Presence Metadata in the Internet of Things: Challenges and Opportunities

Robert Hegarty[1] and John Haggerty[2]

[1]School of Science, Engineering and Environment, University of Salford, The Crescent, Salford, U.K.
[2]DC2D, Beeston, Nottingham, U.K.
r.c.hegarty@salford.ac.uk, john@dc2d.co.uk

Keywords: Internet of Things, Security, Privacy, Metadata.

Abstract: The Internet of Things is an emerging computing paradigm that promises to revolutionise society. The widespread capture and aggregation of data from sensors and smart devices combined with processing using machine learning in cloud computing platforms provides unrivalled insights into our environment. In addition to the numerous benefits (smart healthcare, cities, transportation, etc.) such insights potentially jeopardise the privacy of individuals, organisations, and society as whole. This is despite UK and EU regulations attempting to mitigate the risk of individuals' data exposure and the impact of it on their security. To demonstrate the exploitation of metadata and its threat to privacy, this paper presents Meta-Blue, a Bluetooth Low Energy metadata capture, analysis, and visualisation tool. The results of a case study are combined with an overview of literature on IoT privacy to provide a holistic overview of the challenges and opportunities presented by IoT metadata.

1 INTRODUCTION

The Internet of Things (IoT) has captured the attention of academics and industry alike. Businesses from start-ups to multinational enterprises are embracing and developing IoT devices and services. As these devices and services evolve, they generate, gather and analyse ever increasing quantities of data and metadata about their users. This revolution in data capture, coupled with the processing power of cloud computing and machine learning is enabling the realisation of highly proactive and reactive environments such as smart cities, autonomous vehicles and personalised health care.

Debate continues around the definition of IoT. Whether IoT is a new paradigm or the evolution of embedded systems, wireless networks, artificial intelligence, data mining and other existing technologies is arguable. Regardless of such debate, the market for IoT devices and services is exploding. Gartner predict the market will grow from 4.9 Billion devices to over 20 Billion devices by 2020 (Gartner, 2017). In addition, there is exponential growth in the amount of data collected about users that is stored, analysed and traded by a range of services that make use of this technology.

As with most large-scale systems, there is conflict between the requirements of the three main stakeholders in IoT; end users, governments, and service providers. All have very different perspectives and goals. End users require their data to be secure and easily accessible via apps and web services. Governments also require security. However, there is increasing pressure to provide access to data by state agencies such as law enforcement, intelligence services, etc. which impacts the security of the wider public. Service providers seek to aggregate and monetise data while balancing user privacy with regulatory and legal compliance. Traditionally, metadata has been a grey area in these three perspectives whereby content, such as the text of an email, is separated from communications data, such as events in a network recording the sending of that email.

However, the importance of metadata has increasingly been recognized in Europe, which was enshrined in law by the General Data Protection Regulations (GDPR) (European Parliament, 2016). These regulations impose responsibilities on collectors and processors of data, including metadata, to ensure transparency, proportionality, and appropriateness of data collection. Moreover, GDPR also ensures that data collectors and processors ensure

the security of any personal identifying data that they store. In addition to the legal requirements, the diversity of technology utilised by IoT complicates matters further.

The importance of the IoT paradigm should not be underestimated, nor should the potential for security and privacy challenges and benefits in this emerging domain. This article considers issues regarding security of IoT in general and reviews the established and emerging protocols, infrastructure and services used in this domain. We also identify the security challenges and potential benefits of existing real-world IoT systems to support our position through a novel case study.

This paper is organised as follows. Section 2 provides background on the IoT paradigm and considers the broader issues of metadata and privacy. Section 3 presents an overview of recent and emerging threats, challenges and opportunities affecting IoT devices. Section 4 presents Meta-Blue our tool for visualising Bluetooth Low Energy advertising packets; along with a case study to determine the prevalence of IoT devices that persistently broadcast uniquely identifiable metadata (allowing users to be tracked) and analysis of some popular fitness tracking devices. Finally, we make our conclusions and discuss future work.

