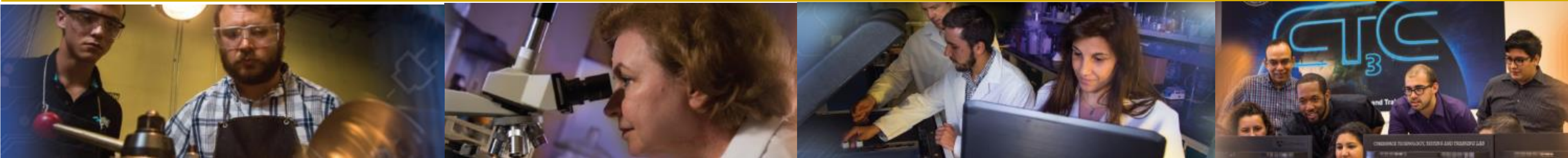
Secure Data Logging and Processing with Blockchain and Machine Learning

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

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

Florida International University



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:

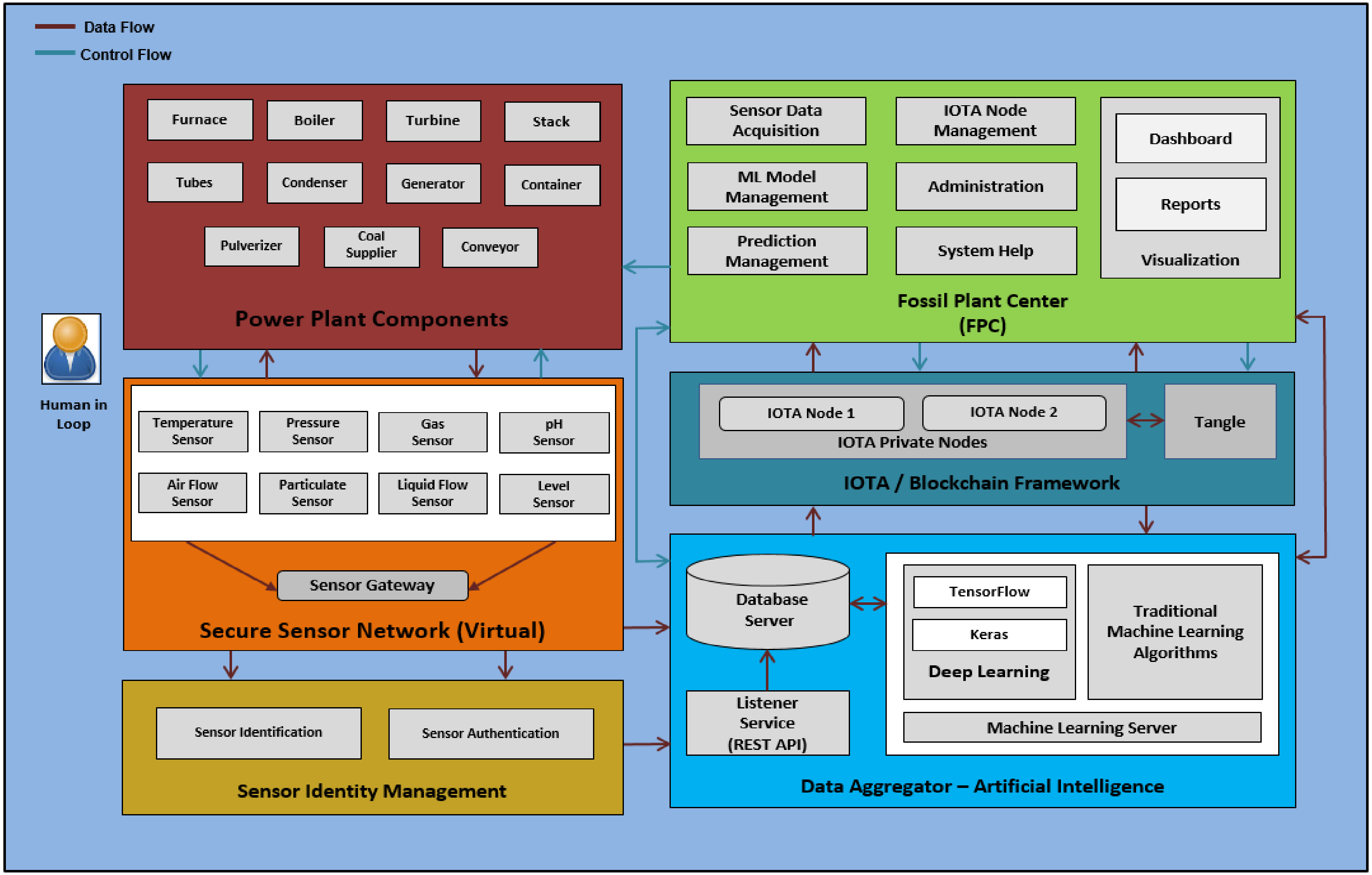
Objective 1: Secure authentication and identity verification of sensor nodes, actuators, and other equipment within a network

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

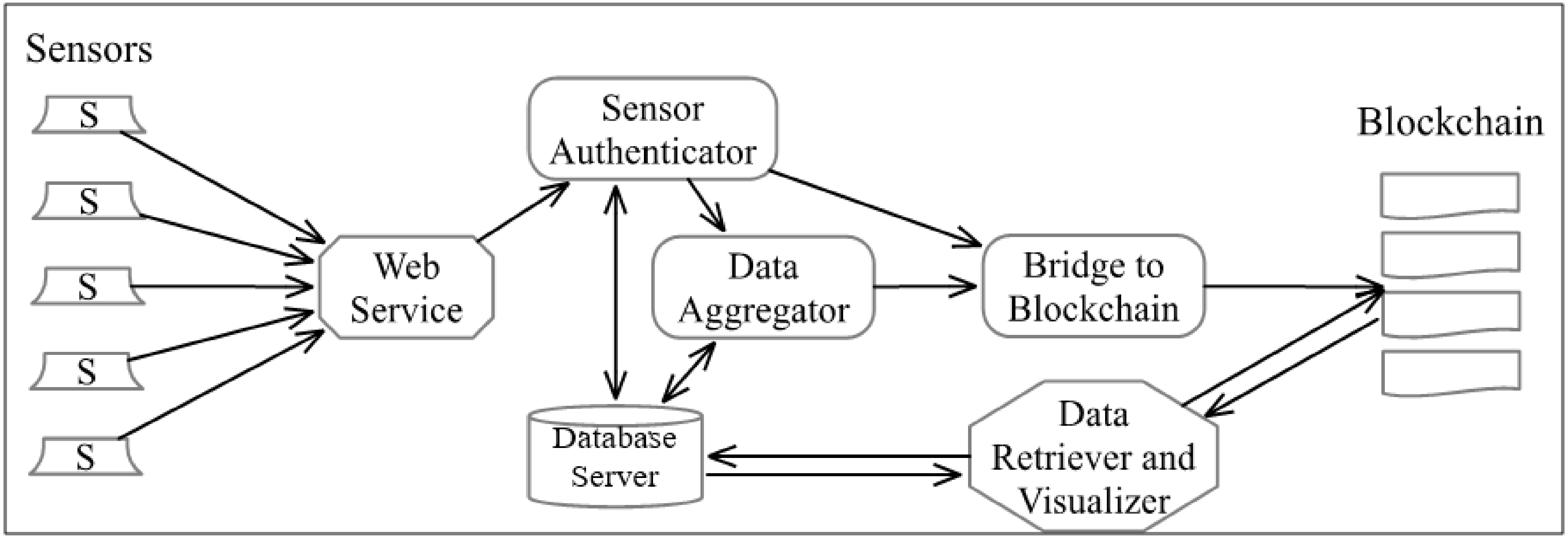
Objective 4: Implement the blockchain technology to provide data security using secured IOTA framework & nodes

