AmbiLoc: A year-long dataset of FM, TV and GSM fingerprints for ambient indoor localization

Andrei Popleteev
Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg, Luxembourg
E-mail: contact@popleteev.com

***Abstract*—Ambient indoor localization — an approach that leverages ambient radio signals — has been previously shown to provide promising positioning performance using the globally available infrastructure of FM, TV and cellular stations. However, the need for specialized equipment and laborious data collection constitute a high entry barrier for follow-up studies.**

This paper presents AmbiLoc — a dataset of radio signals for ambient indoor localization research. The dataset has been systematically collected in multiple testbeds, including large-scale and multi-floor buildings, over the course of one year. Due to the use of a software-defined radio receiver, raw signal samples in AmbiLoc allow extraction of arbitrary fingerprinting features. The first edition of AmbiLoc, introduced in this paper, includes received signals strength (RSS) fingerprints of FM, TV and GSM signals, along with the relevant metadata (such as weather conditions). The dataset is available online at `AmbiLoc.org`.

As the first public dataset of ambient localization signals, AmbiLoc provides an easy entry and a common reference for researchers exploring novel indoor localization methods.

***Index Terms*—Indoor positioning, database, fingerprinting, FM radio, TV, DVB-T, GSM, cellular, signals of opportunity, software-defined radio, performance evaluation, RSS, CSI, SDR.**

I. Introduction

State-of-the-art indoor positioning systems already achieve remarkable accuracy using conventional technologies such as Wi-Fi networks and Bluetooth beacons [1], [2]. However, the reliance on short-range radio technologies limits the coverage of such systems to specific buildings with the required hardware infrastructure.

Ambient indoor localization, in contrast, relies on ambient radio signals (such as TV and FM broadcasts, and cellular networks). These signals, specifically designed for indoor reception and transmitted with high power, are widely available in populated areas worldwide. The feasibility and the high accuracy of ambient indoor localization has been demonstrated in several studies [3]–[5]. However, the short duration of these proof-of-the-concept experiments provides little insight into the long-term stability of ambient indoor localization. Moreover, high-effort data acquisition and specific equipment requirements complicate further research in this area.

This paper presents AmbiLoc — the first public dataset of TV, GSM and FM radio signals for ambient indoor localization[1]. The dataset has been collected in three testbeds: a private apartment and two large-scale multi-floor buildings. Each testbed was sampled approximately bi-weekly over the course of 3 months (apartment) and 1 year (large buildings). Each measurement session also includes information about testbed state (empty or populated) and aviation-grade reports about weather conditions.

The rest of the paper provides a brief overview of state of the art in ambient indoor localization, followed by the details of data acquisition setup and dataset structure.

