2016 IEEE 27th International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC): Services Applications and Business

PerfLoc (Part 1): An Extensive Data Repository for Development of Smartphone Indoor Localization Apps

Nader Moayeri*, Mustafa Onur Ergin†, Filip Lemic†, Vlado Handziski†, and Adam Wolisz†
*National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA
†Telecommunication Networks Group (TKN), Technische Universität Berlin, Berlin, Germany
Email: moayeri@nist.gov, {ergin,lemic,handziski,wolisz}@tkn.tu-berlin.de

***Abstract*—This paper describes a major effort to collect smartphone data useful for indoor localization. Data has been collected with four Android phones according to numerous scenarios in each of four large buildings. The data includes time-stamped traces of various environmental, position, and motion sensors available on a smartphone, Wi-Fi and cellular signal strengths, and GPS fixes, whenever available. Quantitative evidence is presented to validate the collected data and attest to its quality. This unique, extensive data repository is made available to the R&D community through the PerfLoc web portal to facilitate development of smartphone indoor localization apps and to enable performance evaluation of such apps according to the ISO/IEC 18305 international standard in the near future.**

I. Introduction

The Global Positioning System (GPS) has been a phenomenal success with applications in a wide range of domains, but it does not work inside buildings / structures and in urban canyons. The next frontier is to provide localization and tracking capability indoor as a key technology enabler for Location Based Services (LBS), which is anticipated to be a multi-billion dollar market. Indoor localization and tracking has applications in many areas, including emergency response for better coordination of operations and to save lives of first responders and civilians; E-911; military operations, such as search and clear operations and prevention of friendly fire; tracking in underground coal mines, particularly in the aftermath of explosions and roof collapses; asset and personnel tracking in warehouse, hospitals, and factories; tracking of children on school grounds and the elderly; offender tracking; navigation in museums, shopping malls, and large office buildings; urban search and rescue in the aftermath of natural / man-made disasters; and many applications related to the Internet of Things (IoT).

Yet, lack of standardized testing has been an impediment to the wide adoption and deployment of indoor Localization and Tracking Systems (LTSs). It is not possible to compare the performance of various systems presented in academic conferences or developed by industry, because they are evaluated in different environments and according to different and typically inadequate testing criteria and methodologies. This has made it difficult to set minimum performance requirements[1] for indoor LTSs. Consequently, the user community has often been unable to ascertain whether a given indoor LTS meets its requirements. There has been a strong demand by the user community for development of standardized testing procedures for indoor LTSs, which would also help the industry to improve the performance and effectiveness of their indoor LTS products.

In response to this demand, NIST led the development of the international standard ISO/IEC 18305, Test and evaluation of localization and tracking systems [1]. The primary focus of ISO/IEC 18305 is on indoor localization, which is a much harder problem than its outdoor counterpart, for which GPS and to a lesser extent the use of terrestrial cellular technology are the dominant solutions. The standard deals with a challenging test and measurement problem, because the performance of such systems is affected by a wide range of factors, including construction material, size and floor plans of buildings; mobility mode of the entity to be localized and tracked (stationary, walking, running, sidestepping, walking backwards, crawling on the floor, transportation on a cart or forklift, transportation in elevators, etc.); availability of coordinates of the boundary (footprint) of buildings; availability of floor plans and heights of different floors of buildings. While it might be possible to test the "components" of an LTS in a laboratory setting, the proper way to test the "system" is to do so in several large buildings, including high rises and subterranean structures, according to a variety of mobility scenarios. In addition, it is important to determine how many entities the system can localize and track. LTSs can vary significantly in terms of the assumptions under which they can operate, cost, size, weight, battery life, electromagnetic compatibility, intrinsic safety, etc. Therefore, one has to be careful when comparing the performance of different LTSs to ensure fairness. Tailoring the testing procedure to the LTS is not scalable, because there are many LTSs available. ISO/IEC 18305 treats the LTS under test as a black box and yet it provides a comprehensive testing methodology that adequately tests LTSs.

This paper focuses on smartphone indoor localization apps, which can use only the sensors available on a smartphone and the Radio Frequency (RF) signals that a smartphone can receive. In contrast, a general LTS can use other sensors and RF technologies, such as Ultra Wideband (UWB) ranging, angle of arrival estimation, and LiDAR. However, with billions of smartphones in use around the world, the smartphone platform is very important. The main objective of this effort is to create a level playing field for comparison of indoor localization apps. We are doing this by (i) making available to

The work of the first author has been supported by the Public Safety Communications Research (PSCR) Division at NIST.

