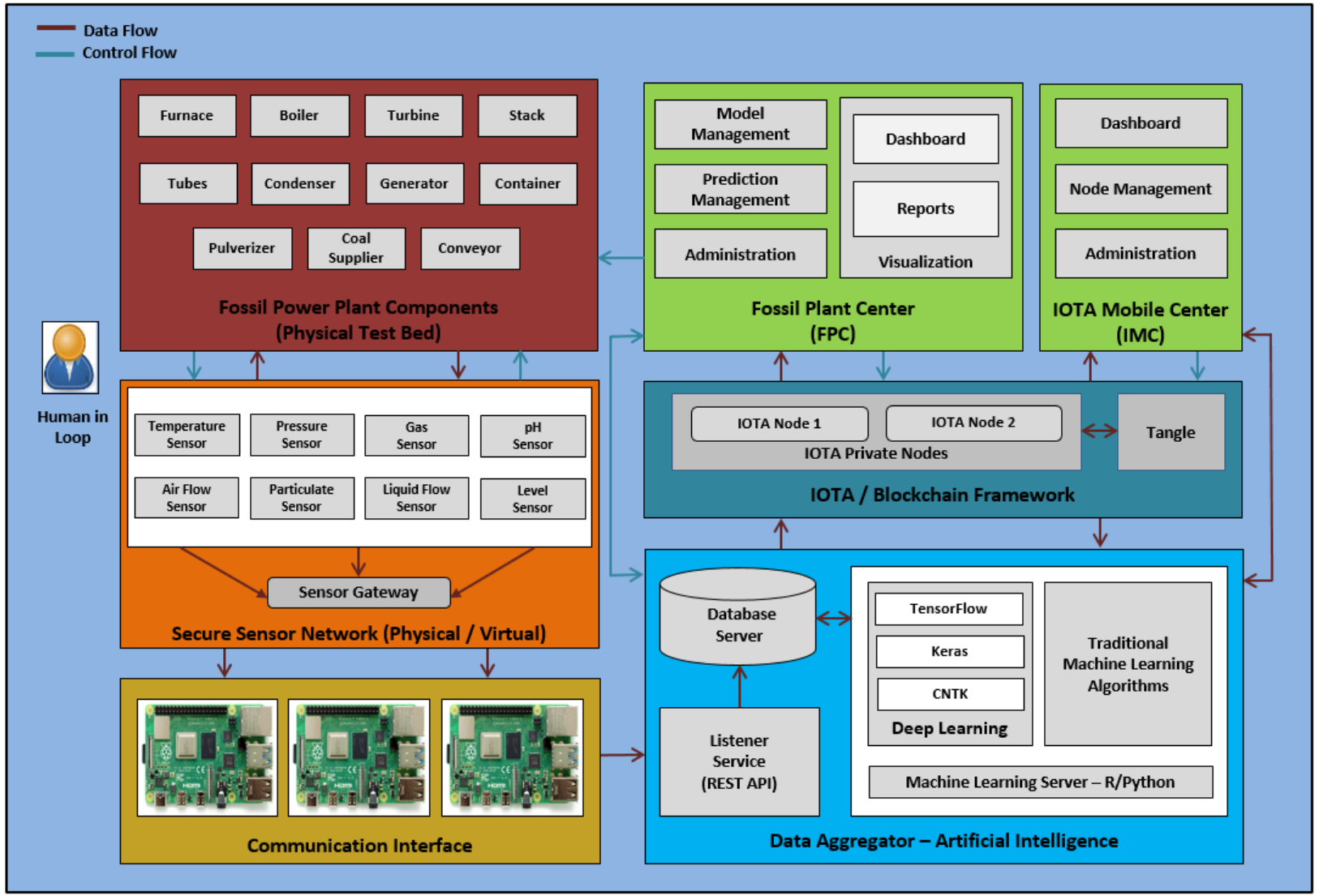
Subtask 3.3: Visualization of Logged Data: Summative data placed on the blockchain should be easily visualized to understand the current situation and the potentially past incidents. This will involve data retrieval from the blockchain, building visual data structures and dashboard for analytics results

Proposed System Architecture

- Sensors on the Test Bed
- Data Aggregator
- Machine Learning (ML) Server
- ML Model Development
- ML Model Deployment On Edge Gateway
- Blockchain



Secure Data Logging & Analysis System



System Platforms

Secure Data Logging & Analysis System consists of following platforms

- **Fossil Power Plant Components** (Physical Test Bed) – Mock-up for the Fossil Power Plant
- **Secured Sensor Network** – Physical & Virtual Sensors
- **Communication Interface** (Raspberry Pi)
- **Data Aggregator** (SQL Server & Machine Learning Server)
- **IOTA / Blockchain Framework**
- **Fossil Plant Center (FPC)** – Web based application to manage the Secure Data Logging & Analysis System
- **IOTA Mobile Center (IMC)** – Mobile Application to Manage IOTA / Blockchain Nodes and Data

Fossil Power Plant Components – Physical Test Bed

Fossil Power Plant components to develop mock-up physical test bed are listed below:

- **Furnace**
- **Boiler**
- **Turbine**
- **Stack**
- **Tubes**
- **Conveyor**
- **Condenser**
- **Generator**
- **Transformer**
- **Pulverizer**
- **Coal Supplier**
- **Container**

