



IOTTB

An Automation Testbed for IoT Devices

Sebastian Lenzlinger

University of Basel
Department of Mathematics and Computer Science
Privacy-Enhancing Technologies Group

2024-07-17

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

Introduction

— **auto**

Introduction

— Why are we here?



https://tse3.mm.bing.net/th?id=OIP.o3AVQNkQCCG_2cmhQzD1zQHaeW&pid=Api

Project Description

To study the privacy and security aspects of IoT devices

- *systematically* and
- *reproducibly*,

we need an easy-to-use

- *testbed*

that

- *automates*

(some aspects of) the process of experimenting with IoT devices.

In this presentation I describe an implementation of such a testbed: IOTTB

Introduction

— Why are we here?

- *systematically*: standardization,
- *reproducible*: a systematic approach promises more reproducible experiments, and thus better verifiable results.
- *testbed*: and environment which fixes certain parameters
- *automates*: beyond reproducibility, the level of manual involvement influences feasibility w.r.t. reproduction

Objectives

Key objectives:

- 1. *Automation recipes* [1] for repeated execution of experiments, including data collection and analysis.
- 2. *FAIR* data storage (Findable, Accessible, Interoperable, Reusable) (see [2], [3] and [4]).

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

Motivation

— **auto**

1 Manual setup and configuration of tools

- e.g. tcpdump, Wireshark, Frida
- configurations not interoperable between tools

Motivation

— Problem(s)

1 Manual setup and configuration of tools

- e.g. tcpdump, Wireshark, Frida
- configurations not interoperable between tools

2 Ad-hoc decisions

- file/artefact naming
- measured/extracted data features
- metadata recorded

Motivation

— Problem(s)

1 Manual setup and configuration of tools

- e.g. tcpdump, Wireshark, Frida
- configurations not interoperable between tools

2 Ad-hoc decisions

- file/artefact naming
- measured/extracted data features
- metadata recorded

3 Tailored utilities

- lack interoperability
- require adaptation depending on project

Motivation

— Problem(s)

Problem(s)		University of Basel		2. Motivation	
1 Manual setup and configuration of tools					
• e.g. tcpdump, Wireshark, Frida • configurations not interoperable between tools					
2 Ad-hoc decisions					
• file/artefact naming • measured/extracted data features • metadata recorded					
3 Tailored utilities					
• lack interoperability • require adaptation depending on project					
4 Scattered data and lack of standardization					
• Inconsistent data collection and storage • Difficult to maintain compatibility across projects					

Motivation

— Problem(s)

Motivation

— Problem(s)

1 Manual setup and configuration of tools

- e.g. tcpdump, Wireshark, Frida
- configurations not interoperable between tools

2 Ad-hoc decisions

- file/artefact naming
- measured/extracted data features
- metadata recorded

3 Tailored utilities

- lack interoperability
- require adaptation depending on project

4 Scattered data and lack of standardization

- Inconsistent data collection and storage
- Difficult to maintain compatibility across projects

5 Onboarding challenges

- New members create ad-hoc solutions
- Perpetuates inefficiency and inconsistency

- Problems with current approach:
 1. Inconsistent data collection
 2. Lack of standardized tools and methods
 3. Issues with file naming and data structuring
- Resulting difficulties:
 1. Compatibility across projects
 2. Onboarding new members
 3. Ad-hoc solutions perpetuating inefficiency

Motivation

— Challenges Faced

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

Background

— **auto**

Background – IoT Devices



Figure 1: Smart Lighting



Figure 2: Smart Speakers



Figure 3: Home Surveillance
Camera



Figure 4: VR Headset



Figure 5: Dall-E Diagram of a
Smart Home Network

IoT devices offer **benefits**:

- Home lighting control
- Remote video monitoring
- Automated cleaning

and more! But, they because

1. Used in Homes
2. Connected
 - LAN only
 - Internet
 - ▶ May lead to information leakage

⇒ Security and
privacy **risks**

- Surveillance potential
- Unauthorized data sharing
- Vulnerable to bugs and security failures



Figure 6: Dall-E Schematic Smart
Home Network

- **IoT Devices Overview:**
 - Devices connected to the internet (voice assistants, smart watches, smart home gadgets)
 - Embedded with microprocessors and software
- **Examples of IoT Devices:**
 - Security cameras
 - Home lighting systems
 - Children's toys
- **Importance of IoT:**
 - Physical dimension (sensors, controllers)
 - Internet connectivity