Secure Data Logging and Processing with Blockchain and Machine Learning System Architecture



Implementation of a Blockchain-Enabled Secure Sensing Data Processing and Logging System

- This system integrates sensor authentication and identification, data aggregation, storage of raw data in the database server, storage of aggregated data on blockchain, and visualization



NATIONAL ENERGY TECHNOLOGY LABORATORY

Project Tasks

FIU Applied Research Center

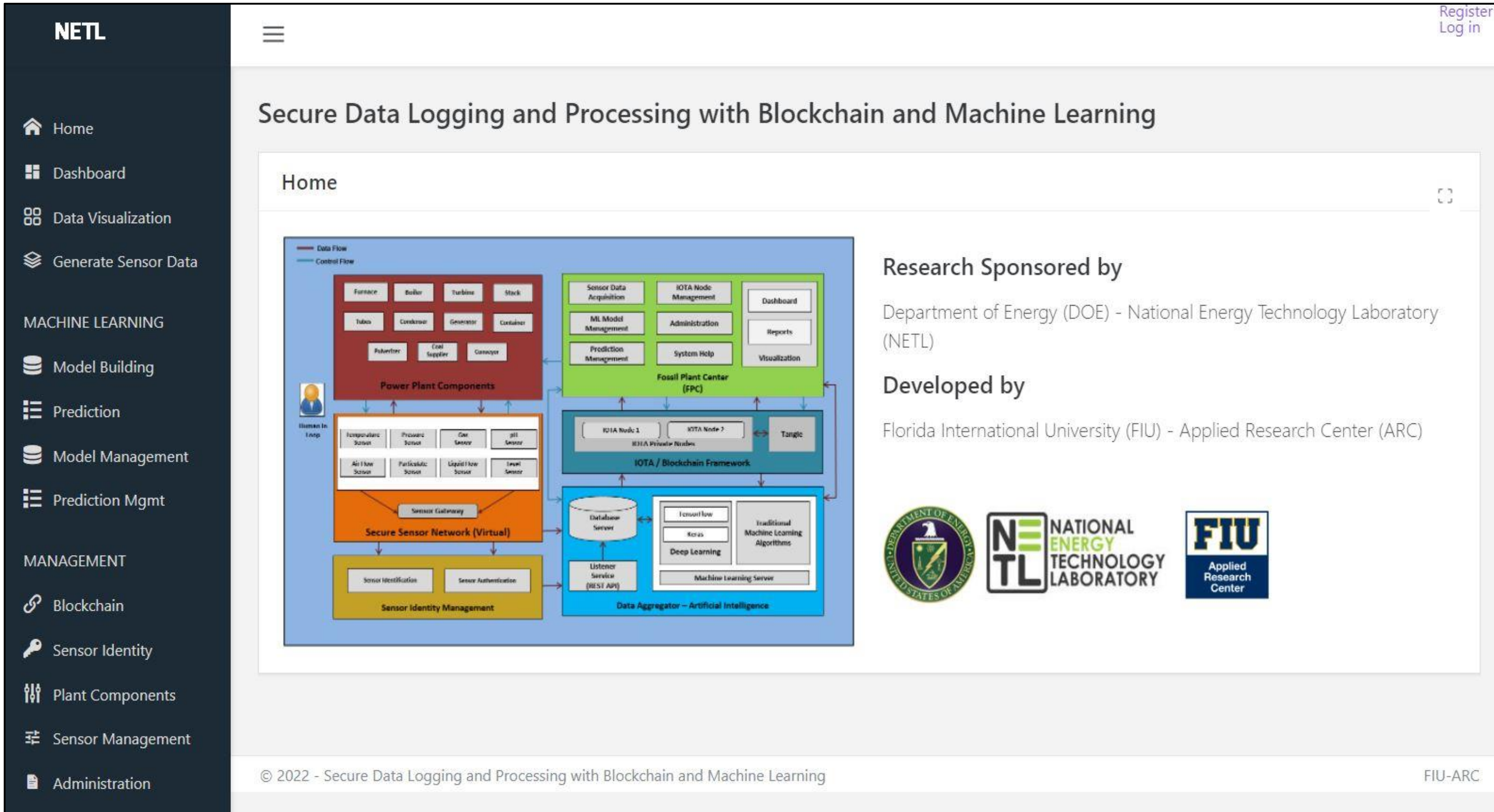
Task 1 - Secure Authentication and Identity Verification of Virtual Sensor Nodes

Task 2 - Data Aggregator / Machine Learning Platform

Task 3 - Secure Logging with Blockchain

CLEVELAND STATE UNIVERSITY 1964

Secure Data Logging and Processing with Blockchain and Machine Learning



NATIONAL ENERGY TECHNOLOGY LABORATORY

Secure Authentication and Identity Verification of Virtual Sensor Nodes

FIU Applied Research Center

- **Sensor Data Generation:** Generation of the sensor data based on the range of the sensors on the Fossil Power Plant components (Turbine, Furnace, Boiler, Stack)
- **Sensor Identity Management:** Sensor identification and authentication of every message received from sensors with Elliptic Curve Cryptography
- **Data Storage:** Storage of raw sensor data and aggregated data from data aggregator and IOTA framework
- **Machine Learning:** Build machine learning models and anomaly detection of sensor data

Cleveland State University 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Fossil Power Plant Components and Sensors Considered for Research

FIU Applied Research Center

The following fossil fuel power plant components and sensors were identified

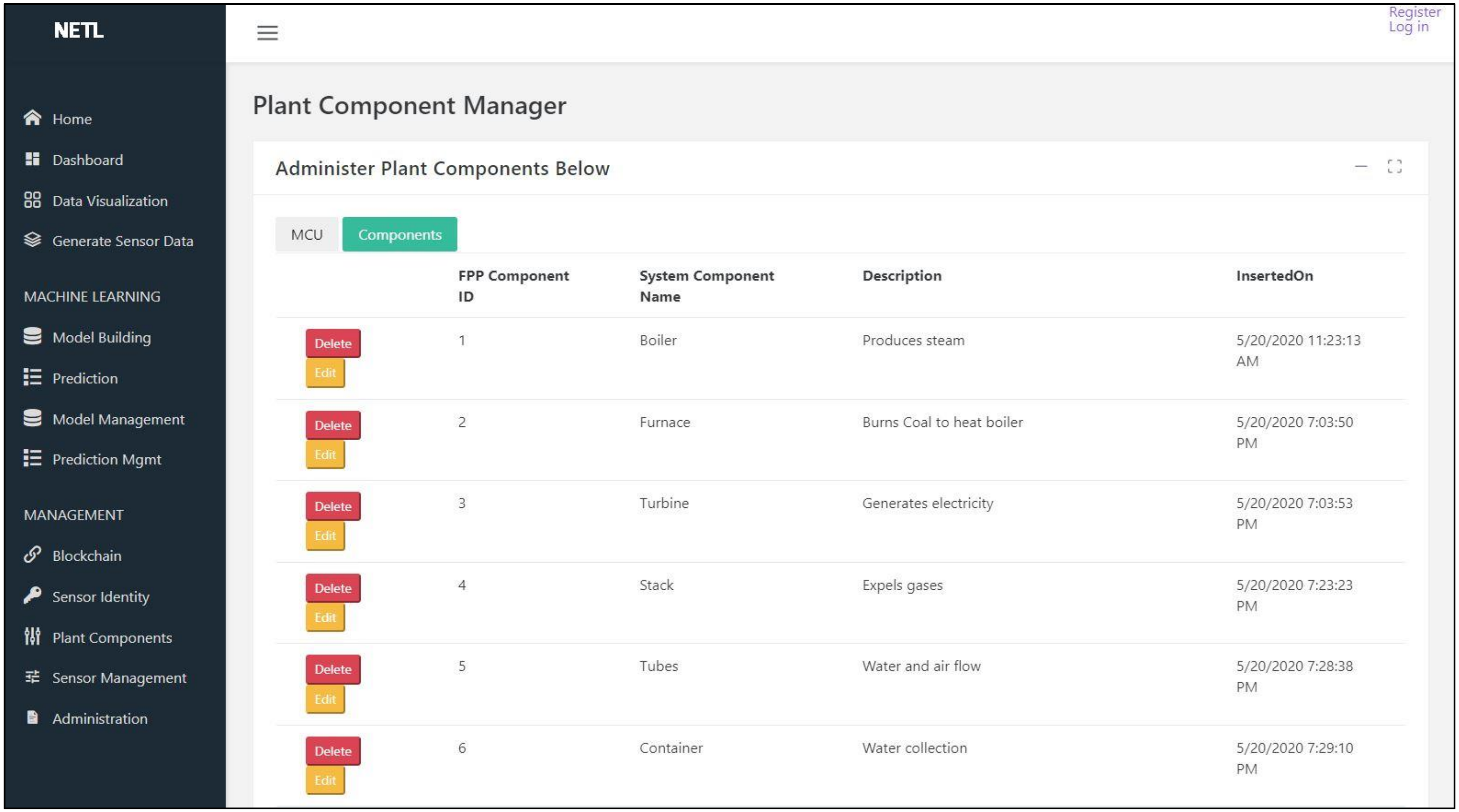
- Temperature sensor (Furnace)
- Pressure sensor (Boiler)
- Vibration sensor (Turbine)
- Gas sensor (Stack)
- Gas sensor (Furnace)
- Air Flow sensor (Furnace)
- Particulate sensor (Furnace)
- pH sensor (Boiler)
- Water Level sensor (Boiler)