Fossil Power Plant Sensors – Secure Sensor Network (Physical / Virtual)

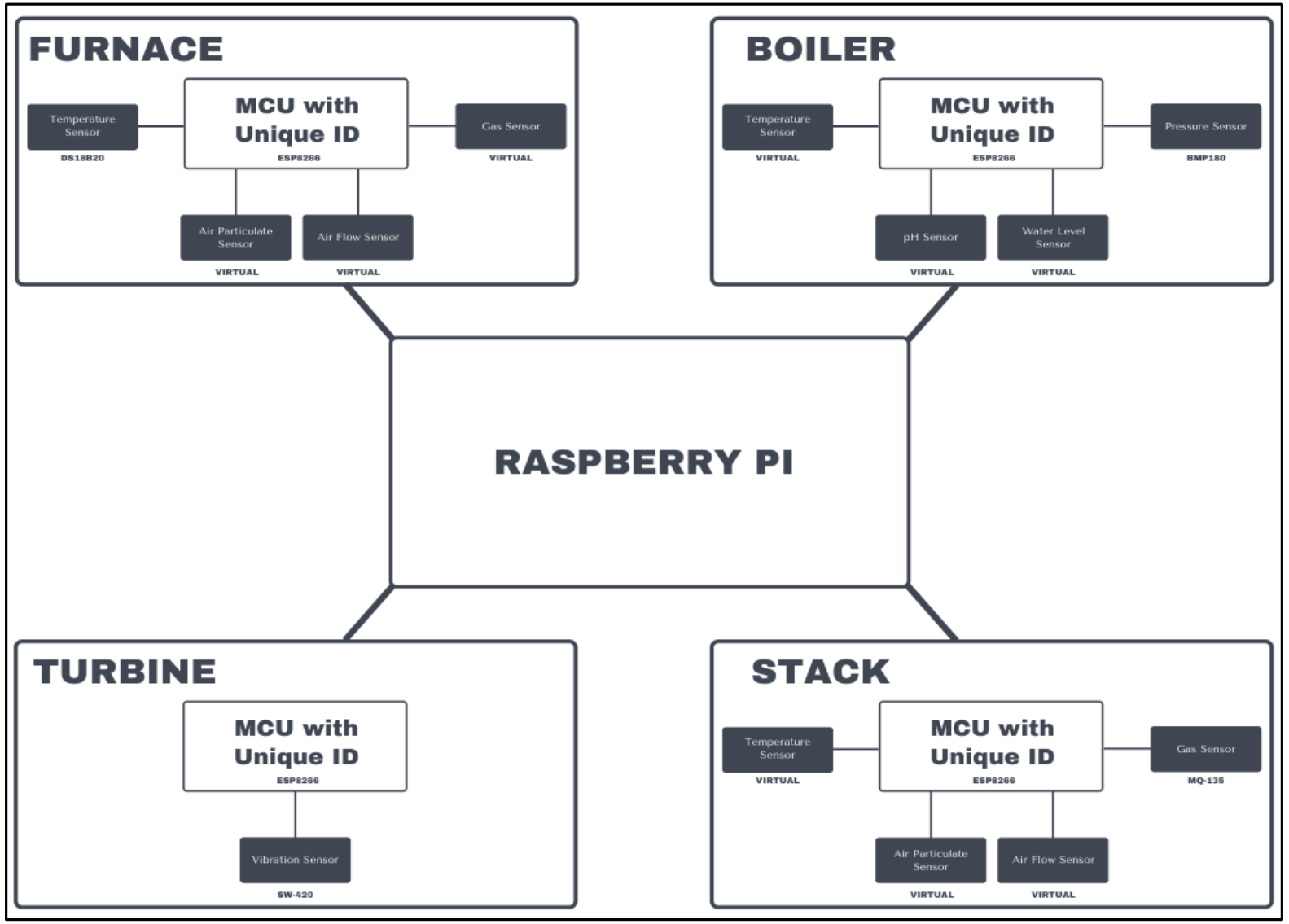
Fossil Power Plant sensors of Secure Sensor Network (Physical / Virtual Sensors) are listed below:

- **Temperature Sensor**
- **Pressure Sensor**
- **Gas Sensor**
- **pH Sensor**
- **Air Flow Sensor**
- **Particulate Sensor**
- **Liquid Flow Sensor**
- **Level Sensor**