2 BACKGROUND

Despite a large amount of literature surrounding IoT and the effort made to define it (Minerva et al, 2015), there is no singular, universal definition. The National Institute of Standards and Technology have published a report on the Network of 'Things' (Voas, 2016) which encompasses the foundations of IoT:

- IoT involves sensing, computing, communication and actuation;
- Things can occur in physical space (e.g. people, vehicles, switches, smart devices, etc.) or virtual space (e.g., data streams, software, files, etc.);
- There are five primitives that describe all aspects of IoT systems; Sensors, Aggregators, Communication Channels, External Utilities and Decision triggers.

As we can see, these foundations are wide and varied, making it problematic to define security and privacy in this domain. Despite the legal changes in Europe and the UK, there is still no 'silver bullet' to address this problem.

To develop an understanding of the importance of IoT, we must understand its characteristics. IoT systems are typically distributed and heterogenous. Much of the computational power used to aggregate, process and provide data to external utilities, such as other services or humans, resides in the core of the network in the form of cloud platforms. Towards the edge of the network, computational power drops off as we shift from intermediate aggregators, such as smart phones or local base stations, to low powered sensors with restricted memory. This facilitates the widespread distribution of sensors in an energy efficient manner.

The three-tiered architecture proposed in IEEE P2413 (IEEE, 2015) encompasses these foundations and is illustrated in figure 1.
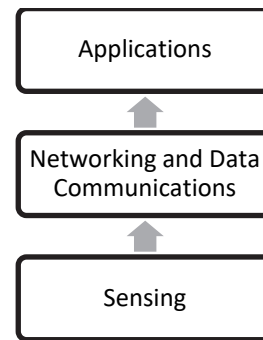

Figure 1: IoT three-tiered architecture, IEEE P2413.

As figure 1 demonstrates, there are three distinct layers with the Application layer accessing data from the Sensing layer via the Networking and Data Communications layer. The authors of (Minerva et al, 2015) note that the definition of IoT varies amongst its various proponents and stakeholders. For this reason, we shall take an evidence-based approach to our analysis of IoT and use this analysis to provide context to the security challenges identified in IoT environments.

IoT devices typically serve as data capture devices which send data to be aggregated and processed by a cloud service. The metadata associated with IoT devices has some similarities with conventional computing devices; for example, it may contain; IP address, UUID, MAC address, etc. These artefacts are used to enable connectivity between the IoT device, intermediate devices and cloud services.

While not all of these characteristics are considered PII (Personally Identifiable Information) under GDPR legislation, the proposed European ePrivacy legislation (European Commission, 2017) directly targets communications metadata and non-personal data. It is hoped that this ePrivacy legislation will bring about a change in the IoT market and enhance the privacy of IoT communications.

Whilst much concern in the media and wider literature has focused on user and data security, metadata provides a greater challenge. For example, users cannot access it without specialist equipment, software or knowledge and may not even be aware that it exists. If it is not open to the users, they may not understand where their data is ultimately collected and stored despite legal obligations to make this transparent. If they do not understand this, then there is very little awareness of how their data is being processed or mined for events, locations or relational information associated with their activities.

In addition to metadata associated with applications, there is other metadata that is created due to the wide variety of protocols required for IoT communications over and above those used more generally in TCP/IP networking. Moreover, each step of data transfer from the device to its ultimate storage may add metadata, for example, MAC (Media Access Control) addresses, IP (Internet Protocol) addresses, UUID (Universal Unique Identifier), etc.

Bluetooth Zigbee, Z-Ware, and WiFi facilitate short range low power communication for IoT devices and all make use of MAC addresses to uniquely identify devices. With the exception of Zigbee these addresses are 48 bits in length with the first 6 bits representing the vendor ID. Zigbee an 802.11.4 standard protocol uses 64 bits for MAC addressing. MAC address randomisation is supported but not mandated by Bluetooth and WiFi, however it has been demonstrated that fingerprinting techniques that negate MAC address randomisation and facilitate the tracking of devices via alternative attributes used to create fingerprints (Vanhoef et al 2016).

