InSecTT:
Intelligent Secure Trustable Things

Open datasets from the InSecTT project

Document Type	White Paper
Primary Author(s)	Per Erik Strandberg | Westermo
Document Version / Status	1.0 | Final
Distribution Level	PU (public)

Project Acronym	InSecTT
Project Title	Intelligent Secure Trustable Things
Project Website	https://www.insectt.eu/
Project Coordinator	Michael Karner | VIF | michael.karner@v2c2.at
JU Grant Agreement Number	876038
Date of latest version of Annex I against which the assessment will be made	2022-09-05

InSecTT has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 876038. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Sweden, Spain, Italy, France, Portugal, Ireland, Finland, Slovenia, Poland, Netherlands, Turkey

CONTRIBUTORS

Name	Organization	Name	Organization
Per Erik Strandberg	Westermo		

REVIEWERS

Name	Organization	Date
P Priller	AVL	2023-10-19
J Kallio	VTT	2023-10-20
J Karoliny	SAL	2023-10-20

DOCUMENT HISTORY

Rev.	Date	Author / Organization	Description
0.1	2023-10-18	P E Strandberg, Westermo	First Draft
0.2	2023-10-20	J Karoliny	Split SAL datasets into three separate
0.3	2023-10-23	P E Strandberg, Westermo	Minimal changes after reviews
1.0	2023-11-02	P E Strandberg, Westermo	Prepare for public release

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

This white paper aims at briefly summarizing the datasets that are known to have been created and used in the 3-year Europan research project InSecTT. By making this list public, we hope to make the datasets more FAIR (findable, accessible, interoperable and reusable).

The remainder of this document lists the datasets and provides a brief description of them.

Nr	Description
1	UWB weak-NLOS structured dataset
2	UWB dynamic localization dataset
3	AD-EYE open-KTH
4	Modular Ice-cream Factory Dataset on Anomalies in Sensors (MIDAS)
5	ICSFlow: an open-source dataset for Intrusion Detection purposes
6	5G Measurement Data
7	Public transport anomaly data
8	BLE Indoor Zone-Based Localization Dataset
9	WSN Power Consumption Dataset
10	TDMA and BLE Interference Prediction Datasets
11	The Westermo Network Traffic Dataset
12	Air quality dataset collected over one year in the Brindisi airport (Italy).

Table 1: Overview of InSecTT data sets

2 DATA SETS

This section lists the data sets of the InSecTT project.

2.1 UWB weak-NLOS structured dataset

Partner: Johannes Kepler Universität Linz (Austria)

Context: Automotive, but also any other domains in which UWB is used for ranging/localization

Link: https://github.com/ppeterseil/UWB-weak-NLOS-structured-dataset

2.2 UWB dynamic localization dataset

Partner: Johannes Kepler Universität Linz (Austria)

Context: Automotive, but also any other domains in which UWB is used for ranging/localization

Link: https://github.com/ppeterseil/UWB-AEC-trust-dataset

2.3 AD-EYE open-KTH

Partner: KTH Royal Institute of Technology (Sweden)

Context: Sensor data (incl. LIDAR) from vehicles driving on KTH campus

Link: https://www.adeye.se/open-kth

2.4 Modular Ice-cream Factory Dataset on Anomalies in Sensors

Partners: Mälardalen University and ABB AB (both in Sweden)

Context: Sensor data with and without anomalies from a factory

Link: https://github.com/vujicictijana/MIDAS

Reference: Markovic, T., Leon, M., Leander, B., & Punnekkat, S. (2023). A modular ice cream factory dataset on anomalies in sensors to support machine learning research in manufacturing systems. *IEEE Access, 11*, 29744-29758.

2.5 ICSFlow: an open-source dataset for Intrusion Detection purposes

Partner: RISE Research Institutes of Sweden AB (Sweden)

Context: Network data and process state variables logs for supervised and unsupervised ML-based IDS assessment

Link: https://www.kaggle.com/datasets/alirezadehlaghi/icssim

Reference: Dehlaghi-Ghadim, A., Moghadam, M. H., Balador, A., & Hansson, H. (2023). Anomaly Detection Dataset for Industrial Control Systems. *arXiv preprint arXiv:2305.09678*.

2.6 5G Measurement Data

Partner: Munster Technological University (Ireland)

Context: Real-life 5G channel parameters and uplink throughput. From measurements conducted in Cork City and suburbs.

Link: https://github.com/MTU-Insectt/Measurements5G

2.7 Public transport anomaly data

Partner: University of Modena and Reggio Emilia (Italy)

Context: Anomaly dataset collected from bus transportation.

Link 1: https://aimagelab.ing.unimore.it/imagelab/page.asp?IdPage=42

Link 2: https://motchallenge.net/data/MOTSynth-MOT-CVPR22/

2.8 BLE Indoor Zone-Based Localization Dataset

Partner: Silicon Austria Labs GmbH (Austria)

Context: Dataset with more than 20 000 labeled Received Signal Strength Indicator (RSSI) measurements for indoor zone-based localization with Bluetooth Low Energy (BLE)

Link 1: https://doi.org/10.5281/zenodo.4073072

Link 2: https://github.com/juliankaroliny/SAL-Autarkic-Localization-RSSI-BLE-Dataset/tree/1.0

References:

- Karoliny, J., Blazek, T., Ademaj, F., Bernhard, H. P., & Springer, A. (2021, June). Rssi-based machine learning with pre-and post-processing for cell-localization in IWSN. In 2021 IEEE 7th World Forum on Internet of Things (WF-IoT) (pp. 604-609). IEEE.
- Blazek, T., Karoliny, J., Ademaj, F., & Bernhard, H. P. (2021, June). RSSI-Based Location Classification Using a Particle Filter to Fuse Sensor Estimates. In 2021 17th IEEE International Conference on Factory Communication Systems (WFCS) (pp. 27-32). IEEE.