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Fossil Power Plant Components Management

FIU Applied Research Center



NETL

Register
Log in

Home
Dashboard
Data Visualization
Generate Sensor Data
MACHINE LEARNING
Model Building
Prediction
Model Management
Prediction Mgmt
MANAGEMENT
Blockchain
Sensor Identity
Plant Components
Sensor Management
Administration

Plant Component Manager

Administer Plant Components Below

MCU | Components

	FPP Component ID	System Component Name	Description	InsertedOn
Delete Edit	1	Boiler	Produces steam	5/20/2020 11:23:13 AM
Delete Edit	2	Furnace	Burns Coal to heat boiler	5/20/2020 7:03:50 PM
Delete Edit	3	Turbine	Generates electricity	5/20/2020 7:03:53 PM
Delete Edit	4	Stack	Expels gases	5/20/2020 7:23:23 PM
Delete Edit	5	Tubes	Water and air flow	5/20/2020 7:28:38 PM
Delete Edit	6	Container	Water collection	5/20/2020 7:29:10 PM

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Sensor Management

FIU Applied Research Center

NETL

Register
Log in

- Home
- Dashboard
- Data Visualization
- Generate Sensor Data

MACHINE LEARNING

- Model Building
- Prediction
- Model Management
- Prediction Mgmt

MANAGEMENT

- Blockchain
- Sensor Identity
- Plant Components
- Sensor Management
- Administration

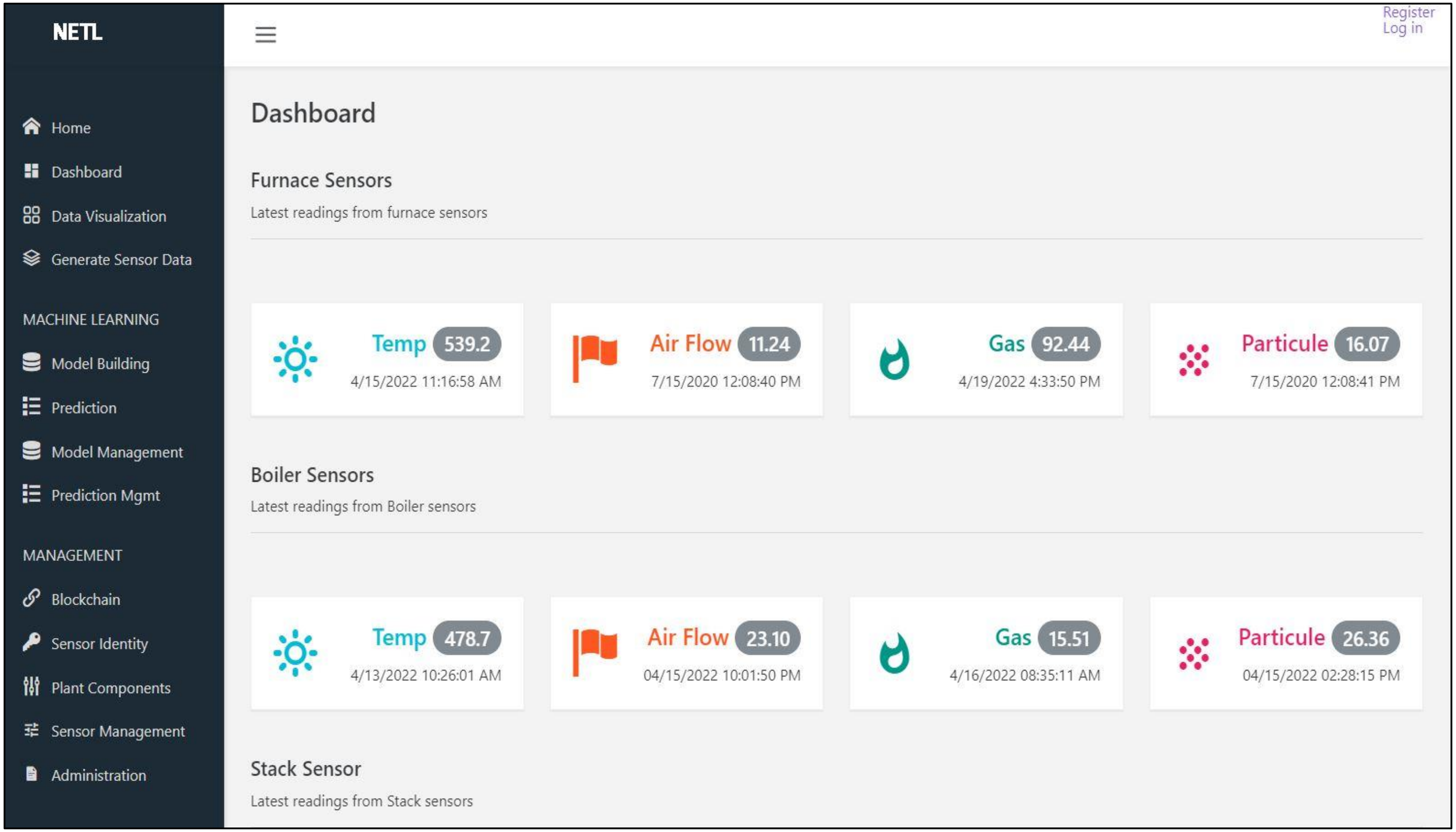
Sensor Management

Administer Sensor Information Below

	Sensor ID	Sensor Unique ID	System Component ID	Sensor Name	Description	Sensor Type	Sensor Input
Delete Edit	1	00f8d84c	1	Pressure	Boiler sensor	Pressure	Virtual
Delete Edit	2	00f882b9	2	Temperature	Furnace sensor	Temperature	Virtual
Delete Edit	3	00f88983	4	Gas	Stack sensor	Gas	Virtual
Delete Edit	4	0085a58c	3	Vibration	Turbine sensor	Vibration	Virtual
Delete Edit	7	00f88983	2	Gas	Furnace sensor	Gas	Virtual

