Modbus 2023 Dataset: Detailed Analysis and Literature Review

### **1. Introduction**

The Modbus protocol is a widely used communication standard in industrial control systems (ICS). In the field of cybersecurity, these systems are often vulnerable to threats targeting critical infrastructures. The Modbus 2023 Dataset is designed to analyze attack scenarios and detect anomalies using machine learning models.

This report focuses on examining the Modbus 2023 dataset, supporting it with a literature review, and analyzing its potential applications.

### **2. Dataset Overview**

#### **Source Information and Purpose of the Dataset**

The Modbus 2023 Dataset was developed by the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick. This dataset aims to capture normal and abnormal behaviors in networks that operate using the Modbus protocol.

* **Purpose:** To detect attacks, analyze anomalous traffic, and ensure secure data communication.
* **Significance in Cybersecurity:** It provides guidance for the development of defense tools against attacks targeting critical infrastructures.

#### **Features of the Dataset**

The dataset includes features at both the network level and protocol level. Key features include:

* **Timestamp:** Records the time of each traffic instance.
* **Source and Destination IP Addresses:** Identifies the parties involved in communication.
* **Connection State:** Indicates whether the traffic is secure or malicious.
* **Protocol Type:** Specifies the type of protocol, such as Modbus.
* **Bytes per Packet:** Measures the volume of traffic.

Each feature provides critical insights for understanding network behavior.

### **3. Literature Survey**

#### **Machine Learning Techniques and Use of the Modbus Dataset**

Research on the Modbus 2023 Dataset often employs the following methods:

##### **Supervised Learning Algorithms**

* **Support Vector Machines (SVM):** Effective for classifying normal and abnormal traffic.
* **Decision Trees:** Analyzes traffic based on specific features.

##### **Deep Learning Techniques**

* **Neural Networks:** Particularly useful for detecting attacks.
* **LSTM (Long Short-Term Memory):** Analyzes time-series data in network traffic.

##### **Anomaly Detection**

* **K-Means:** Clusters anomalous data points.
* **DBSCAN:** Identifies malicious behaviors in network traffic.

#### **Successful Strategies**

* Feature engineering enhances the detection accuracy at the protocol level.
* Advanced preprocessing steps, such as normalization and outlier removal, improve performance.

### **4. Findings and Future Work**

#### **Key Findings**

* The dataset's extensive feature set provides a strong foundation for machine learning models.
* The simulation of diverse attack scenarios enhances its practical relevance.

#### **Future Work and Recommendations**

* **Protocol Customization:** Incorporating additional features specific to the Modbus protocol.
* **Transfer Learning:** Reusing models across different datasets.
* **Time-Series Modeling:** Improving detection of anomalies based on temporal data.

### **5. Appendices and References**

* **Dataset Link:** [Modbus 2023 Dataset](https://www.unb.ca/cic/datasets/modbus-2023.html)
* **References:**
  + [UNB CIC: Modbus 2023 Dataset](https://www.unb.ca/cic/datasets/modbus-2023.html)
  + [Modbus Protocol Overview and Features](https://www.mikrodev.com/tr/modbus-protokolu-ve-tum-ozellikleri/)

ÖMER FARUK GÜNDOĞDU 210408040