[1] AmbiLoc.org

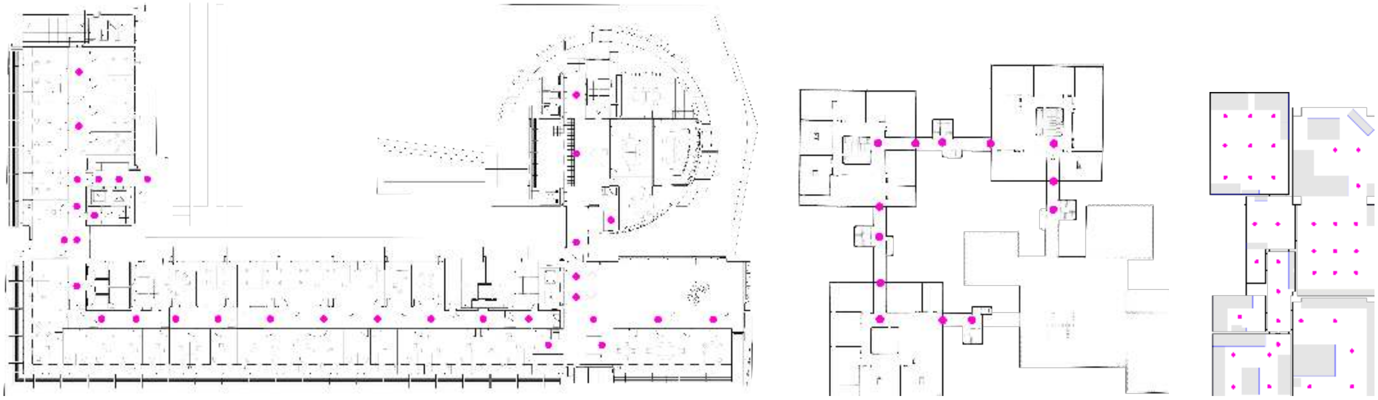
II. Related Work

Ambient indoor localization has been largely inspired by Wi-Fi based positioning. However, the two approaches have a number of important differences. Firstly, ambient radio waves propagate mainly outdoors, where they are exposed to external factors, such as weather and dynamic obstacles. Wi-Fi signals, in contrast, propagate mainly indoors and over relatively short distances. Secondly, carrier frequencies of ambient broadcasts are substantially lower than 2.4 GHz employed by Wi-Fi networks; this leads to different wave interaction with small indoor objects and lower attenuation by building materials [6]. As a result, specific characteristics of ambient indoor localization systems cannot be directly extrapolated from Wi-Fi based solutions and thus require dedicated studies.

The feasibility of ambient indoor localization has been demonstrated in several studies. The ranging approaches used in early TV positioning systems provided rather low — tens of meters — indoor accuracy, explained by the multiple signal reflections inside buildings [7]–[10]. Other publications, in turn, employed variations of the fingerprinting approach which *leverages* multipath interference (instead of suffering from it); the authors reported meter-scale accuracies achieved with FM radio [11]–[14], DVB-T [5] and GSM signals [3], [15], [16].

A common observation for the fingerprinting-based localization systems was the significant benefit of extra signal features, such as advanced physical-level FM properties [13] or additional GSM carriers [3], [16], [17]. However, such features are only available from particular receiver models, thus severely affecting the reproducibility of the experiments.

While pioneering studies demonstrated promising results over the short term (few days or weeks), long-term performance of ambient indoor localization remains an open research question. Answering it requires specialized equipment and the high effort for long-term data acquisition.



(a) Office Building (Floor 1) (b) Campus Building (Floor 1) (c) Apartment

Fig. 1. Testbed floorplans and test point locations.

TABLE I
TESTBED CHARACTERISTICS.

Testbed	Dimensions	Number of test points	Sampling period	Number of sessions	Samples collected
Office Building	100×50 m	$33 + 36 + 16$ (floors 1, 0, -2)	12 months (Jan–Dec 2016)	23	1955
Campus Building	80×80 m	$13 + 13$ (floors 1, 0)	12 months (Jan–Dec 2016)	20	520
Apartment	14×7 m	37 (floor 3)	3 months (Jan–Mar 2016)	6	222

The AmbiLoc dataset, presented in this paper, removes these barriers and provides an easy entry for new studies on ambient indoor localization. AmbiLoc complements the existing large-scale indoor localization datasets, such as UJIIndoorLoc [18] and PerfLoc [19], which contain Wi-Fi signals and inertial traces collected by off-the-shelf smartphones. In contrast, AmbiLoc contains ambient FM, TV and GSM signals, collected in multiple fixed locations over a long period.

III. APPROACH

AmbiLoc data collection has been guided by two main requirements: 1) long-term data from large-scale testbeds, and 2) future-proof data samples, not limited to one specific signal feature. This section presents an overview of our data collection approach (further details are available in [20]).

A. Testbeds

The large-scale testbeds for the study have been provided by the University of Luxembourg. One building hosted administration and research offices, while the other one featured mainly lecture halls and research labs. These testbeds are further referred to as “Offices” and “Campus”. Then, a number of fixed sampling locations (*test points*) were defined for each testbed, sparcely covering the complete floors. Additionally, in line with some previous studies [11], [21], the project included a small-scale apartment testbed with a dense placement of test points. Detailed testbed characteristics and floorplans are presented in Table I and Figure 1, respectively.