CLEVELAND STATE UNIVERSITY 1964

Sensor Dashboard



NATIONAL ENERGY TECHNOLOGY LABORATORY

Fossil Power Plant Components and Sensors Value Ranges

FIU Applied Research Center

Sensor	Unit	System Component	Benign Value Range
Temperature	°C	Turbine	1.81 – 37.11
Temperature	°C	Boiler	540 – 570
Pressure	hPa	Turbine	992.89 – 1033.3
Pressure	hPa	Boiler	600 – 2465
Vibration	µm (pk-pk)	Turbine	13.31 – 14.07
O_2	% in flue	Stack	1.6 – 4.1
Gas	ppm	Stack	10 – 1000
Gas	ppm	Boiler	0 – 100
CO_2 Gas	% in flue	Stack	8 – 10
NO_2 Gas	%in flue	Stack	10 – 12
pH	-	Boiler	7 - 9

CLEVELAND STATE UNIVERSITY 1964

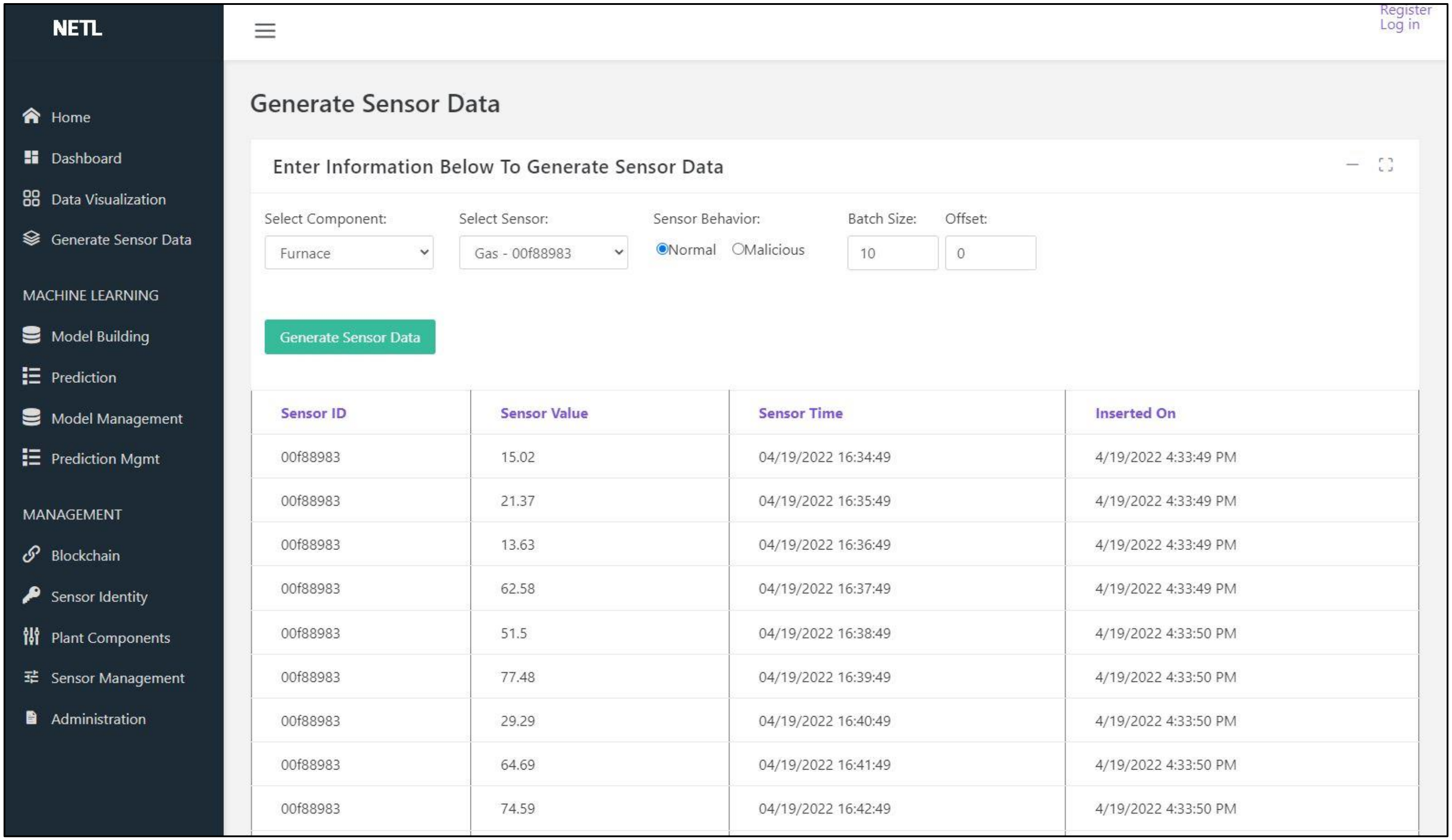
Sensor Data Generation

- Research was conducted to analyze the various sensors mounted on the FPP system components
- Investigated normal sensor value ranges for different sensors
- Synthetic data generated based on the range (Normal / Benign and Malicious) by selecting between lower range – offset and higher range + offset

NATIONAL ENERGY TECHNOLOGY LABORATORY

Generate Sensor Data

FIU Applied Research Center



NETL

Register
Log in

- Home
- Dashboard
- Data Visualization
- Generate Sensor Data

MACHINE LEARNING

- Model Building
- Prediction
- Model Management
- Prediction Mgmt

MANAGEMENT

- Blockchain
- Sensor Identity
- Plant Components
- Sensor Management
- Administration

Generate Sensor Data

Enter Information Below To Generate Sensor Data

Select Component: Furnace
Select Sensor: Gas - 00f88983
Sensor Behavior: Normal / Malicious
Batch Size: 10
Offset: 0