Bluetooth has developed over the years to enable a wide variety of features. As of 2017 Bluetooth Low Energy (BLE) (Bluetooth.com, 2019) is the dominant standard adopted by most Personal Fitness Tracking Devices (PFTDs). This subset of IoT devices is of interest as these devices are typically always on and attached to a human subject. Moreover, these devices are associated with highly sensitive data, such as health, temporal geo-location and pattern of life. Combined with communications protocols associated with BLE, unique personal identifying information is generated.

Before considering the details and specification for BLE and the emerging Bluetooth Mesh standards, we must understand how PFTDs utilise BLE. Figure 2 illustrates how PFTDs use BLE and are integrated into the IoT architecture.
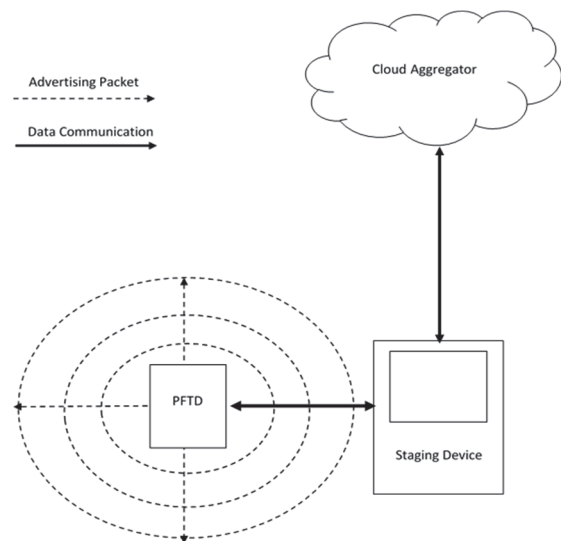

Figure 2: PFTD architecture.

PFTDs are typically constrained devices. Their internal architecture consists of an embedded processor responsible for retrieving data such as step count, temperature and heart rate from sensors and placing it in a small onboard storage area. This memory is used to store data for approximately one week, depending on capabilities and Internet connectivity, to facilitate data collection whilst the device is disconnected from the network. To synchronise with the cloud service, the PFTD connects to a staging device, typically a smart phone or computer, using BLE to transfer data. This is in turn transferred from the staging device to the cloud service using cellular or TCP/IP networks for aggregation and processing. The staging device then retrieves the results from the aggregating and processing service to display the data in graphical or numerical format for the user to view.

To facilitate an *ad-hoc* connection between the PFTD and a staging device, the BLE component(s) on the PFTD periodically broadcasts advertising packets. BLE advertising packets can contain various types of metadata including; Advertisement Address (often a MAC address), List of Services, Service Class UUID, Shortened Local Name, Complete Local Name and Custom Payload (e.g. Power Level). This metadata is used to enable staging devices to distinguish between devices in their vicinity and determine which one to listen to (Argenox, 2019).

It is noted that many devices send advertising packets containing a static MAC address which is persistently used by the device (Hilts et al, 2016). This is despite BLE privacy features being introduced in version 4.0 and improved in version 4.2 of the

Bluetooth protocol. Hilts et al used RaMBLE a Bluetooth advertisement packet sniffing tool created by Contextis (Contextis, 2019) to gather advertising packets from Bluetooth devices. To validate their findings we developed Meta-Blue, a Bluetooth sniffing tool using open source libraries. Details of the tool and case study can be found in section 4.

3 PRESENCE METADATA CHALLENGES AND BENEFITS

IoT provides an ever-increasing list of benefits to society. However, the rapid deployment of complex, heterogeneous, distributed IoT platforms has also brought with it a variety of challenges. In the race for market share, products have been created using the simple formula: take an existing appliance; add a network stack and embedded operating system; and connect it to the Internet for control by an application and/or service. There are some obvious caveats and flaws to this formula that have resulted in unforeseen consequences to the security of the devices, data privacy and the broader Internet. The authors of (Kambourakis, 2017), highlight this describing how the Mirai family of malware has infected vulnerable IoT devices causing unprecedented large scale, high intensity Denial of Services attacks against global Internet infrastructure, impacting not only on the end users of IoT devices but society as a whole.