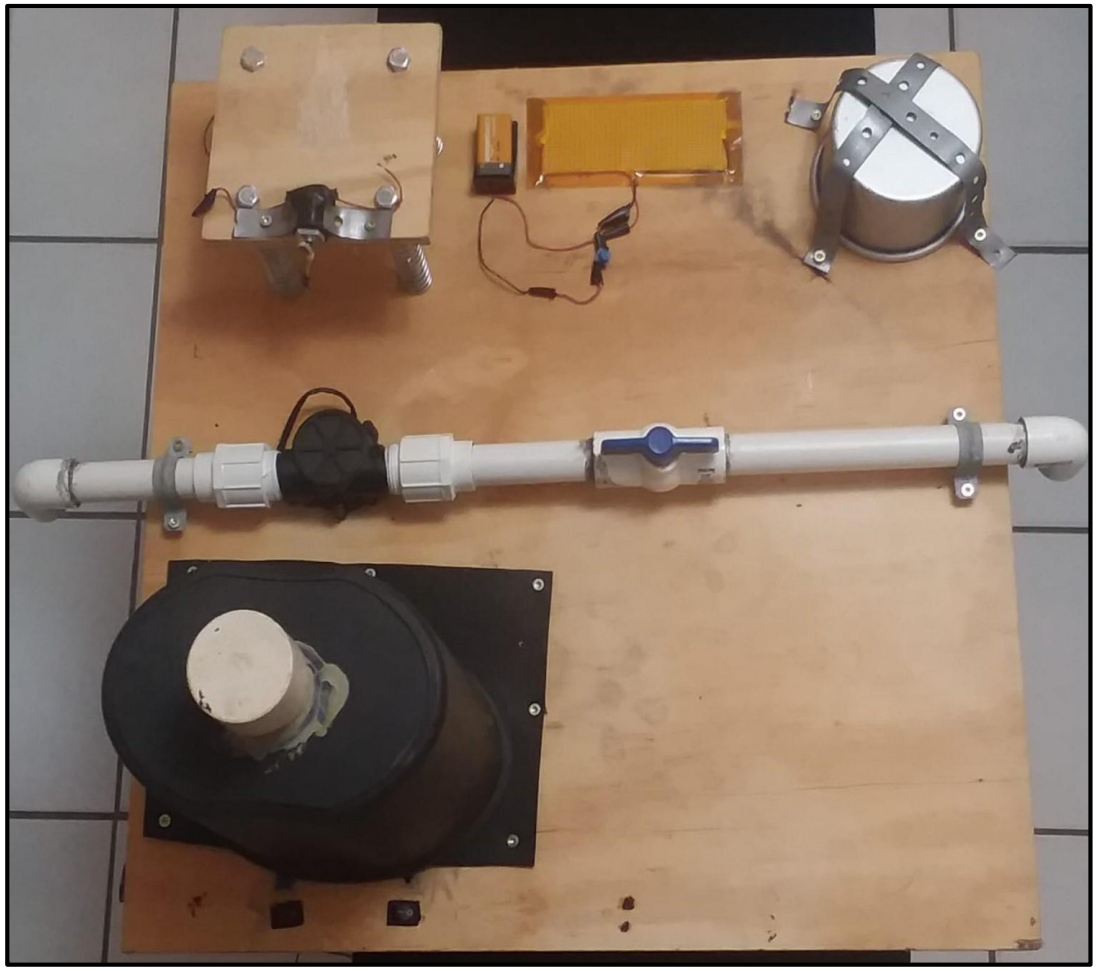
Physical Test Bed Design and Implementation

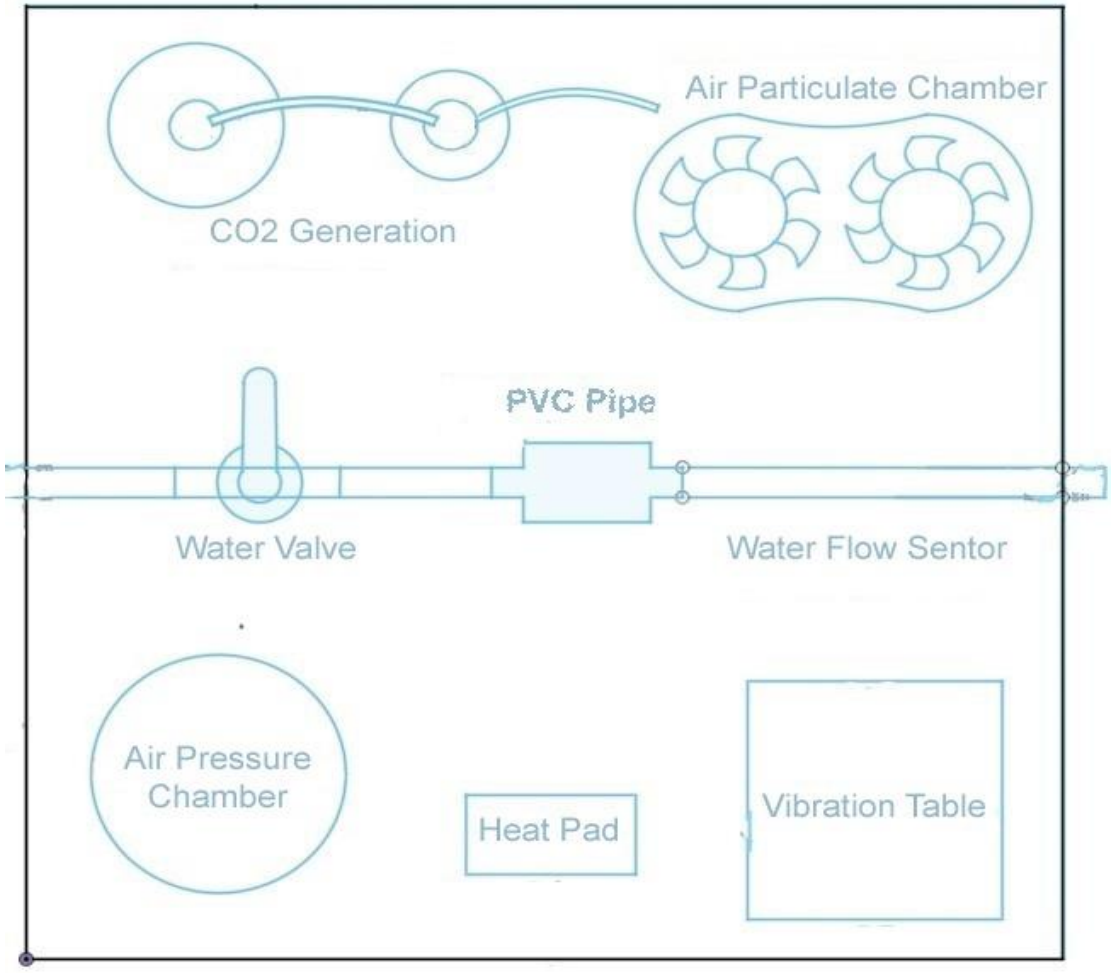
- The hardware platform of the physical test bed includes a variety of sensors to represent the operations of fossil fuel power plant
- All the sensors are compatible with the ESP8266 boards & the Raspberry Pi
- In the initial design, the physical test bed maps to four components of fossil fuel power plant integrated with physical & virtual sensors
- In year 1, following four components are implemented as part of the test bed:
  - Furnace
  - Boiler
  - Turbine
  - Stack