Generate Sensor Data

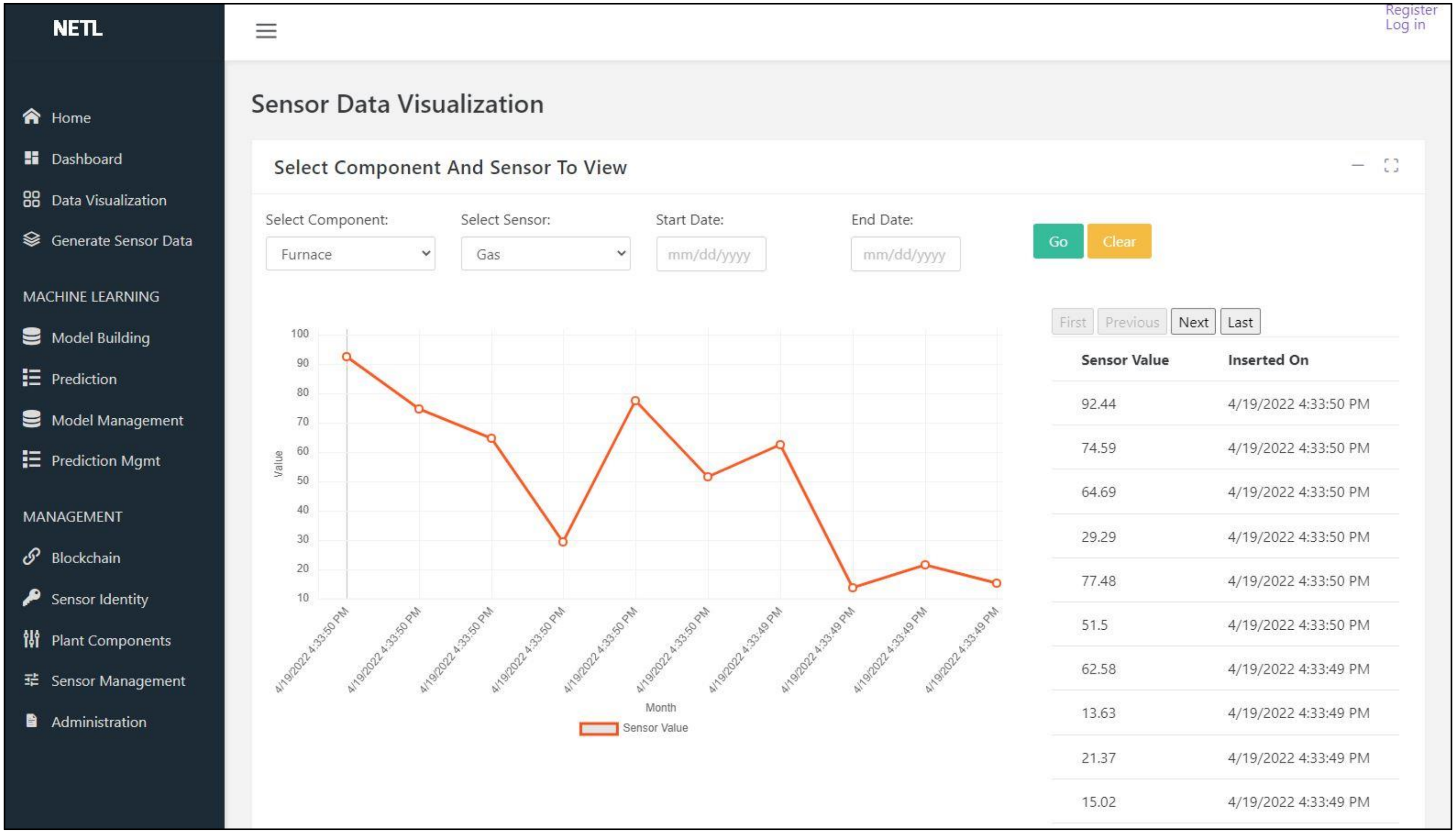
Sensor ID	Sensor Value	Sensor Time	Inserted On
00f88983	15.02	04/19/2022 16:34:49	4/19/2022 4:33:49 PM
00f88983	21.37	04/19/2022 16:35:49	4/19/2022 4:33:49 PM
00f88983	13.63	04/19/2022 16:36:49	4/19/2022 4:33:49 PM
00f88983	62.58	04/19/2022 16:37:49	4/19/2022 4:33:49 PM
00f88983	51.5	04/19/2022 16:38:49	4/19/2022 4:33:50 PM
00f88983	77.48	04/19/2022 16:39:49	4/19/2022 4:33:50 PM
00f88983	29.29	04/19/2022 16:40:49	4/19/2022 4:33:50 PM
00f88983	64.69	04/19/2022 16:41:49	4/19/2022 4:33:50 PM
00f88983	74.59	04/19/2022 16:42:49	4/19/2022 4:33:50 PM

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Sensor Data Visualization

FIU Applied Research Center



Cleveland State University 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Sensor Identification and Authentication

FIU Applied Research Center

- Sensor identity management is essential in mitigating the different attacks which exploit weakness in sensor authentication
  - Injection attacks: The attacker would inject faulty or fake sensing data in the system
  - Sybil attacks: A compromised sensor would impersonate as many legitimate sensors as possible
  - Spoofing attacks: An attacker would pretend to be a legitimate sensor
- Sensor identification and authentication is implemented using Elliptic Curve Cryptography

Cleveland State University 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Sensor Identification and Authentication

FIU Applied Research Center

NETL

Register
Log in

- Home
- Dashboard
- Data Visualization
- Generate Sensor Data

MACHINE LEARNING

- Model Building
- Prediction
- Model Management
- Prediction Mgmt

MANAGEMENT

- Blockchain
- Sensor Identity
- Plant Components
- Sensor Management
- Administration

Sensor Identity

Select Component And Sensor To View

Select Component: Boiler

Select Sensor: Pressure

Start Date: mm/dd/yyyy

End Date: mm/dd/yyyy

Go Clear

First Previous Next Last

Sensor UniqueID	Sensor Value	Sensor Behavior	Sensor Hash	Sensor Verification	Inserted On
00f8d84c	2074.87	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	1402	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	2046.09	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	641.67	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	1961.17	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	1058.08	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	949.64	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	2418.51	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	1201.26	0	System.Byte[]	1	4/18/2022 4:38:35 PM
00f8d84c	713.22	0	System.Byte[]	1	4/18/2022 4:38:35 PM

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Sensor Identification and Authentication

FIU Applied Research Center

- A private key is generated for every system component
- A signature is generated from the data string which is encoded using UTF-8
- The data string is a concatenation of the properties of the sensor that include the sensor ID, batch number, sensor output value, and time.
- A public key is used to verify the signature and record is updated in the database.

```
def append_with_Signature(data_string, priv_key, public_key):
    signer_priv_key = SigningKey.from_pem(priv_key)

    Encode String
    data_bytes = data_string.encode("utf-8")

    # Sign Data
    signature = signer_priv_key.sign(data_bytes)
    return signature

def validSignatureInt(signature, public_key, data_string):
    verify_key = VerifyingKey.from_pem(public_key)

    data_bytes = data_string.encode("utf-8")


    try:
        valid = verify_key.verify(signature, data_bytes)

        return 1

    except BadSignatureError:
        print("BAD SIGNATURE")

        return 0

    #finally:
        #return valid_signature
```

CLEVELAND STATE UNIVERSITY 1964

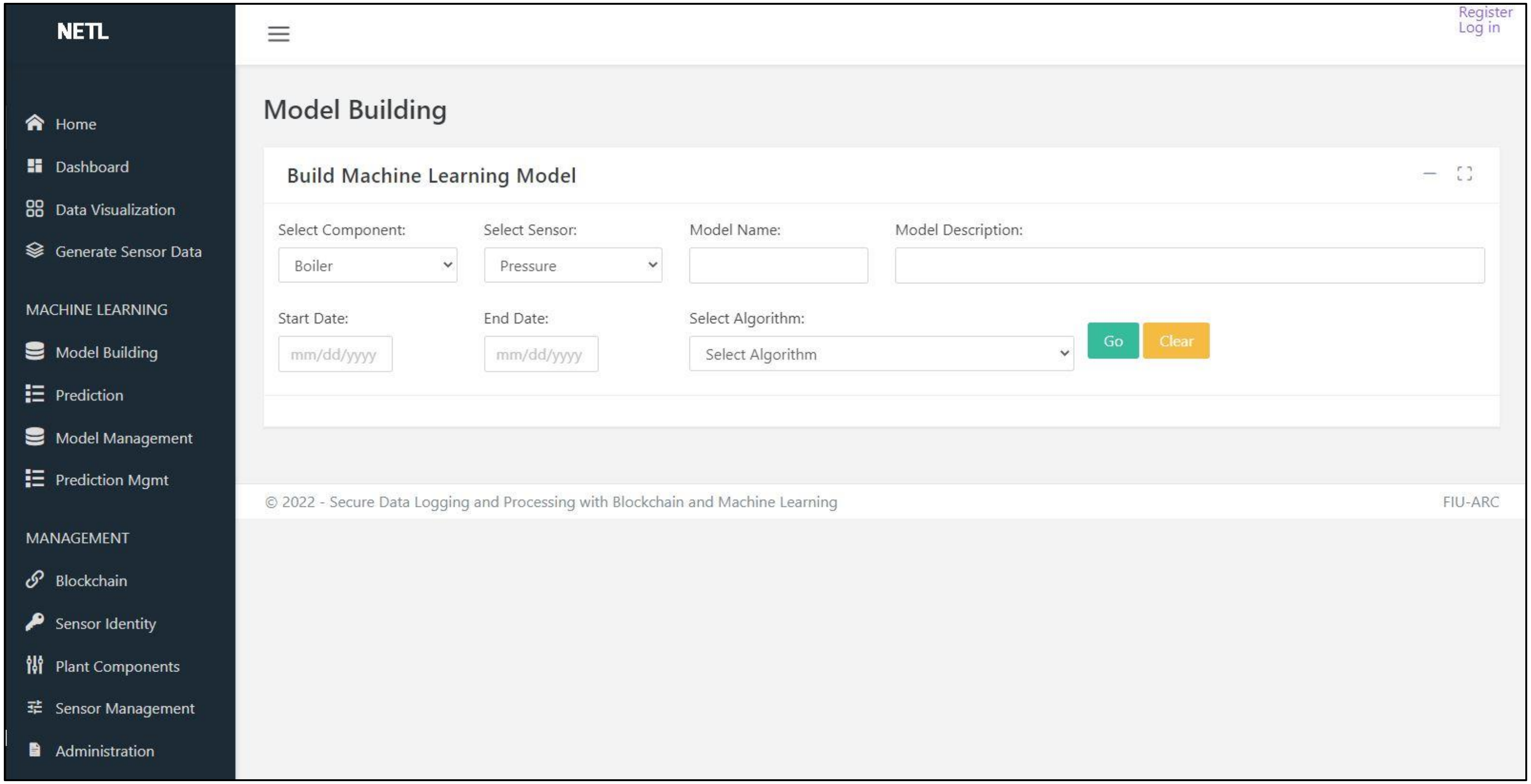
Data Aggregation & Machine Learning Platform