3.1 Threats and Challenges

The data aggregated by fitness tracking platforms such as Strava, FitBit, Garmin, etc. is used to create a competitive online health social network. These networks focus on comparison of fitness data gathered from end user devices. Users of these social networks compare steps taken, calories burned, geo-locations and times, to compete. Like conventional social networks such as Facebook and Twitter, the interactions between the physical and digital worlds have resulted in some unforeseen discoveries.

For example, while analysing Strava's global heatmap (Strava, 2018), security researcher Nathan Ruser identified the locations of secret military installations around the globe (Ruser, 2018). Ruser could infer the locations and layouts of these installations and the routes taken by front line combat troops on active service. This breach of operational security sent shockwaves throughout the military establishment.

The Strava case study is a powerful reminder that malicious actors may seek to infer information from the data aggregated in the core of the network by the providers of IoT services. Our work focuses on covert data aggregation from devices at the edge of the network and aims to highlight some challenges and opportunities. Such data is routinely broadcast by millions of wearables and other IoT devices around the world every day. In addition to fitness trackers, BLE emissions from smart watches, glasses, headphones, vehicles, tags, children's toys, etc. all contribute to these readily observable broadcasts to provide insights into our patterns of life (Craddock, 2016).

Metadata at the edge of the network potentially poses a broader threat to privacy than centralised aggregated datasets controlled by cloud services providers. Users of IoT devices e.g. fitness trackers, have consciously opted in to these services, the metadata is subject to terms of service, and a single entity is accountable for storage and access control. Metadata collection at the edge of the network indiscriminately and covertly captures data that can be used surveil individuals, recording and/or predicting their movements through pattern of life analysis. More insidiously, such data can contribute to mass surveillance of the public by unknown entities.

It is important to consider that the data captured during our case study was from a single capture device at a fixed location. A much more sinister scenario exists when we contemplate the impact of an attack on a smart city in which many fixed and mobile devices could be compromised and used to track the movement of individuals across a large metropolitan area.

Issoufaly and Tournoux, investigate the exploitation of a Botnet of Bluetooth Low Energy devices, for large scale individual tracking in BLEB (Issoufaly, 2017).

Bluetooth Mesh (Bluetooth.com, 2017) is an emerging standard that enables Bluetooth devices to communicate with each other over a mesh network. This presents the opportunity for a bad actor to potentially create or infect an app and create a botnet of tracking nodes in the form of Bluetooth enabled mobile phones to geo-locate individuals across a broad area via BLE advertisement packets.

As will be demonstrated in the case study section of this paper, many of the threats and challenges highlighted in this sub-section are feasible and the tools required to implement such attacks are readily available in the public domain.

3.2 Benefits and Opportunities

The previous sub-section focused on the privacy challenges arising from the use of IoT devices. In this sub-section, we consider the safety and security benefits that IoT devices can provide. It is outside the scope of this paper to discuss the moral, ethical or legal issues that the following raise should they be implemented.

An obvious benefit of using and identifying presence metadata is in the area of disaster recovery. For example, this data could be used to identify and recover individuals who are trapped by serious environmental hazards, such as buildings destroyed by earthquakes or floods. In addition, it could be used to help co-ordinate the emergency services.

With a potential range of 100m, it is possible to detect the presence of devices at a not insignificant distance for asset recovery. When searching for the proceeds of acquisitive crime, the presence of multiple stolen devices in a single location may be used to identify stolen assets. This approach could be enhanced by either embedding battery powered BLE chips into high value items or keeping a registry of stolen device MAC addresses. This would facilitate a targeted search for such devices by law enforcement.