[1]For example, fire departments may wish to have an indoor localization capability in burning buildings with 3 m average accuracy. Similarly, the Federal Communications Commission (FCC) in the US requires the average indoor location accuracy for E-911 calls to be x meters y% of the time.

the R&D community a rich repository of smartphone sensor data, RF signal strength data, and GPS fixes collected based on the guidance provided by ISO/IEC 18305 and (ii) developing a web portal that uses ISO/IEC 18305 to automatically evaluate the performance of indoor localization apps developed based on the data repository and publish the results.

The rest of this paper is organized as follows. Section II presents an overview of related work. We present different aspects of our campaign and the properties of the collected data in Section III. Section IV is on validation of the data we collected. Concluding remarks are given in Section V.

II. Related Work

This section provides an overview of two relevant directions in performance evaluation of indoor localization solutions.

A. Indoor Localization Competitions

Lessons learned during the 2014 IPSN / Microsoft Indoor Localization Competition are reported in [2]. The competition used one scenario to evaluate a broad set of solutions, including RF, magnetic, light, and ultrasound-based systems. EvAAL is another popular series of competitions focused on indoor localization solutions for assisted living [3]. [4] presents the results of an indoor localization competition focused on interference robustness of RF-based localization solutions. Indoor localization competitions are attractive, because they provide the possibility of evaluating and objectively comparing different localization solutions in the same environment and similar conditions. However, such competitions are rare due to their labor, time, and cost intensity. Furthermore, although the evaluation environment is the same, not all solutions can be evaluated at the same time. Hence, the temporal variability of environmental conditions reduces the objectivity of the results.

B. Usage of Raw Data Traces

Use of publicly available data traces is a promising solution for mitigating temporal variability of conditions, which is hardly avoidable in indoor localization competitions. There have been a few efforts to virtualize certain aspects of experimental evaluation of localization solutions. VirTIL testbed [5] is focused on the evaluation of RF range-based indoor localization algorithms by providing range measurements made throughout the evaluation area. The EU EVARILOS project's efforts [6] yielded unprocessed low-level RF data, such as Received Signal Strength Indicator (RSSI) and Time of Flight (ToF) measurements, that are used in a wide range of indoor localization solutions, including fingerprinting, hybrid and proximity-based solutions. This paper builds upon that work and focuses on the extensive set of sensors available in today's smartphones, including not only RF-based ones, but also environmental, position, and motion sensors. Moreover, we significantly expand the scope of the data traces by using several evaluation areas and testing scenarios.

III. Data Collection

This section consists of three parts. First, we describe the environment and the scenarios, selected based on guidance from ISO/IEC 18305, we used for data collection. Second, we provide an overview of the collected data. Third, we provide a synopsis of the collected data.

A. Environment and Scenarios

We selected four buildings for data collection. One was an office building, two were industrial shop and warehouse types of buildings, and the fourth was a subterranean structure. Unfortunately, we did not have access to a high-rise building or a single-family house, as recommended by ISO/IEC 18305 in addition to the above types of buildings. The total space covered by these buildings was about 30,000 m^2.

We instrumented these buildings with more than 900 test points, henceforth called dots, installed on the floors. Each dot is a disk of diameter about 3 cm with its center marked for subsequent surveying. The locations of these dots were precisely determined by a surveying contractor. Therefore, the ground truth locations of the dots are known to NIST.

In an effort to capture the differences in qualities of the sensors and RF circuitry in smartphones, we used four Android[2] smartphones for data collection. These smartphones were LG G4 (LG), Motorola Nexus 6 (NX), OnePlus 2 (OP), and Samsung Galaxy S6 (SG). Since we did not want to repeat each data collection scenario four times, once for each smartphone, we devised a mechanism to collect data with all four phones simultaneously. We used armbands to attach two phones to each arm of the person collecting data, as shown in Figure 1. On each arm, one phone faced forward and the other faced to the side away from the person. We connected four cables in parallel to a push-button switch. The other side of the cables were connected to the audio jacks of the four phones. We used this mechanism to send a signal to the phones to create a timestamp in each of them whenever the person collecting data was on top of a dot.

The data that we collected is described in detail in the next section. We collected two types of data, one for training and the other for testing purposes. The data in each type is timestamped, but the training data is also annotated with the ground truth locations of the dots at which the push-button switch was pressed. The training data would allow the app developers to check how well their apps are performing by comparing the location estimates provided by their apps at time instances when the push-button switch was pressed with the ground truth locations at those time instances. In the case of the test data, we do not provide the ground truth locations at time instances at which the push-button switch was pressed. Instead, we ask each app developer to upload the location estimates provided by the app at those time instances on the PerfLoc web portal so that the portal can evaluate the app's performance. Naturally, a developer would do this step when he/she is convinced that his/her app is as good as it can be.