- The following Machine learning and Deep Learning algorithms for anomaly detection are implemented:
  - One-class Support Vector Machines.
  - Isolation Forest
  - Elliptic Envelope
  - Local Outlier Factor
  - Agglomerative Clustering
  - AutoEncoders
- Scikit-Learn, Keras, and TensorFlow frameworks are used for implementation

NATIONAL ENERGY TECHNOLOGY LABORATORY

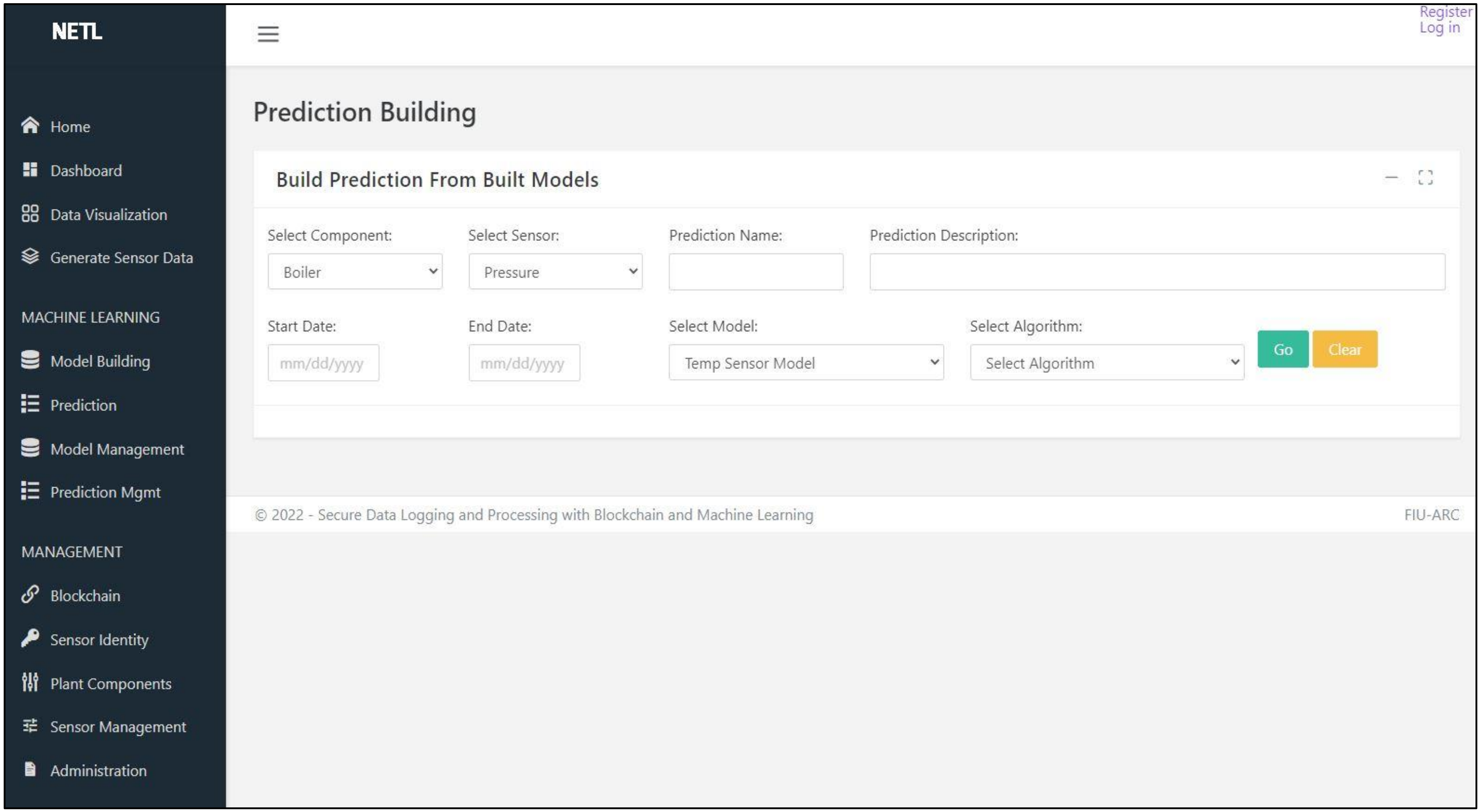
Machine Learning Model Building

FIU Applied Research Center

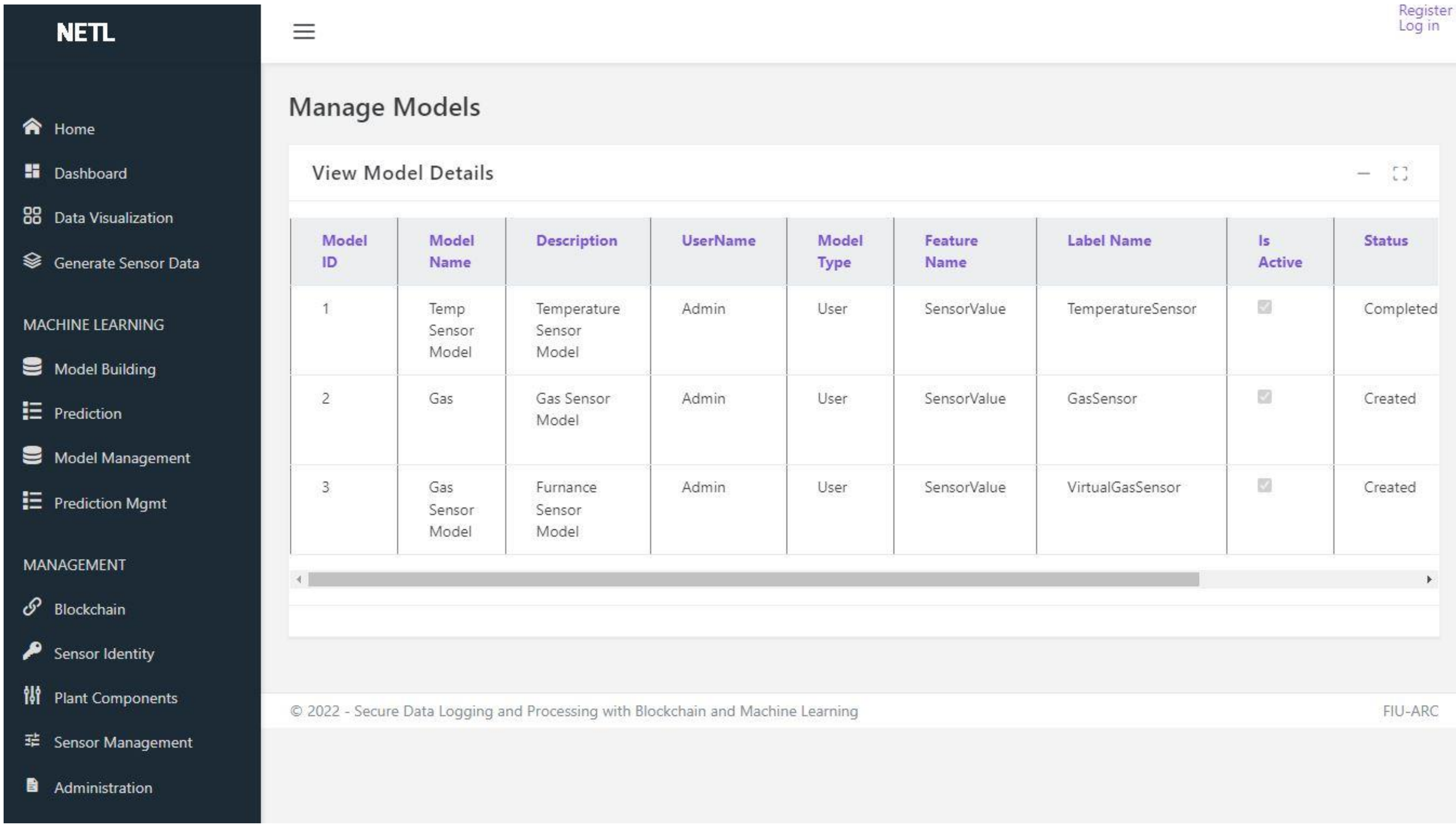


CLEVELAND STATE UNIVERSITY 1964

Machine Learning Prediction



Machine Learning Model Management

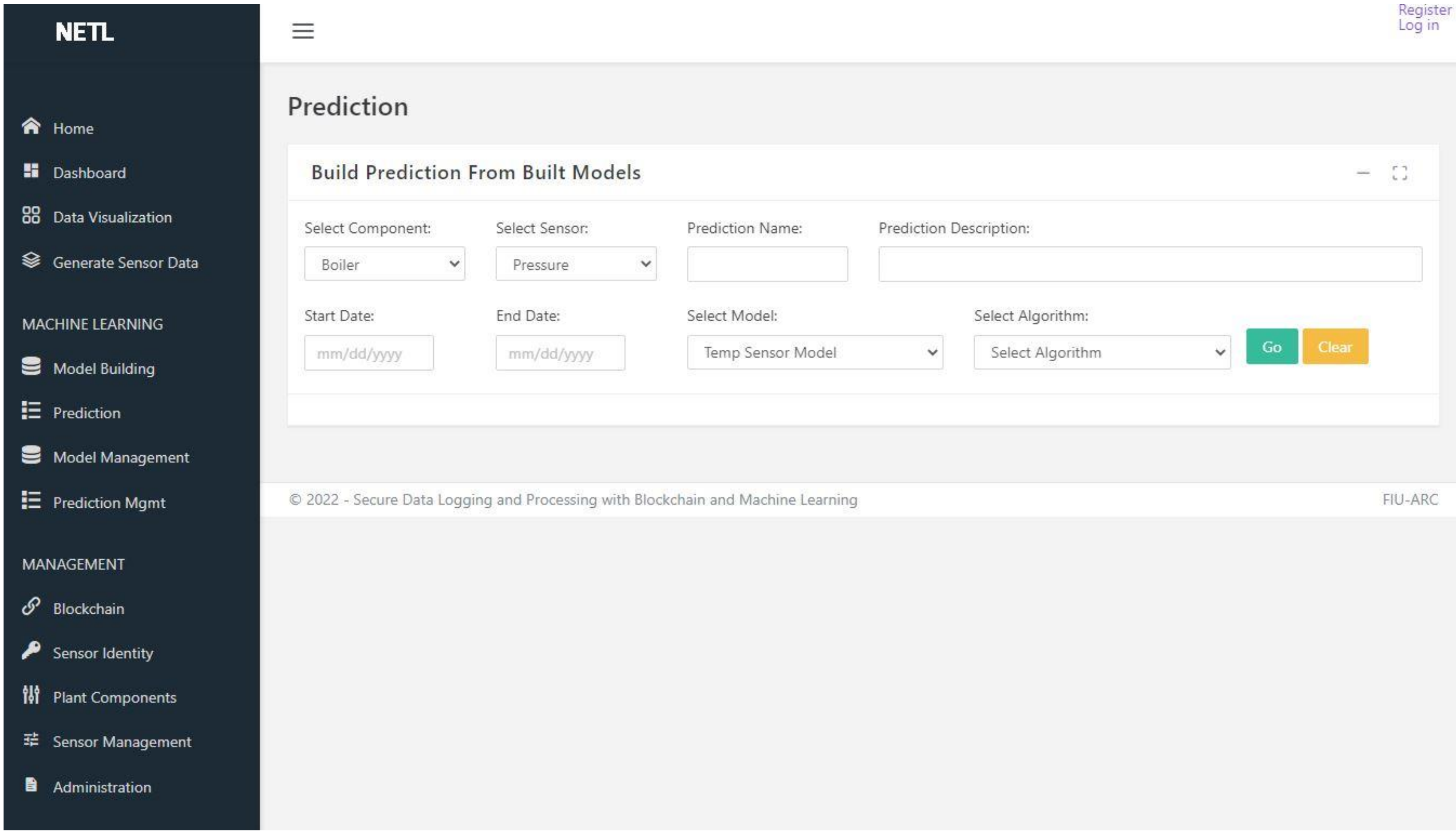


Model ID	Model Name	Description	UserName	Model Type	Feature Name	Label Name	Is Active	Status
1	Temp Sensor Model	Temperature Sensor Model	Admin	User	SensorValue	TemperatureSensor	☑	Completed
2	Gas	Gas Sensor Model	Admin	User	SensorValue	GasSensor	☑	Created
3	Gas Sensor Model	Furnance Sensor Model	Admin	User	SensorValue	VirtualGasSensor	☑	Created

NATIONAL ENERGY TECHNOLOGY LABORATORY

Machine Learning Prediction Management

FIU Applied Research Center



CLEVELAND STATE UNIVERSITY 1964

Anomaly Detection Results

Model	Accuracy
One-class SVM	0.9875
Elliptic Envelope	0.93375
Isolation Forest	0.9875
Local Outlier Factor	0.9875
Agglomerative Clustering	0.4825
AutoEncoder	0.9775