Background

— IoT Devices

- **What is a Testbed?**
 - Controlled environment for experiments
 - Ensures reproducibility and standardization
- **Examples of Testbeds:**
 - Industry and Engineering: Platforms for product development
 - Natural Sciences: Laboratories (e.g., climate chambers, wind tunnels, see [5])
 - Computing: Software testing environments (unit tests, IDEs)
 - Interdisciplinary: Complex systems (e.g., smart electric grid testbeds, see [6])

Background — Testbeds

FAIR Data Principles	 University of Basel 3. Background	Background — FAIR Data Principles	
	• FAIR Data Principles: [4], [3] ▸ Findability: Data should be easy to find ▸ Accessibility: Data should be accessible under well-defined conditions ▸ Interoperability: Data should be integrated with other data ▸ Reusability: Data should be reusable for future research • Purpose: ▸ Improve reusability of scientific data ▸ Guide for designing <i>data storage</i> systems	Findability: • Ensuring data is easily locatable and identifiable. • Use of persistent identifiers like DOIs. • Metadata should be richly described to enable precise searching. • Positive Example: A dataset with a DOI and comprehensive metadata that is indexed in major search engines. • Negative Example: A dataset stored on a personal computer with no metadata and no persistent identifier. Accessibility: • Data should be retrievable by authorized users. • Use of standardized protocols for data access. • Clear access conditions and usage licenses. • Positive Example: A dataset available through a well-documented API with clear access guidelines and permissions. • Negative Example: A dataset stored in a proprietary format that requires special software to access, with unclear or restrictive access conditions.	Interoperability: • Data should integrate with other datasets. • Use of standardized formats and vocabularies. • Ensure compatibility with existing data and tools. • Positive Example: A dataset in CSV format using standardized column headers that align with other datasets in the field. • Negative Example: A dataset in a non-standard format with custom jargon that is difficult to merge with other data sources. Reusability: • Data should be well-documented to allow future use. • Include clear licensing for reuse. • Ensure data quality and provenance are maintained. • Positive Example: A dataset with a clear Creative Commons license, detailed documentation, and a version history. • Negative Example: A dataset with no documentation, unclear provenance, and no stated reuse policy.
Sebastian Lenzlinger	IOTTB	2024-07-17	12 / 34

- **Importance of Network Traffic in IoT:**

1. Captures communication patterns (device-to-server (internet), device-to-device (LAN, e.g., companion apps))
2. Essential for evaluating performance and identifying unauthorized communications

- **Protocol Analysis:**

1. Understand device operation and communication protocols
2. Identify compatibility, efficiency, and security issues

- **Flow Monitoring:**

1. Detect potential security threats (data breaches, unauthorized access, malware)
2. Monitor for anomalies indicating security incidents or vulnerabilities

- **Information Leakage:**

1. Adversaries can passively observe traffic and extract sensitive information
2. Even encrypted traffic can leak information about the smart environment and users

see [7], [8], [9], [7] and [10]

Background

— Network Traffic

- Nw traffic important for various reasons for us
- due to data being encrypted in many cases now adays
 - most methods build down to some type of network traffic analysis

Examples:

- **Leakage:** Personal data and device usage patterns. [7]
 - **Details:** The study found that IoT devices often leak personal data and detailed usage patterns to third-party servers.
- **Leakage:** Home device interactions and usage. [8]
 - **Details:** This research revealed that interactions with home devices can be intercepted, providing insights into daily routines and activities.
- **Leakage:** Device/Network communication *patterns*. [9]
 - **Details:** Sniffing tools can capture communications between IoT devices. WiFi packets expose usage patterns regardless of encryption [10]. Those patterns contain features which can be extracted (i.e. leaked) and fed into machine learning models which are capable of exposing more meaningful information (e.g., identifying devices and their functionality) [11].

In the end these are all some aspect of the same issue: even encrypted traffic leaks information which can be valuable to adversaries.