Kao et al proposed an asset tracking system that utilises Bluetooth Low Energy and WiFi (Kao, 2017) to assist users tracking assets in large buildings.

Following the 2016 terror attack in Manchester, many children were left unaccounted for as they were evacuated from the scene of the attack. IoT devices could be used for the identification of missing children or vulnerable adults. By simply recording the MAC address of the devices the children were carrying (PFTDs, mobile phones, etc.) it would have been possible to locate them by issuing simple instructions to the many hotels that had provided accommodation. This approach could be further enhanced through the implementation of a Smart City sensor grid or Bluetooth Mesh application that responds to missing children reports in a similar way to the US Amber Alert system. Identification information could be obfuscated to preserve the privacy of missing individuals once they have been recovered.

Many serious offenders on licence are required by law to wear a tag to enable the enforcement of curfews and tracking of their location. IoT devices could aid offender probation compliance checks. Given the range of Bluetooth devices, it would be possible to install BLE tracking devices at the entrance of schools to enable staff to be alerted in real time to the presence of a sex offender and take appropriate action.

Roadside capture of advertising packets could be used to ascertain the congestion levels of a route. Such data could be aggregated and combined with existing road sensors to better measure the traffic at key junctions or intersections. This data could be sent to the emergency services to improve emergency response times. Transport for London carried out a similar trial using Wi-Fi to measure foot traffic across various routes on the London underground (TfL, 2017).

Lin et al proposed the use of Bluetooth low energy tags to track dementia patients and keep them safe (Lin, 2015).

It is noted that like many tools and techniques in cyber security, what can be used for good can also be subverted for malicious purposes. Therefore, careful consideration would need to be made about implementation of the examples above. The use of ephemeral keys and end to end encryption in the broadcast of hashed device identities and encrypted device locations is used by some mobile device tracking systems (Apple, 2020) to facilitate recovery of devices while preventing unauthorised tracking. Such an approach may be adapted to facilitate the tracking of individuals in the scenarios previously described.

4 META-BLUE CASE STUDY

This section demonstrates the extent to which PFTDs advertising broadcasts may be employed to identify individuals through presence metadata. The case study captured and collated advertising packets from BLE devices to determine if it is feasible to identify an individual device over a sustained period.

To capture BLE advertising packets and validate the results we developed Meta-Blue (Meta-Blue 2019), and licenced it freely under GNU 2.0 with the hope that other academics and researchers will use it help spread awareness on IoT security and privacy issues. Meta-Blue uses the open source library pyBluez (pyBluez, 2015) based on the Bluez Bluetooth official Linux Bluetooth implementation (Bluez, 2019). The tool captures and stores MAC addresses from BLE advertising packets. The stored MAC addresses are processed and visualised using the MatPlotlib library (MatPlotLib, 2019) and illustrated in figure 3.
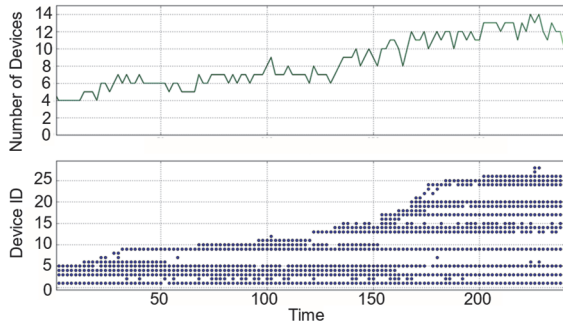
Figure 3: BLE device visualisation.

The plots in Figure 3 visualise the total number of devices visible for each iteration of the scan, and the device identification numbers associated with each device where device IDs are derived from MAC addresses. The scan was captured at the beginning of a lecture session and more than 25 unique devices where recorded over the duration of the capture. Even without the context being supplied it is evident that many devices arrived in the location in a short period of time and could be individually identified.