Physical Test Bed Block Diagram



Power Plant Test Bed Mock-up Design & Implementation





Power Plant Test Bed Hardware

- The physical testbed was created to map fossil power plant
- All the required fossil plant conditions are simulated using the physical hardware mentioned below:

Temperature Setup	Water System
Heat Pad	Water Pump
Potentiometer	PVC Pipe
Switch	PVC Elbow
9V battery	PVC Sch. 40 Reducer Bushing
	PVC Sch. 40 Female S x FPT Adapter
	PVC Sch. 40 Slip-Joint x Slip-Joint Ball Valve
	Food Grade Vinyl Tubing
Vibration System	**CO2 System**
Spring Pack	Homer Bucket
Washers	Vinegar
Bolts	Baking Soda
Nuts	Satin Nickel Chest Latches
Plywood	In-Line Shut Off Valve
Vibration Motor	

Physical Test Bed Sensors and Parameters

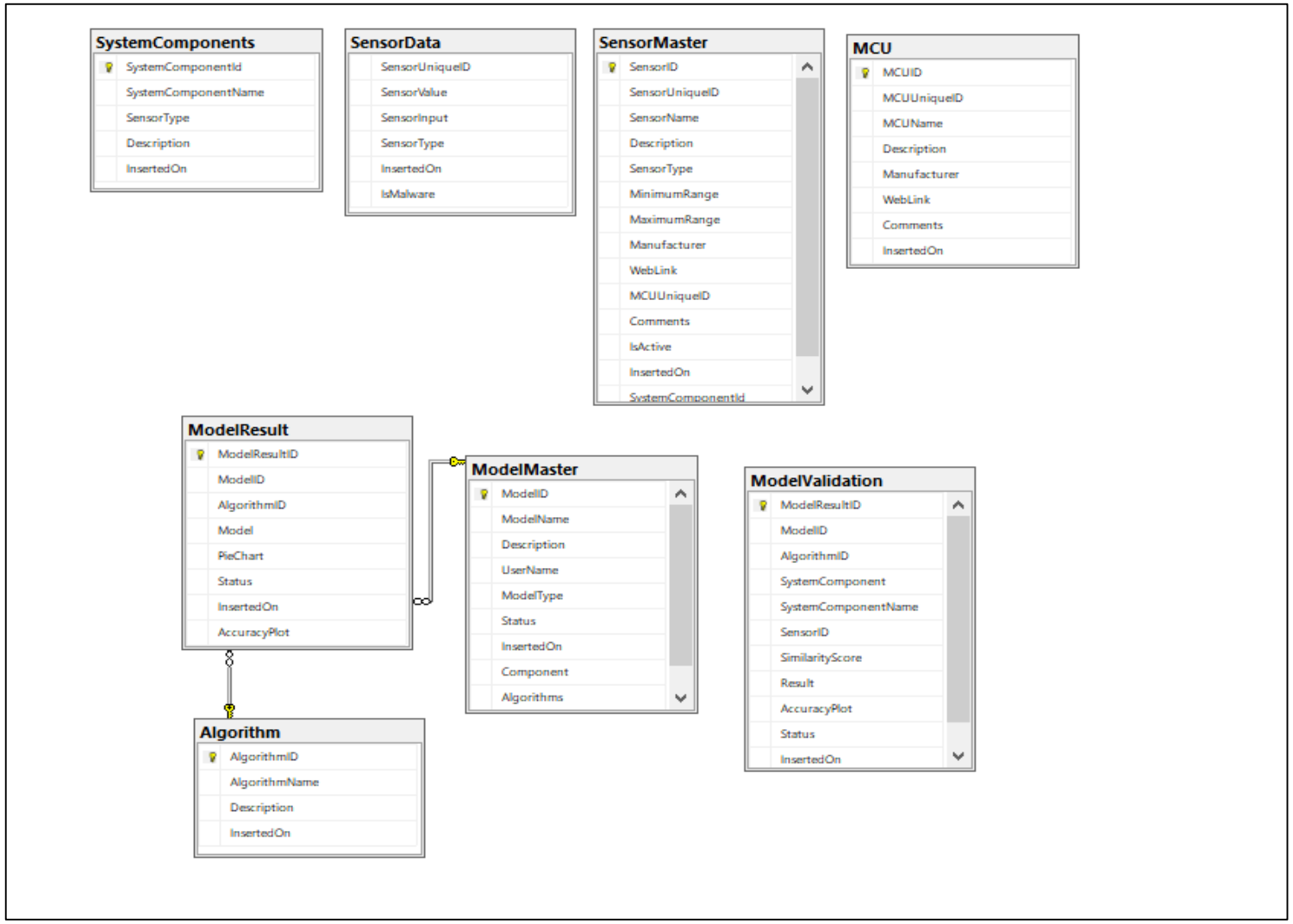
- The test bed prototype consists of temperature, vibration, gas, pressure, air flow, and liquid flow sensors
- The prototype included simulation of six different parameters to be measured by the physical and virtual sensors mounted on each component
- The six parameters measured through physical and virtual sensors are:
  - **Vibration**
  - **Heat (Temperature)**
  - **Pressure**
  - **Water Flow**
  - **Air Particles**
  - **CO2**
- Micro Controller Unit (MCU) is used to uniquely identify the sensor in the network