Background

— Findings from Key Studies

Examples:

- how many people live in a household
- how many devices are in the household
- when which devices are on line
- when, who is home

- **Network Packet Capture:**
 1. Intercepting and storing data packets on a network
 2. Principal technique for studying device behavior and communication patterns
- **Importance in IoT Security Research:**
 1. Main data collection mechanism
 2. Essential for analyzing network traffic

Background

— Packet Capture

- data collection for network traffic

- **Automation Recipes:**
 - Platform agnostic automation
 - e.g., install tool y, retrieve dataset x
 - Integrate with existing scripts/tools
 - Examples in ML
 - *Collective Mind Framework*: [12], [1]
 - Provides reusable recipes for building, running, benchmarking, and optimizing applications
 - Platform-independent or supplemented with user-specific scripts

Background

— Automation Recipes

- **Importance of Automation:**
- Automates workflows irrespective of underlying tools
- the agnostic part is just the goal
- these recipes must be able to integrate well with existing tools, personal scripts
- Enhances reproducibility and efficiency in experiments
- Underlying data has a standardized (w.r.t. to tooling) format, if tool is available

Summary of Key Points	University of Basel 3. Background	Background — Summary of Key Points
• Key Issues Identified: 1. Manual setup and configuration of tools 2. Ad-hoc decisions in file naming, data features, and metadata 3. Tailored utilities lacking interoperability 4. Scattered data and lack of standardization 5. Onboarding challenges for new members • Importance of Addressing These Issues: 1. Improve reproducibility and reliability of experiments 2. Enhance data quality and interoperability 3. Facilitate easier onboarding and collaboration		

- **How IOTTB Addresses These Issues:**

- 1. **Automation Recipes:**

- Standardize the setup and configuration of tools
 - Ensure consistent data collection and analysis processes

- 2. **FAIR Data Storage:**

- Enhance findability, accessibility, interoperability, and reusability of data
 - Improve data management and sharing practices

- 3. **Testbed Design:**

- Provide a controlled environment for reproducible experiments
 - Simplify onboarding and collaboration through standardized procedures

Background

— Return to ...

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

IoTDB

— **auto**

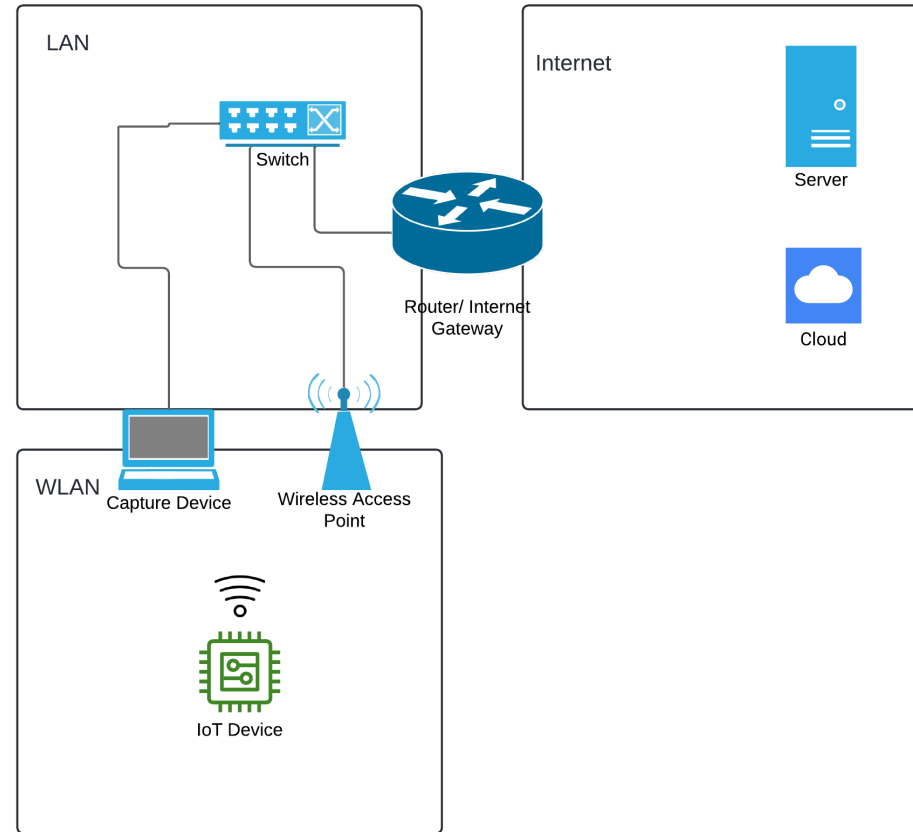


Figure 7: Common capture setup. Separate AP, switch and capturing device.