In order to replicate this first experiment in a public setting, scans were conducted at a railway station. This is a more likely scenario whereby individuals were not known to the researchers. The plots in Figure 4 illustrate the arrival and departure of trains at a major UK railway station in Manchester. Over 400 unique devices where recorded over a 10-minute time period.
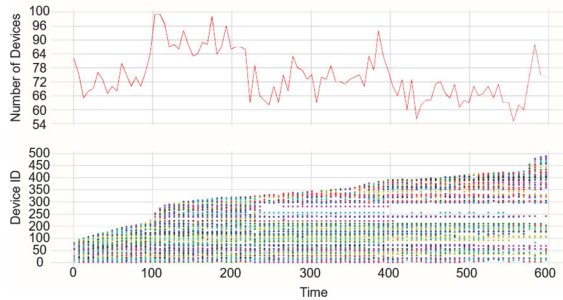

Figure 4: Train arrival and departure visualised with Meta-Blue.

From the capture alone, figure 4 demonstrates that it is possible to infer the arrival and departure of individuals by train, as indicated by the peaks and toughs in upper graph (indicating the number of devices visible at each time segment), and addition or reduction of coloured markers in the lower graph that represent individual devices. For example, between 100 and 200 seconds in the capture, train arrival is marked by an increase in total devices, and the

emergence then disappearance of device ID's in the 250 to 300 range, illustrating the transit of 50 devices/individuals past the capture device in the busy railway station. Nearly 500 devices were observed in the 10 minute (600 second) capture, illustrating the popularity of BLE devices.

To better understand the circumstances under which BLE devices persistently broadcast static metadata a separate visualisation was created using Meta-Blue and control devices. Figure 5 illustrates the results of this experiment.
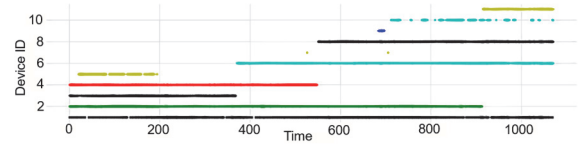

Figure 5: Control device visualisation with Meta-Blue.

Two control PFTDs were monitored for a fixed period of time. Each of the devices was periodically connected to its staging device, a mobile phone, to synchronise data. The first control device with ID 1 persistently broadcast its MAC address regardless of its connection with the staging device. The second control device with ID's 3, 6, 8 broadcast different ID's each time it was connected and disconnected from its staging device, demonstrating adherence to the Bluetooth LE Privacy standard. Devices ID's 2, 4, 5, 7, 9, 10, 11 belong to unknown devices transiting the vicinity of the experiment (demonstrating the prevalence of BLE devices). It may be the case that device ID's 2 and 11 related to the same device, however this could not be confirmed as these broadcasts were not from the control devices. Both of the control devices where periodically monitored using Meta-Blue and RaMBLE (Contextis, 2019) over a period of months. Despite power cycling and battery depletion, device ID 1 never changed the MAC address that was persistently used and broadcast for months. It is clear from the data in Figure 4, and through other evidence gathered using Meta-Blue and RaMBLE (Contextis, 2019), that many devices opt to broadcast persistent MAC addresses.

As demonstrated by this case study, it is possible to identify unique devices through presence metadata. This could be further exploited to impact an individual's security and privacy by correlating this metadata with other data.

5 CONCLUSIONS AND FUTURE WORK

BLE devices present a number of privacy and security issues, both of which are at the heart of recently proposed EU regulatory control. First and foremost, devices are powered on by default and often cannot be turned off. As such, most users are unlikely to be aware of the advertising packets frequently broadcast by their devices and their ability to identify individual devices. This data can be used to identify personal identifying information, particularly when correlated with other data. In addition, users may not know whether high standards in data encryption are being employed by the device's software or turned on at all. Moreover, in many cases it is unclear as to how or where the data is stored, or the level of encryption and/or anonymization applied to discrete or aggregated data stored in the cloud by the provider. This goes against the transparency required by regulatory bodies and may have serious implications for controllers and processors of such data.