B. Data collection

Data collection has been performed using a dedicated data acquisition platform (DAQ) described in [20]. Each testbed has been sampled approximately every two weeks; all target floors were sampled within the same day. At the beginning of each session, the operator specified whether the testbed was populated or empty. To ensure consistent results, all the measurements were performed by the same person with the same equipment (see Fig. 2). To reduce the impact of operator’s body on received signals, the receiver was raised above the head level. The test points were typically visited in the same order, and receiver orientation was consistent for each point throughout the study.

Ground truth locations of the test points were initially defined with regard to local references (such as office doors). However, due to the relatively low precision of such human-defined ground truth [22] and its possible impact on localization performance [23], later on the ground truth locations were established more precisely using laser rangefinders.

C. Raw signal sampling

Normally, multi-band radio acquisition would require an array of specialized receivers. However, such receivers can only provide a limited and strictly fixed set of signal parameters, at a limited rate, and possibly missing important details (such as Cell ID of GSM channels [15]). To avoid these limitations, we employed a software-defined radio (SDR) approach.

In contrast to the traditional receivers, SDR hardware provides only a simple radio-frequency (RF) frontend, delegating most of signal processing to software. Raw RF samples from the SDR represent a complete signal “snapshot” that can be stored and processed later. This approach provides very high flexibility for extracting arbitrary signal features — not only received signal strength (RSS) or channel state information

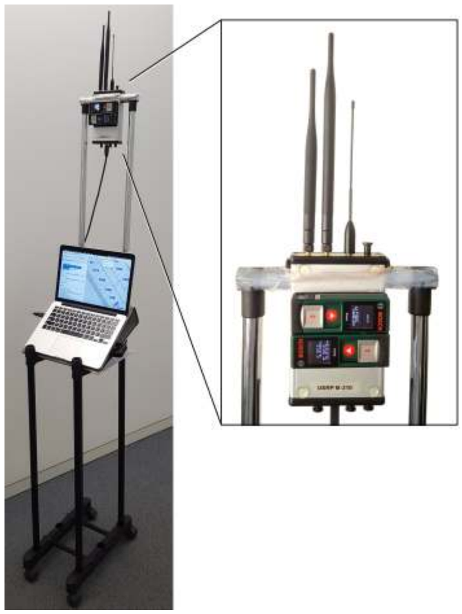


Fig. 2. Data acquisition cart with a USRP B210 receiver and ground-truth rangefinders [24].

(CSI), but also features not even foreseen at the time of signal acquisition.

In our experiments, at each test point the DAQ recorded 2-second long raw RF samples from multiple radio bands: FM radio (87.5–108.5 MHz), DVB-T (six active channels between 490 and 610 MHz), eGSM-900 (925–960 MHz), GSM-1800 (1805–1880 MHz), and Wi-Fi (13 channels in the 2.4 GHz band). The process took about 70 s per test point (including re-tuning and storage delays). Overall, the project collected about 12 TB of raw RF samples in more than 2600 measurements across all the testbeds (see Table I).

IV. Dataset Structure

The current public edition of AmbiLoc contains RSS fingerprints of ambient FM, TV and GSM stations. In order to ensure better compatibility with the wide variety of data analysis tools, the dataset is published in text-based formats. It should be noted, however, that preliminary conversion of AmbiLoc files into a suitable binary format can significantly speed up the loading procedure.

A. RSS data

RSS measurements were calculated from the raw RF samples using GNU Radio toolkit [25] with a GSM-specific module [26], and are sampled at 10 Hz, resulting in 20 fingerprints per test point and radio type.

