

# Bluetooth Low Energy in the Wild Dataset

Thomas Zachariah  
University of California, Berkeley  
Berkeley, California  
tzachari@berkeley.edu

Meghan Clark  
University of California, Berkeley  
Berkeley, California  
mclarkk@berkeley.edu

Prabal Dutta  
University of California, Berkeley  
Berkeley, California  
prabal@berkeley.edu

## ABSTRACT

In 2015, we performed a study to learn which Bluetooth Low Energy peripheral devices were most prevalent among consumers, as well as which services these devices provided and utilized in practice. Additionally, we sought to investigate the real-world usage of standard Bluetooth services versus custom protocols among developers. The study involved a continuous month-long scan taken on two floors of an academic building. The resulting dataset consists of scan results from approximately 3000 unique devices.

## CCS CONCEPTS

• **Networks** → **Network protocols**; **Network properties**; **Network services**; • **Computer systems organization** → *Embedded and cyber-physical systems*;

## KEYWORDS

Internet of Things, Bluetooth Low Energy, Services, Characteristics

## 1 INTRODUCTION

Since its introduction in 2011, the Bluetooth Low Energy (BLE) radio has become a commonplace component in mobile devices and has made possible many of the resource-constrained peripherals that comprise the Internet of Things. The Bluetooth specification provides well-defined generic access and attribute protocols, as well as a data model consisting of a range of standardized services and characteristics [1]. The specification also defines a set of pre-assigned profiles that correspond with particular applications (e.g. Heart Rate, Environmental Sensing, Running Speed). However, developers are also permitted to implement their own custom services and characteristics. It is not well-understood how often developers decide to use a custom solution versus one provided natively by the BLE specification.

Our dataset helps shed light on the real-world distribution of device types and their utilization of existing BLE standards versus custom protocols and data models. The dataset could potentially yield interesting insights into how well the BLE standards match developer needs in practice.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

DATA'18, Shenzhen, China

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. 978-1-4503-6049-4/18/11...\$15.00  
DOI: 10.1145/3277868.3277882

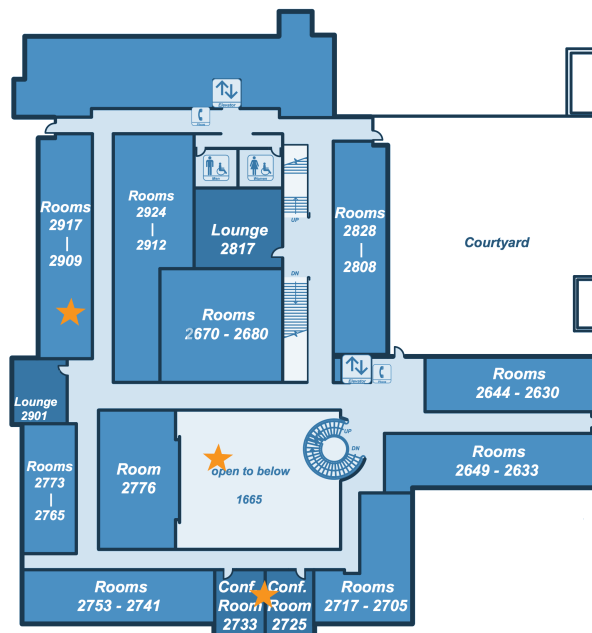
## 2 DATA

In 2015, to gauge the types of BLE peripheral devices that were becoming common in the consumer space and to determine both the pre-defined and custom services and characteristics that were utilized, a month-long scan was performed inside an academic building. A dataset consisting of scans of approximately 3000 unique devices was collected.

### 2.1 Collection and Format

The data consists of parsed Bluetooth scan results taken continuously over a month on the first two floors of the building. The primary scanners were three Raspberry Pi boards with USB Bluetooth dongles, placed in the building as shown in Figure 1. Additionally, two relatively stationary Android devices, that were, at the time, acting as mobile Internet gateways for BLE peripherals (a deployment for a related study on BLE-based IoT gateway architecture [2]), provided supplementary data results from their regular scans.

Each result contains at minimum the parsed advertisement packet from a peripheral device, which consists of Generic Access Profile (GAP) data. If devices were in range for enough time, the scanner attempted to connect with the device to inspect its services and



**Figure 1: Scanner placement in the building.** During the study, BLE scanners were placed in the first floor atrium and in an office and conference room on the second floor.

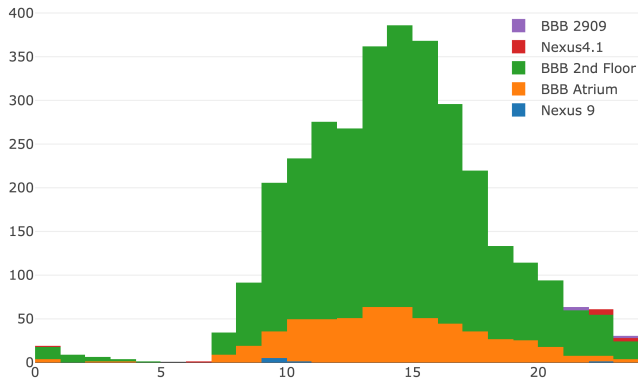


Figure 2: Number of peripherals discovered by hour of day.

characteristics. If any were discovered, their UUIDs and names (for the known pre-assigned profiles) were attached to the result’s entry.

Each entry also includes metadata conveying the time, scanner location, and signal strength when the device’s broadcast was detected and received. Figure 2 depicts how many peripherals were discovered by each scanner during each hour of the day. We also fingerprinted devices that belonged to certain companies and manufacturers based on the services, characteristics, and GAP data present in results, and added these affiliations to the metadata.

The data has been anonymized by replacing device addresses with unique codes (e.g., AB:CD:EF:01:23:45 → 00118) and obfuscating any personal names associated with devices (e.g., Jane Smith’s iPhone → iPhone #12). While connections with peripherals did occur, scanners only retrieved the publicly broadcast GAP data and the rudimentary details of the services and characteristics on the device.

### 3 ACKNOWLEDGMENTS

This work was supported in part by the CONIX Research Center, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA, as well as the TerraSwarm Research Center. This material is based upon work partially supported by the National Science Foundation under the NSF Graduate Research Fellowship and grants CNS-1111541, CNS-1239031, CNS-1350967, and CNS-1824277, as well as by the NSF/Intel Partnership on CPS Security and Privacy.

### REFERENCES

- [1] Bluetooth Special Interest Group. 2016. Bluetooth Core Specifications. (Dec 2016). <https://www.bluetooth.com/specifications/bluetooth-core-specification>
- [2] Thomas Zachariah, Noah Klugman, Bradford Campbell, Joshua Adkins, Neal Jackson, and Prabal Dutta. 2015. The Internet of Things Has a Gateway Problem. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile '15)*. ACM, New York, NY, USA, 27–32. DOI: <https://doi.org/10.1145/2699343.2699344>