It is important that when considering any of the scenarios discussed in this paper that MAC address collisions can occur. While they are rare, they could have a significant impact in some use cases. As such, when employed in cases where identification of an individual device has significant consequences, a secondary check should be carried out to validate the presence of a device or individual. This will be addressed in future work.

The emergence of a Bluetooth Mesh standard will enable existing BLE devices to communicate via a mesh thereby significantly increasing the range at which a device can be detected. This has implications for both the security challenges and benefits considered in this paper. Further research is also required to determine the consequences of the Bluetooth Mesh standard. As with the BLE standard the way manufacturers implement the new standard will play a key role in determining the privacy and security of users. In addition to the privacy risks outlined in this paper, the IoT poses a large threat to societal privacy and trust. A much broader range of threats to privacy are emerging as IoT matures; to give a single example; private corporations are constructing large scale unregulated surveillance networks, marketed as a feature of smart connected door bells. Aside from the recent attacks on these devices and potential for them to disrupt the Internet through Mirai type attacks. The threat to the Universal Declaration of Human Rights Article 12 (United Nations, 1948) the right to privacy, by private corporations with global reach demonstrates that further regulation or enforcement of existing legislation is required to balance the interests of the market and the privacy of the individual.

REFERENCES

Hung, M., Gartner, 2017, Leading the IoT, https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf, accessed November 2019

European Parliament, 2016 , General Data Protection Regulation, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 , accessed November 2019

Minerva, R., Biru, A., Rotondi, D., 2015 Towards a definition of the Internet of Things (IoT), IEEE Internet of Things

ePrivacy Proposal https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017PC0010&from=EN accessed January 2020

Voas, J., 2016, Network of 'Things', NIST Special Publication 900-183

www.bluetooth.com, accessed November 2019

https://www.argenox.com/a-ble-advertising-primer/ , accessed November 2019

Hilts, A., Parsons, C., Knockel, J., 2016, Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security, Open Effect Report

https://www.contextis.com/en/resources/tools/ramble-ble-app , accessed November 2019

Kambourakis, G., Kolias, A., 2017, The Mirai botnet of the IoT Zombie Armies, IEEE Military Communications Conference MILCOM.

https://www.strava.com/heatmap , accessed November 2019

Ruser, N., 2018, https://twitter.com/Nrg8000, accessed November 2019

Craddock R., Watson D., Saunders W., 2016 Generic Pattern of Life and behaviour analysis, IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)

Issoufaly, T., Tournoux, P., 2017, BLEB : Bleutooth Lowe Energy Botnet for large scale individual tracking, 1st International Conference on Next Generation Computing Applications (NextComp)

Vanhoef, M., Matte, C., Cunche, M., Cardoso, L., Piessens, F., Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms, ASIA CCS '16 Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security, 2016.

https://www.bluetooth.com/blog/introducing-bluetooth-mesh-networking/ , accessed November 2019

Kao C., Hsiano, R., Chen, P., Pan, M., 2017, A Hybrid indoor positioning for asset tracking using Bluetooth low energy and Wi-Fi, IEEE International Conference of Consumer Electronics, Taiwan (ICCE-TW)

Sager Weinstein, L., 2017, Review of the TfL WiFi Pilot, http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf , accessed November 2019

Lin, Y., Chen, M., 2015, A cloud-based Bluetooth low energy tracking system for dementia patients, 6th International Conference on Mobile Computing and Ubiquitous Networking (ICMU)

Apple Find My, https://www.apple.com/icloud/find-my/, accessed 2020

Hegarty, R., Haggerty J., https://github.com/arc2h/Meta-Blue accessed January 2020

Karulis, P., 2015, https://github.com/karulis/pybluez , accessed November 2019

Bluez Official Linux Bluetooth protocol Stack, http://www.bluez.org/ , accessed November 2019

Matplotlib, https://matplotlib.org/ , accessed November 2019

United Nations, Universal Declaration of Human Rights 1948, https://www.un.org/en/universal-declaration-human-rights/ , accessed January 2020