The fingerprints are stored in a matrix form in human-readable CSV (comma-separated values) files, one file for each radio type. Each row represents one full-width RSS fingerprint, where each column corresponds to one radio channel. Channel characteristics — such as carrier frequencies for FM and TV, and Cell ID information for GSM — are provided in supplementary CSV files.

TABLE II
AmbiLoc RSS fingerprint parameters.

	Carrier frequencies	Channel width	Fingerprint size
FM radio	87.5–108.5 Mhz	100 kHz	210
GSM-900	925–960 MHz	200 kHz	175
TV (DVB-T)	498 to 602 MHz	8 MHz	6

B. Structure

RSS data files are further organized in a hierarchy of folders: `testbed` / `floor` / `session_timestamp` / `test_point`, where test points are represented by their latitude and longitude in WGS-84 coordinate system [27]. The set of test points is fixed for each floor and the coordinates remain the same across all the sessions.

C. Metadata

Finally, every data session includes a JSON-formatted file that describes the testbed, its state (populated or empty) and ground truth precision (manual or laser-based). The metadata also includes weather reports published by the nearby Luxembourg Findel airport (situated 3 to 4 km from the large testbeds and 7 km from the small one). Weather information includes temperature, humidity, precipitation and several other parameters that can be useful for analyzing weather impact on localization performance.

Further technical details, auxiliary Python scripts, and the actual dataset are available online at `AmbiLoc.org`.

V. Conclusion

The paper presented AmbiLoc, a long-term dataset of ambient indoor localization signals from TV, FM radio and GSM stations. Collected in 2016, AmbiLoc covers two large multi-floor buildings (over 12 months) and a private apartment (over 3 months). As a first public dataset of ambient indoor localization signals, AmbiLoc provides a common reference for a fair comparison of different localization approaches. Moreover, the dataset enables the researchers to avoid laborious data collection, focusing instead on development of novel ambient indoor localization methods with worldwide coverage.

The first AmbiLoc edition is dedicated to RSS fingerprinting. However, the dataset is not limited to RSS. Indeed, the RSS values are derived from raw RF signal samples, recorded by a software-defined receiver; this allows extraction of arbitrarily complex fingerprints well after the data acquisition stage. Our ongoing work focuses on studying the localization performance of the ambient approach [28], [29], as well as further processing of the collected signal samples (in particular, from the Wi-Fi and GSM-1800 bands) and extending AmbiLoc with new location-sensitive features.

VI. Acknowledgments

This work was supported by the National Research Fund of Luxembourg (INDOORS project, grant number C14/IS/8311593). The author gratefully acknowledges Dominique Kogue's assistance with GSM data processing.

REFERENCES

[1] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[2] R. Faragher and R. Harle, "Location fingerprinting with Bluetooth Low Energy beacons," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, 2015.

[3] A. Varshavsky, E. de Lara, J. Hightower, A. LaMarca, and V. Otsason, "GSM indoor localization," *Pervasive and Mobile Computing*, vol. 3, no. 6, pp. 698–720, December 2007.

[4] A. Popleteev, "Indoor positioning using FM radio signals," Ph.D. dissertation, University of Trento, 2011.

[5] R. Carvalho, S.-H. Yang, Y.-H. Ho, and L.-J. Chen, "Indoor localization using FM and DVB-T signals," in *Proc. 13th IEEE Annual Consumer Communications Networking Conf. (CCNC)*, Jan. 2016, pp. 862–867.

[6] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, 1993.

[7] M. Rabinowitz and J. J. Spilker Jr, "A new positioning system using television synchronization signals," *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 51–61, 2005.

[8] D. Serant, P. Thevenon, M.-L. Boucheret, O. Julien, C. Macabiau, S. Corazza, M. Dervin, and L. Ries, "Development and validation of an OFDM/DVB-T sensor for positioning," in *Proceedings of IEEE PLANS-2010*. IEEE, 2010, pp. 988–1001.