Test Bed Communication Interface

- Raspberry Pi is used as a communication interface to transfer the data from sensors to the database located in FIU's Artificial Intelligence & Big Data Hub
- A communication interface was developed to read the incoming data from the physical and virtual sensors and store in the SQL Server database
- Message Queuing Telemetry Transport (MQTT) wireless protocol is used to send the sensor data through the communication interface to the server

Database Design and Implementation

- Developed required database objects to receive and store the sensor data
- Developed multiple stored procedures to perform different database activities



Fossil Plant Center

- The Fossil Plant Center(FPC) pilot application is developed using Windows Presentation Foundation (WPF)
- Graphs / plots are developed using LiveCharts open-source library
- WPF application is used to plot/graph real time and historical sensor data, develop machine learning models and predictions
- The live graphs shows the real time data (Raw data) recevied from the sensors and historical graph shows the historical data from the database
- There is also a tree view navigation structure that divides the application into two major sections, the machine learning module, and the power plant components
- By clicking on the power plant component, a tree list appears that contains all the components of the power plant

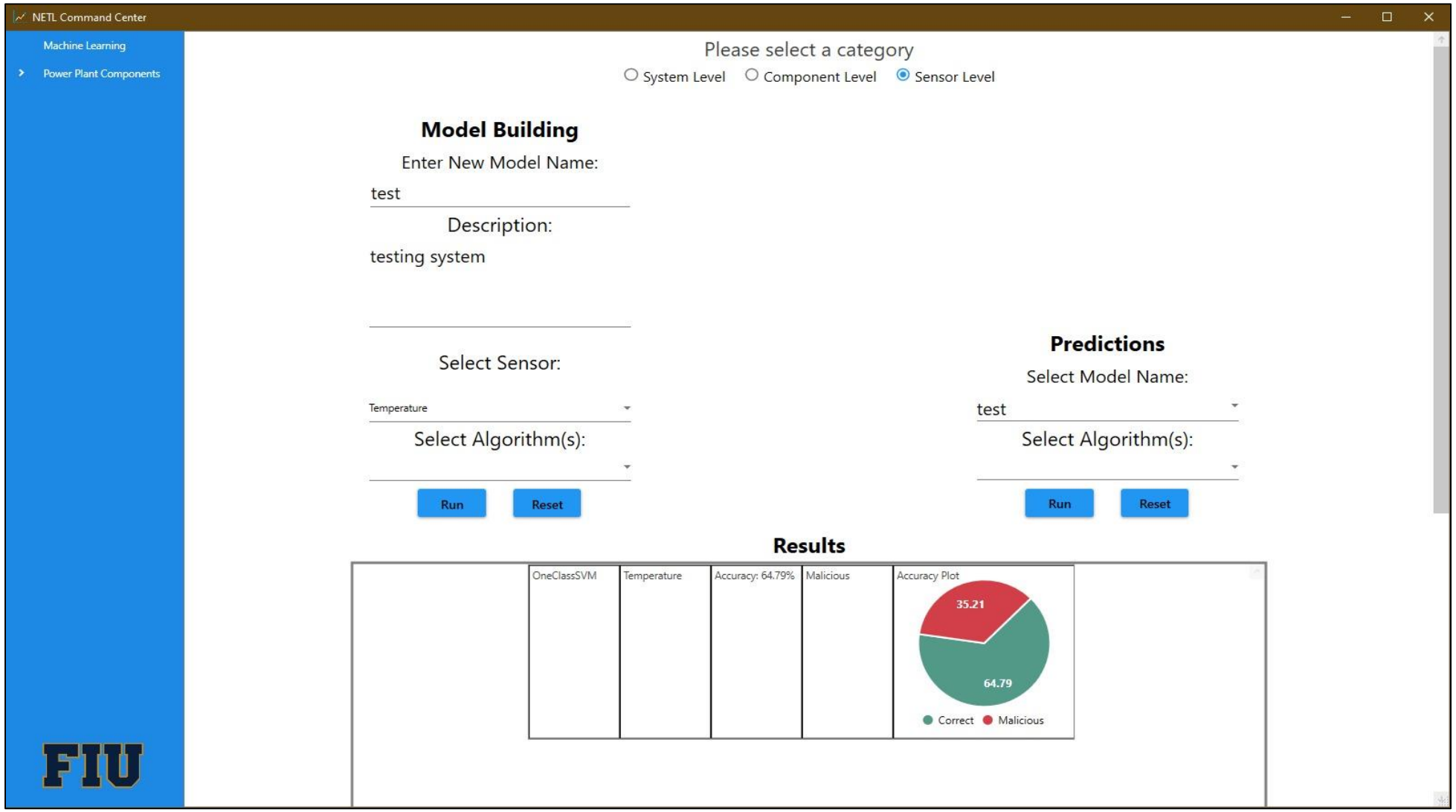
Fossil Plant Center

- The user can filter data for specific sensor on a power plant component by expanding the tree
- User can select the component name to view all the sensors that are active for the selected component
- The machine learning component of the application allows the user to build models and perform predictions
- The data for the machine learning models resides in the database and used to perform in-memory analytics.
- The results of the model include the accuracy score, similarity score, F1 score, component, and the algorithm used to generate the model

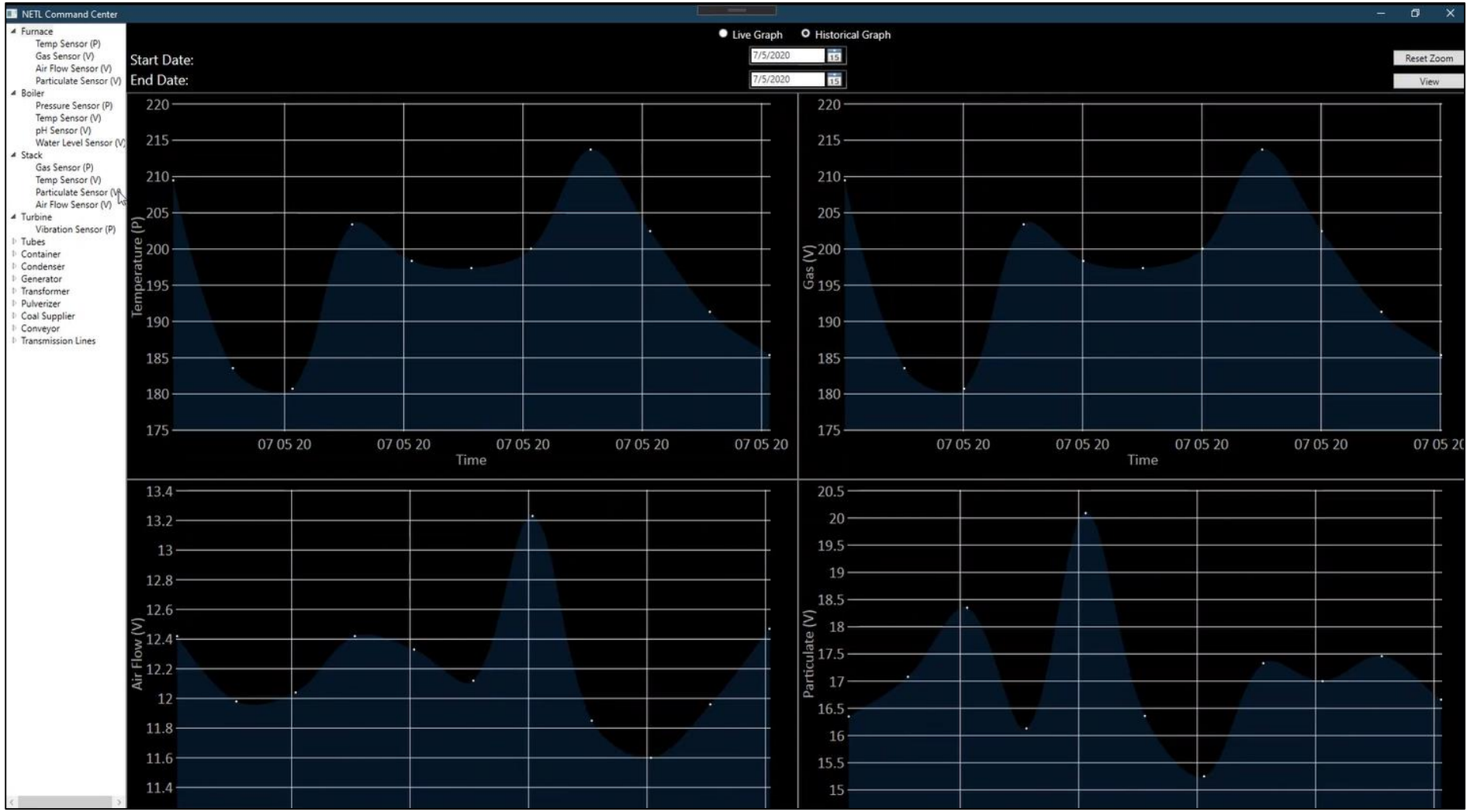
Machine Learning Implementation

- Researched on multiple machine learning algorithms to perform sensor data analytics
- Implemented best-in-class anomaly detection algorithms
- The machine learning model will detect the anomalies in the power plant sensor data and alert the user.
- The algorithms used in developing the model are Isolation Forest, One-Class Support Vector Machine, and Local Outlier Factor
- Machine learning models are created using baseline data from the sensors
- Prediction is performed using the compromised data received from the sensors