2.9 WSN Power Consumption Dataset

Partner: Silicon Austria Labs GmbH (Austria), Johannes Kepler Universität Linz (Austria), Linz Center of Mechatronics GmbH (Austria)

Context: Power consumption measurements of four different Wireless Sensor Network protocols (Bluetooth Low Energy (BLE), Thread, the EPhESOS protocol, and UWB) for a wireless sensor node use case.

Link 1: https://doi.org/10.5281/zenodo.7762712

Link 2: https://github.com/juliankaroliny/InSecTT_WSN_Power_Consumption_Dataset/tree/1.0

2.10 TDMA and BLE Interference Prediction Datasets

Partner: Silicon Austria Labs GmbH (Austria)

Context: Datasets with measurements of typical traffic on a BLE channel. The first dataset is used as proof-of-concept to demonstrate the possibility of tracking and predicting unknown sources of periodic interference [1]. The second data set includes typical BLE traffic and is used to demonstrate the prediction of the BLE channel hopping pattern [2].

Link 1: https://doi.org/10.5281/zenodo.7152044

Link 2: https://doi.org/10.5281/zenodo.6306288

Link 3: https://github.com/juliankaroliny/InSecTT_BLE_Channel_Sniff_Dataset/tree/v1.1.1

Link 4: https://github.com/juliankaroliny/InSecTT_TDMA_Interference_Dataset

References:

- Karoliny, J., Blazek, T., Ademaj, F., Springer, A., & Bernhard, H. P. (2023). Time slotted multiple-hypothesis interference tracking in wireless networks. IEEE Internet of Things Journal.
- Karoliny, J., Blazek, T., Springer, A., & Bernhard, H. P. (2023, May). Predicting the Channel Access of Bluetooth Low Energy. In IEEE International Conference on Communication (ICC 2023) (p. 1).

2.11 The Westermo Network Traffic Dataset

Partners: Westermo Network Techologies AB, Mälardalen University, RISE Research Institutes of Sweden AB, and TietoEVRY (all in Sweden).

Context: Network traffic with anomalies from industrial communication system.

Link: https://github.com/westermo/network-traffic-dataset

Reference: Strandberg, P. E., Söderman, D., Dehlaghi-Ghadim, A., Leon, M., Markovic, T., Punnekkat, S., Moghadam, M. H. and Buffoni, D. (2023). The Westermo network traffic data set. *Data in Brief, 50*, 109512.

2.12 Air quality data set

Partners: University of Parma and Aeroporti di Puglia (both in Italy)

Context: One year of indoor air quality metrics, e.g., concentration of CO_2, PM of different sizes, etc., from the Brindisi airport (Italy).

Link: Will become available through the IoT Lab at University of Parma, https://iotlab.unipr.it/

3 REFERENCES

[1] Blazek, T., Karoliny, J., Ademaj, F., & Bernhard, H. P. (2021, June). RSSI-Based Location Classification Using a Particle Filter to Fuse Sensor Estimates. In 2021 *17th IEEE International Conference on Factory Communication Systems (WFCS)*. IEEE.

[2] Dehlaghi-Ghadim, A., Moghadam, M. H., Balador, A., & Hansson, H. (2023). Anomaly Detection Dataset for Industrial Control Systems. *arXiv preprint arXiv:2305.09678*.

[3] Karoliny, J., Blazek, T., Ademaj, F., Bernhard, H. P., & Springer, A. (2021). RSSI-based machine learning with pre-and post-processing for cell-localization in IWSNs. In *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*. IEEE.

[4] Karoliny, J., Blazek, T., Ademaj, F., Springer, A., & Bernhard, H. P. (2023). Time slotted multiple-hypothesis interference tracking in wireless networks. *IEEE Internet of Things Journal*.

[5] Karoliny, J., Blazek, T., Springer, A., & Bernhard, H. P. (2023). Predicting the Channel Access of Bluetooth Low Energy. In *IEEE International Conference on Communication* (ICC).

[6] Markovic, T., Leon, M., Leander, B., & Punnekkat, S. (2023). A modular ice cream factory dataset on anomalies in sensors to support machine learning research in manufacturing systems. *IEEE Access*.

[7] Strandberg, P. E., Söderman, D., Dehlaghi-Ghadim, A., Leon, M., Markovic, T., Punnekkat, S., Moghadam, M. H. and Buffoni, D. (2023). The Westermo network traffic data set. *Data in Brief.*

A. ABBREVIATIONS AND DEFINITIONS

Term	Definition
BLE	Bluetooth Low Energy
ICS	Industrial communication system
IDS	Intrusion detection system
LIDAR	Light detection and ranging
NLOS	Non-line-of-sight
PM	Particulate matter (e.g. pollen is typically PM_{10})
RSSI	Received signal strength indicator
TDMA	Time-division multiple access
UWB	Ultra-wideband