[9] J. Huang, L. Lo Presti, and R. Garello, "Digital video broadcast-terrestrial (DVB-T) single frequency networks positioning in dynamic scenarios," *Sensors*, vol. 13, no. 8, pp. 10 191–10 218, 2013.

[10] D. Serant, O. Julien, P. Thevenon, L. Ries, and M. Dervin, "Testing OFDM-based positioning using the digital TV signals," in *Proceedings of EUSIPCO-2012*. IEEE, 2012, pp. 539–543.

[11] A. Matic, A. Popleteev, V. Osmani, and O. Mayora, "FM radio for indoor localization with spontaneous recalibration," *Journal of Pervasive and Mobile Computing*, vol. 6, pp. 642–656, 2010.

[12] A. Popleteev, V. Osmani, and O. Mayora, "Investigation of indoor localization with ambient FM radio stations," in *Proceedings of PerCom-2012*. IEEE, 2012, pp. 171–179.

[13] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "FM-based indoor localization," in *Proceedings of MobiSys 2012*, 2012.

[14] S. Yoon, K. Lee, and I. Rhee, "FM-based indoor localization via automatic fingerprint DB construction and matching," in *Proceeding of MobySys-2013*. ACM, 2013, pp. 207–220.

[15] Y. Oussar, I. Ahriz, B. Denby, and G. Dreyfus, "Indoor localization based on cellular telephony RSSI fingerprints containing very large numbers of carriers," *EURASIP Journal on Wireless Communications and Networking*, vol. 81, 2011.

[16] R. Górak, M. Luckner, M. Okulewicz, J. Porter-Sobieraj, and P. Wawrzyniak, "Indoor localisation based on GSM signals: Multistorey building study," *Mobile Information Systems*, vol. 2016, 2016.

[17] Y. Tian, "Practical indoor localization system using GSM fingerprints and embedded sensors," Ph.D. dissertation, Université Pierre et Marie Curie, 2015.

[18] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proceedings of IPIN-2014*. IEEE, 2014, pp. 261–270.

[19] N. Moayeri, M. O. Ergin, F. Lemic, V. Handziski, and A. Wolisz, "PerfLoc (Part 1): An extensive data repository for development of smartphone indoor localization apps," in *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on*. IEEE, 2016, pp. 1–7.

[20] A. Popleteev, "Indoor positioning using ambient radio signals: Data acquisition platform for a long-term study," in *Proceedings of the 13th IEEE Workshop on Positioning, Navigation and Communications (WPNC-2016)*, 2016.

[21] V. Moghtadaiee, A. Dempster, and S. Lim, "Indoor localization using FM radio signals: A fingerprinting approach," in *Proceedings of IPIN-2011*, 2011.

[22] A. Popleteev, "HIPS: Human-based indoor positioning system," in *Proceedings of the 7th International Conference on Indoor Positioning and Indoor Navigation (IPIN-2016)*. IEEE, 4–7 Oct. 2016.

[23] ——, "Poster: Impact of ground truth errors on Wi-Fi localization accuracy," in *Proceedings of MobiSys-2017*, 2017.

[24] Ettus Research, "USRP B-210 USB software defined radio (SDR)," Online: https://www.ettus.com/product/details/UB210-KIT.

[25] "GNU radio toolkit," http://gnuradio.org.

[26] P. Krysik, "Gnuradio blocks and tools for receiving GSM transmissions," https://github.com/ptrkrysik/gr-gsm.

[27] "Department of Defense World Geodetic System 1984: Its definition and relationships with local geodetic systems," Online: http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html, National Imagery and Mapping Agency, Tech. Rep. TR8350.2, 1997.

[28] A. Popleteev, "Indoor localization using ambient FM radio RSS fingerprinting: A 9-month study." in *Proceedings of the 17th IEEE International Conference on Computer and Information Technology (IEEE CIT-2017)*, Helsinki, Finland, Aug. 2017.

[29] ——, "Please stand by: TV-based indoor localization," in *Proceedings of the 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC-2017)*, 2017.