NATIONAL ENERGY TECHNOLOGY LABORATORY

Secure Logging with Blockchain

FIU Applied Research Center

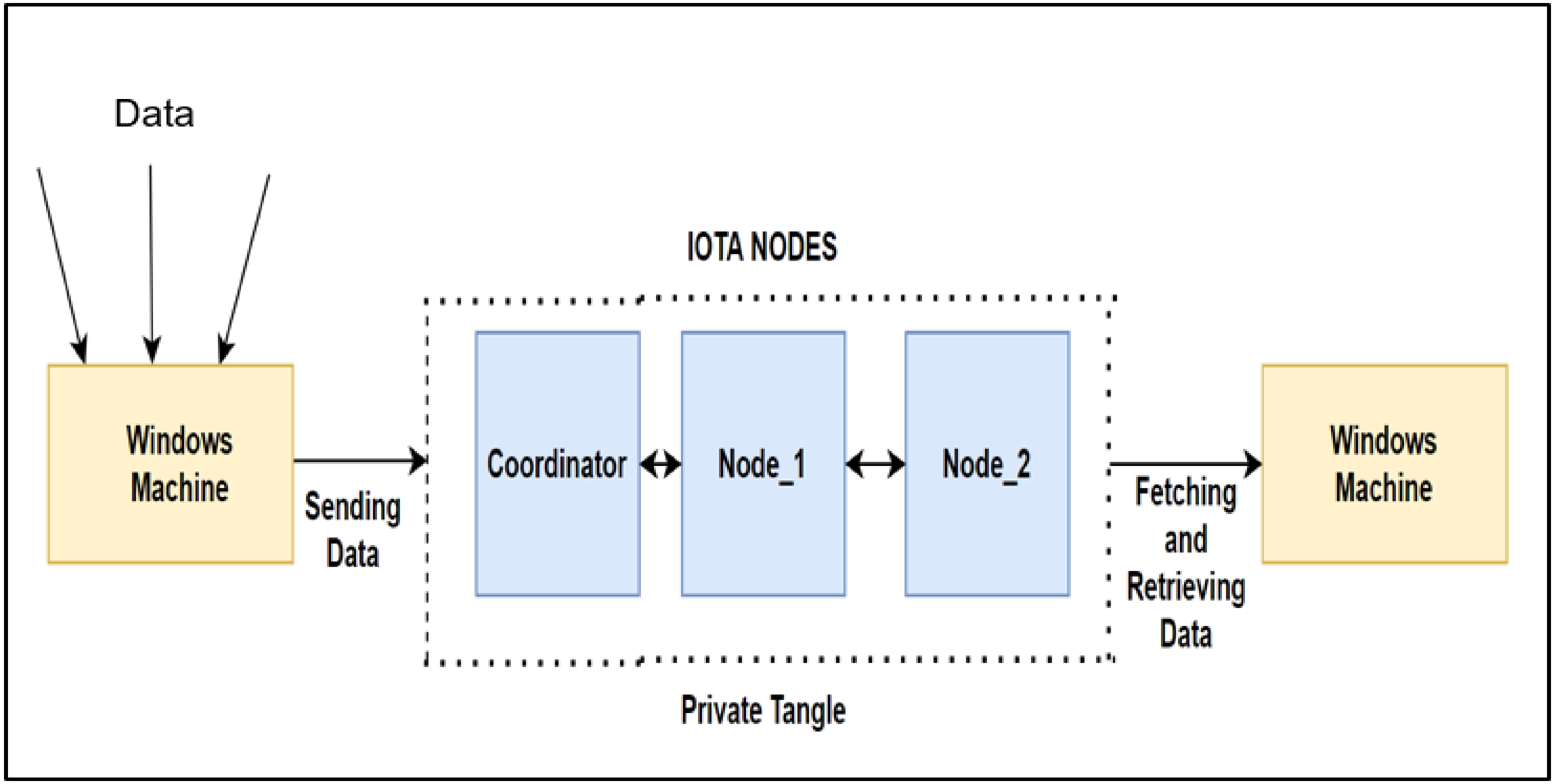
The objectives of the IOTA implementation include:

- Security
- High Scalability and Throughput
- Data Integrity
- Privacy
- Traceability and Auditability
- Secure Storage

CLEVELAND STATE UNIVERSITY 1964

IOTA Private Framework at FIU

A three node (Coordinator, Neighbor node- Node_1, Neighbor node- Node_2) private tangle is implemented at FIU



NATIONAL ENERGY TECHNOLOGY LABORATORY

Coordinator and Neighbor Nodes

FIU Applied Research Center

Coordinator Node:

- The Coordinator is a client that sends signed messages called milestones that nodes trust and use to confirm messages
- Nodes rely on the Coordinator to reach a consensus, therefore each one is hard-coded with the address of the Coordinator
- Nodes use this address to validate the Coordinator's signatures in milestones

Neighbor Nodes:

- Neighbor nodes are mutually connected and communicate directly with each other on the same IOTA network
- To synchronize their ledgers with the rest of the network, all nodes send and receive transactions among their neighbors
- After receiving a new transaction, nodes check that they have the transaction's history in its ledger
- If a node is missing any transactions, it communicates with neighbor node to synchronize with the rest of the network

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

Aggregation and Checksum Generation

FIU Applied Research Center

Aggregation:

- Sensor data aggregation is performed in batches and insert into the tangle
- Calculated the mean, median, mode, Std-Dev, Variance, Min and Max of the generated sensor data batch.
- This gives some meaningful insights about the generated data such as the range, average sensor value and how the sample of data is distributed.

Checksum:

- Checksum is calculated to verify the integrity of the data during transfer.
- Checksum is an aggregated hash value of the sensor data using SHA-256.
- The output hash value will always be the same if the input is same.
- Small change in the input value will result in a completely different hash value.

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

IOTA Storage and Retrieval of the Sensor Data

FIU Applied Research Center

Sensor Data Storage into IOTA

- The IOTA transaction is encoded as JSON structure and stored in the IOTA tangle
- The transaction is broadcasted as a message in the private tangle
- Once the transaction is sent to the IOTA, a Unique Message ID is generated which refers to the current transaction.

Sensor Data Retrieval from IOTA:

- This module queries the private tangle with the Message ID and the IOTA nodes sends the response after fetching the data from the tangle.
- The retrieved data is then decoded into human readable format and stored with transaction timestamp.

```
netl_structure = {
    "sensor_id" : sensor_id,
    "system_batch_id" : batch_id,
    "checksums" : checksum,
    "mean" : sensor_mean,
    "mode" : sensor_mode,
    "median" : sensor_median,
    "min" : sensor_min,
    "max" : sensor_max,
    "standard_deviation" : std_dev,
    "variance" : sensor_var
}
```

IOTA Data Structure

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

IOTA Data Storage in Centralized Database

FIU Applied Research Center

- Once the data was fetched and retrieved from the private tangle, the data is stored in the centralized database

Columns	Value
SensorID	6ad99030
SensorBatchID	6ad99030_1
Checksum	6b186fa5019b4e92bf5713de60efc1f8fdf860 4d44496e833c3e4b48b59fbb05
SensorMean	554.995
SensorMode	544.57
SensorMedian	555.04
SensorStandardDeviation	8.679
SensorVariance	75.32
SensorMin	540.00
SensorMax	569.997
TransactionTimestamp	2022-03-29 21:46:38.000
MessageID	8e23514bc49702c6d9986fb75ac8ae29637d 75e05934e4d680bad16fe33ba5df

CLEVELAND STATE UNIVERSITY 1964

NATIONAL ENERGY TECHNOLOGY LABORATORY

FIU Applied Research Center

System Demonstration "Secure Data Logging and Processing with Blockchain and Machine Learning"

Conclusion

- Prototype for Secure Logging and Processing with Blockchain and Machine Learning for FPP is developed.
- Authentication and Identification of Sensor Data.
- Machine learning platform for Anomaly Detection and Data Aggregation.
- Secure two-level logging with IOTA distributed ledger.
- Blockchain technology to provide sensor data security using IOTA framework.
- Visualization of sensor data with sophisticated analytics capabilities.

Q & A

Thank You

Acknowledgment & Disclaimer

Acknowledgment: "This material is based upon work supported by the Department of Energy Award Number DE- FE0031745."