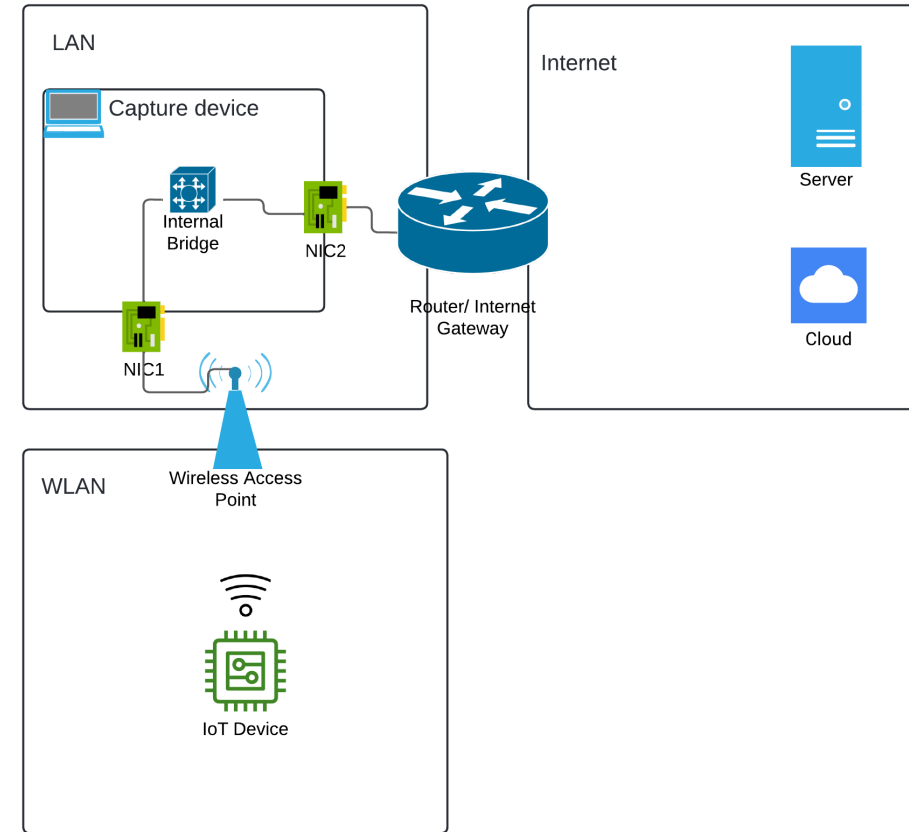


Figure 8: Setup with AP and "Capture Device" on same machine.

[...] testbed for IoT devices which automates aspects of running experiments.

IoT_{DB}
— The testbed

[...] testbed for IoT devices which automates aspects of running experiments.

How is this realized?

IoT_{DB}
— The testbed

[...] testbed for IoT devices which automates aspects of running experiments.

How is this realized?

iottb:

- Python Package
 - Defines Data Storage (implicit in behaviour)
 - Database is a directory hierarchy in a file system
 - DB is a collection of “device”-folders
 - Devices in turn hold some metadata and can have subfolders containing capture data
 -
 -

IoT_{DB}

— The testbed

[...] testbed for IoT devices which automates aspects of running experiments.

How is this realized?

iottb:

- Python Package
 - Defines Data Storage (implicit in behaviour)
 - Database is a directory hierarchy in a file system
 - DB is a collection of “device”-folders
 - Devices in turn hold some metadata and can have subfolders containing capture data
 - Defines a metadata schema for devices, as well as captures
 - Automates collecting of metadata + data

IoTDB

— The testbed

DEMO

IoT_{DB}

—

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

Outlook

— **auto**

FAIR-ness?

Outlook

— Evaluation

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

FAIR-ness?

Findability:

- supported through use of UUIDs, while maintaining human readability.

Outlook

— Evaluation

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

FAIR-ness?

Findability:

- supported through use of UUIDs, while maintaining human readability.

Accessibility:

- to a degree up to user of testbed
- UUID precondition for data met
- metadata makes sense also without data

Outlook

— Evaluation

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

FAIR-ness?*Findability:*

- supported through use of UUIDs, while maintaining human readability.

Accessibility:

- to a degree up to user of testbed
- UUID precondition for data met
- metadata makes sense also without data

Interoperability:

- Used data formats are common and well known (json, pcap)
- Metadata schema understandable given example

Outlook

— Evaluation

1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

FAIR-ness?*Findability:*

- supported through use of UUIDs, while maintaining human readability.

Accessibility:

- to a degree up to user of testbed
- UUID precondition for data met
- metadata makes sense also without data

Interoperability:

- Used data formats are common and well known (json, pcap)
- Metadata schema understandable given example

Reusability:

- Used formats support this.
- Data capture tool (iottb) can be made available
- + rerun with the same configuration

Outlook

— Evaluation

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standard

Automation Recipes?