Machine Learning Model Building and Results



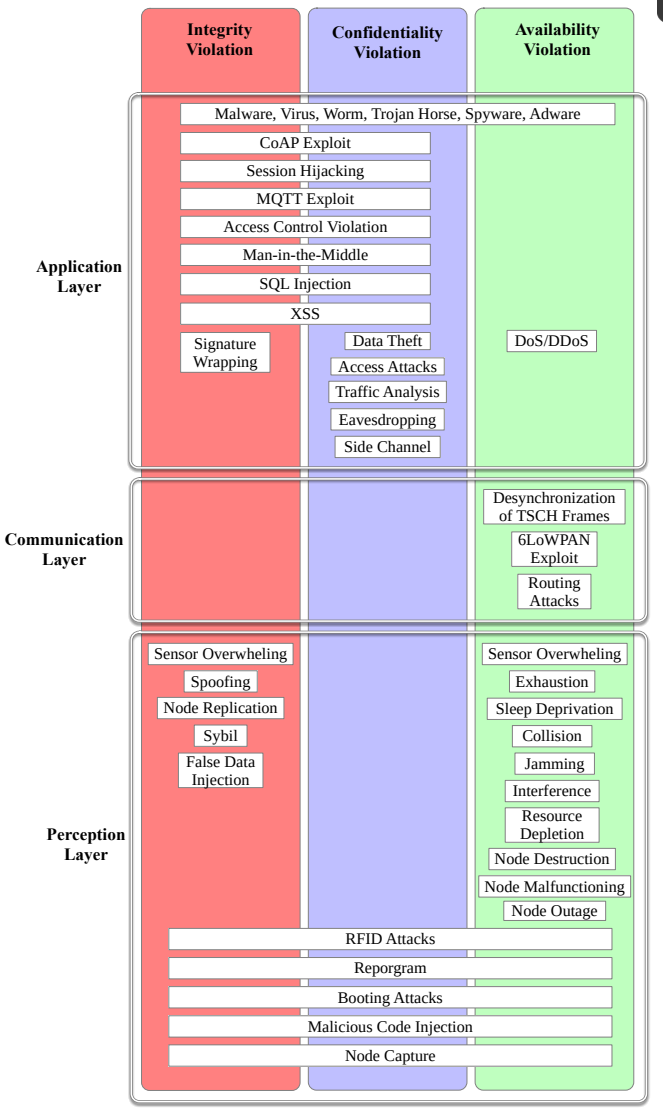
Real-time Visualization of the Sensor Data



Visualization of the Historical Sensor Data

Threat Analysis

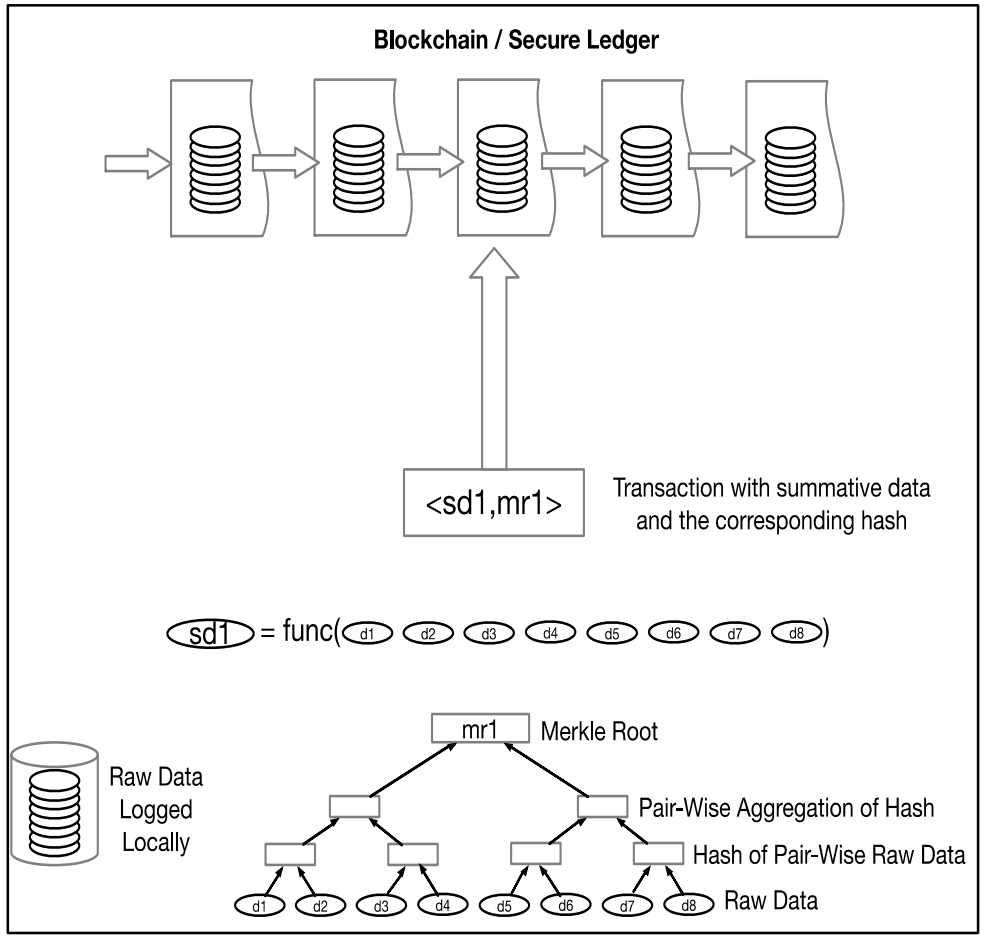
- Performed a thorough threat analysis via a comprehensive literature review
- A conference proceedings paper (titled: *On Threat Analysis of IoT-Based Systems: A Survey*) was accepted and presented (online) at the IEEE International Conference on Smart Internet of Things, August 14-16, 2020, Beijing
- Plan to extend the paper by adding discussions on how Blockchain technology has and could address these threats



