Evaluating Concept Drift Detectors on Real-World Data

Ufuk Erol, Francesco Raimondo, George Oikonomou, et al.

University of Bristol, UK, Toshiba Research Europe Ltd.

25 September 2023

Presented by: James Pope
PRESENTATION OUTLINE

INTRODUCTION

DATASET AND EXPERIMENT OVERVIEW

RESULTS AND ANALYSIS

CONCLUSION
Cloud-IoT deployments are ubiquitous and employed in various application domains, including smart buildings.

Compromised endpoint may send normal looking data but add subtle changes that are difficult to detect. Attacker can see device data but not other device’s data.

Concept drift detection and malicious data injection and with a real-world dataset.
CLOUD-IoT
Device (IoT Endpoint)

- Environmental
- Light
- Accelerometer
- USB Powered
- Microcontroller Radio
Testbed
**Histogram IoT Endpoint Sensor Readings**

- **Temperature**
  - Celsius (°C)
  - Number of Readings

- **Humidity**
  - Percentage in Air (%)
  - Number of Readings

- **Gas**
  - Ohms
  - Number of Readings (Log Scale)

- **Light**
  - Lux (lx)
  - Number of Readings (Log Scale)
**CONCEPT DATA DRIFT**

**Natural Data Drift**
Aggregated over time period and used differencing and threshold to detect deviation when we expect signals to be the same. Used moving average approach HDDM\_W.

**Malicious Data Drift**
Gradual changes assumed. Weighted average of the original data with a constant slope, and linear signal to adjust the trend.

\[
data[t] = \alpha \times data[t] + \beta \times (data[ts] - slope \times t)\]

(1)
Natural Data Drift between Devices 85 and b5 temperatures, March 2022. HDDM_W Predictor $\lambda = 0.01$. 

Figure: Devices b5 and 85, Original Temperature

Figure: Devices b5 and 85, Temperature Difference
RNN/LSTM Architecture

Many-to-one

Sequence Length 50. Threshold. NB: Input is vector.
MALICIOUS DATA DRIFT

Figure 4
Malicious Data Drift, Device 61, March 2022. LSTM.

Figure: Device 61, Original Temperature

Figure: Device 61, Drift (10-13 March) Temperature
LONG TERM HUMIDITY DRIFT

Figure 6

LSTM Data Drift Detection, Humidity

Humidity Sensor

22-03  22-04  22-05  22-06  22-07  22-08
LONG TERM TEMPERATURE DRIFT

Figure 7
LSTM Data Drift Detection, Temperature
CONCLUSION

Contributions

▶ Presented dataset and tools for anomalous data injection
▶ Investigated anomalous data drift detection techniques

Future Work

▶ Evaluating more (type/quantity) malicious attacks
▶ Comparing detection techniques

Take Away

Given assumptions, possible to detect maliciously injected anomalies.
QUESTIONS