- iottb automates capture
- Metadata should allow repeating experiments
- want: configure capture based on metadata

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

Questions

— **auto**

Outline

1. [Introduction](#)
2. [Motivation](#)
3. [Background](#)
4. [IoTDB](#)
5. [Outlook](#)
6. [Questions](#)
7. [Appendix](#)

Appendix

— **auto**

Bibliography

- [1] G. Fursin, “Collective Knowledge: Organizing Research Projects as a Database of Reusable Components and Portable Workflows with Common Interfaces,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2197, p. 20200211–20200212, Mar. 2021, doi: [10.1098/rsta.2020.0211](https://doi.org/10.1098/rsta.2020.0211).
- [2] D. Balenson *et al.*, “Toward Findable, Accessible, Interoperable, and Reusable Cybersecurity Artifacts,” in *Proceedings of the 15th Workshop on Cyber Security Experimentation and Test*, in Cset '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 65–70. doi: [10.1145/3546096.3546104](https://doi.org/10.1145/3546096.3546104).
- [3] “FAIR Principles.” Accessed: Jun. 22, 2024. [Online]. Available: <https://www.go-fair.org/fair-principles/>

- [4] M. D. Wilkinson, M. A. Swertz, and et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data*, vol. 3, no. 1, p. 160018–160019, Mar. 2016, doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- [5] T. Vaughan, S. Battle, and K. Walker, “The Use of Climate Chambers in Biological Research,” *Environmental Science & Technology*, vol. 39, no. 14, pp. 5121–5127, 2005.
- [6] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, “Cyber-Physical Security Testbeds: Architecture, Application, and Evaluation for Smart Grid,” *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 847–855, Jun. 2013, doi: [10.1109/TSG.2012.2226919](https://doi.org/10.1109/TSG.2012.2226919).
- [7] J. Ren, D. J. Dubois, D. Choffnes, A. M. Mandalari, R. Kolcun, and H. Haddadi, “Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach,” in *Proceedings of the Internet Measurement Conference*, in IMC '19. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 267–279. doi: [10.1145/3355369.3355577](https://doi.org/10.1145/3355369.3355577).

- [8] D. Kumar *et al.*, “All Things Considered: An Analysis of IoT Devices on Home Networks,” in *28th USENIX Security Symposium (USENIX Security 19)*, Santa Clara, CA: USENIX Association, Aug. 2019, pp. 1169–1185. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity19/presentation/kumar-deepak>
- [9] K. Friess, “Multichannel-Sniffing-System for Real-World Analysing of Wi-Fi-Packets,” in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, Jul. 2018, pp. 358–364. doi: [10.1109/ICUFN.2018.8436715](https://doi.org/10.1109/ICUFN.2018.8436715).
- [10] A. Acar *et al.*, “Peek-a-Boo: I See Your Smart Home Activities, Even Encrypted!,” in *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, Jul. 2020, pp. 207–218. doi: [10.1145/3395351.3399421](https://doi.org/10.1145/3395351.3399421).
- [11] M. Alyami, I. Alharbi, C. Zou, Y. Solihin, and K. Ackerman, “WiFi-based IoT Devices Profiling Attack Based on Eavesdropping of Encrypted WiFi Traffic,” in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA: IEEE, Jan. 2022, pp. 385–392. doi: [10.1109/CCNC49033.2022.9700674](https://doi.org/10.1109/CCNC49033.2022.9700674).

- [12] “Toward a Common Language to Facilitate Reproducible Research and Technology Transfer: Challenges and Solutions,” Jun. 28, 2023. doi: [10.5281/zenodo.8105339](https://doi.org/10.5281/zenodo.8105339).

Appendix

— **auto**

Introduction¹

- IoT Network Diagram: https://tse3.mm.bing.net/th?id=OIP.o3AVQNkQCCG_2cmhQzD1zQHaeW&pid=Api
- Figure 2: <https://i0.wp.com/thegroyne.com/wp-content/uploads/2018/04/Amazon-Echo-Dot-Altavoces-inteligentes-04.jpeg>
- Figure 1: <https://www.multimediaplayer.it/wp-content/uploads/kit-philips-hue.jpg>
- Figure 3: <https://d.otto.de/files/bd42f6e9-ac45-5e1c-8d5f-ac3affcee9d6.pdf>²

¹Images licenced for free share and use to the best of my knowledge.

²Unclear licence