Design Authentication and Identification Mechanisms

- A mechanism to authenticate a sensor note is designed and implemented in Python using a quantum-computing-proof one-time digital signature algorithm called Winternitz
- https://github.com/sea212/winternitz-one-time-signature

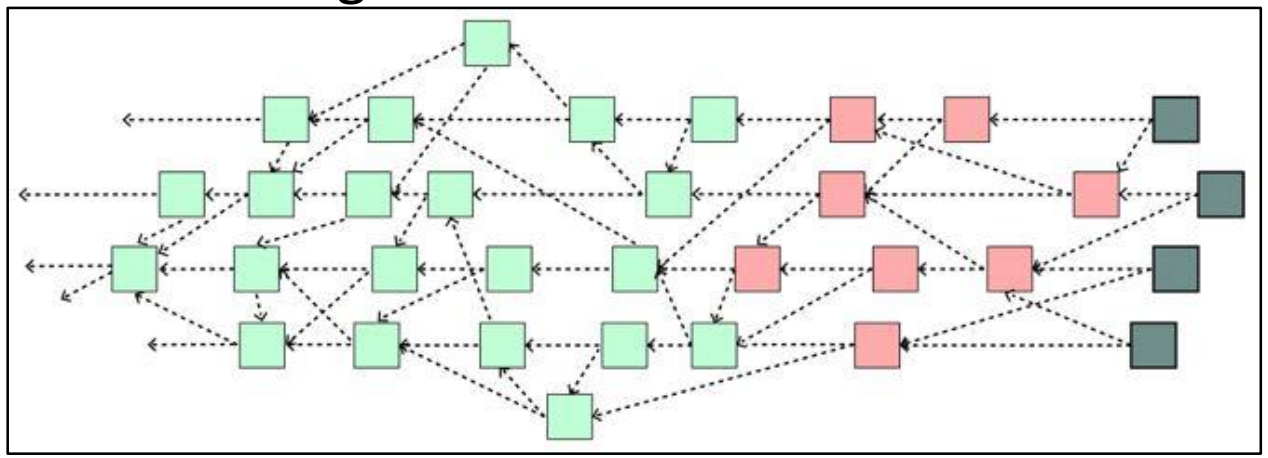
Creating a Strong Linkage between Two Levels of Logging



- Developed the mechanism to enable hierarchical logging where a blockchain is used to store summative sensor data, and raw data are stored locally at replicated data stores
- The linkage is established by organizing the raw sensor records in a Merkle tree and calculate a Merkle tree for each batch of records that will produce a summary
- The raw data is stored locally at the aggregation node using either MongoDB or SQL database. The database should be replicated for fault tolerance.
- The aggregated record is placed on a private IOTA node (for now)

Energy-Aware Blockchain Solution - IOTA/Blockchain Framework

- Completed design of IOTA/Blockchain framework to store the aggregated sensor data
- IOTA distributed ledger framework is implemented
- IOTA is very energy friendly without using expensive proof of work to secure the ledger
- IOTA uses a directed acyclic graph (DAG) as the data structure for the ledger. Hence, it has higher throughput
- IOTA has excellent community support on software development
- Furthermore, IOTA does not charge transaction fees



Visualization of Logged Data – IOTA Mobile Center

- Completed design of IOTA Mobile Center – Mobile App to visualize the data logged in to IOTA platform
- The CSU team is developing a mobile app using the React Native framework
- The app would pull data directly from the IOTA network (current an IOTA private node), store the data on the local SQLite database, and display the aggregated data based on sensor data

Conclusion – Path Forward

- Secure authentication of sensor nodes
  - Explore using SRAM-based physically unclonable functions (PUFs) to generate unique “digital fingerprints” as device IDs. The cryptographic hash of each device ID is uploaded to a blockchain instance
  - While accepting the sensor data from a device locally, it is authenticated by comparing the hash included in the message and the one that is present in that blockchain
- Secure two-level logging with IOTA distributed ledger
- Integration of Deep Learning algorithms for anomaly detection and data aggregation
- Visualization of sensor data with sophisticated analytics functionalities
- Incorporate additional sensors and Fossil Power Plant Components in the test bed.
- Develop integrated pipeline with test bed and Data Aggregator platform
- Develop web-based interface for Data Aggregator –AI platform and test bed to manage sensor data, perform analytics, machine learning model development and prediction

Question and Answers

Thank You