[2]DISCLAIMER- Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

We collected data using a subset of the 14 Test & Evaluation (T&E) scenarios described in ISO/IEC 18305. We did not use all the scenarios, because some did not apply to our data collection campaign. For example, two scenarios in ISO/IEC 18305 call for multiple people to collect data simultaneously. This would be useful if the localization device has peer-to-peer ranging capability or we wish to test the Medium Access Control (MAC) layer issues and determine how many localization devices can be localized by the system simultaneously. Smartphones do not have a peer-to-peer ranging capability and we are focusing on indoor localization apps that passively listen to RF signals of interest, and hence there are no MAC layer issues. Including the training data, we collected data over 38 T&E scenarios in the four buildings. Each scenario called for the person collecting the data to start at a given location outside a building before entering the building and following a predetermined course while pressing the push-button switch at select dots on the course. The scenarios involved different modes of mobility: 1) walking to a dot and stopping for 3 s before moving to the next dot; 2) walking continuously and without any pause throughout the course; 3) running / walking backwards / sidestepping / crawling part of the course; 4) "transporting" the four phones on a pushcart; 5) using elevators, as opposed to stairs, to change floors; 6) leaving the building a few times during a scenario and then reentering through the same door or another.

B. Types of Data Collected

For each scenario in each building we collected six types of data on each smartphone, namely, Wi-Fi, Cellular, GPS, Dots, Sensors, and Metadata, as described below. This data is stored as one or more Google Protocol Buffer Messages [7] in a separate file for each data type

The collected data is composed of:

1) Wi-Fi data: Signal strengths measured from Wi-Fi access points (APs) in range and other information provided by the APs operating at 2.4 and 5 GHz channels. The Wi-Fi scans were performed back-to-back. The frequency of the scans depended on the Wi-Fi chip's firmware and the Wi-Fi driver implementation on a given smartphone. Since these implementations are provided by the manufacturers, the duration of a Wi-Fi scan varied from device to device. Android platform's Wi-Fi connectivity API supports passive scanning only. Scan results were saved as a single protocol buffer message in repeated fields for each AP.

2) Cellular data: Identity information and signal strengths measured from cellular towers in range. The Android operating system continuously scans for cellular signals and a scan cannot be triggered by a user-level application. Therefore, our application periodically requests all observed cell information from the operating system at a 1 Hz update rate and saves the detected cellular tower related information into the file system as repeated fields of the relevant protocol buffer message.

3) GPS data: Detected geophysical position using GPS. During the data collection campaign, whenever the smartphone got a location fix, the GPS information along with the accuracy of that fix was saved as a separate protocol buffer message. We used "fine location" information, which is derived directly from the GPS signals. Since the measurements were performed indoors most of the time, it was not possible to receive the GPS signal the whole time. If the device could not derive a location from the GPS signals, then no data was stored.

4) Dots: Timestamps at dots visited during a scenario. To make sense of the collected data, it is important to know where and when the data was collected. For each dot visited during a scenario, we store a dot-index and a timestamp. The dot-indices are simply successive non-negative integers with 0 reserved for the starting point of the scenario, which may not be a dot, and 1 through n used for the n dots visited during the scenario after the starting location. For each dot-index, there is a timestamp that designates the time when that location was visited. For the scenario that required waiting for three seconds at each of the dots 1 through n, we stored two dot-indices and two timestamps. Specifically, for the i^{th} dot, we used dot-indices $2i-1$ and $2i$ for the time instances we reached that dot and we left it, respectively. Each dot is represented in a separate protocol buffer message with its index and timestamp.

5) Sensors: Built-in environmental, position, and motion sensors that are found in smartphones. The Android platform defines and supports a number of sensors, which are either hardware-based or software-based. Hardware-based sensors, such as accelerometer, magnetometer, or light, are physical components in the device. Software-based sensors, such as linear acceleration and gravity, are virtual sensors that derive their values from one or more hardware-based sensors.

The Android platform does not specify a standard sensor configuration. Hence, the manufacturers can choose any set of sensors to install in their devices. In our measurements we collected data from all of the sensors that were available on any given device. The list of Android-defined sensors, their descriptions, and whether they are hardware- or software-based are given in Table I. Some sensors are "uncalibrated" versions of others with the same name. These sensors provide additional raw values along with some bias. These types of sensors can be useful when an application conducts its own sensor fusion. More information on uncalibrated sensors can be found in [9]. A smartphone may contain device-specific, non-standard sensors that are not defined in the above list. The information collected by any of the sensors, including the non-standard ones, is saved as protocol buffer messages. Some devices also implement software-based one-shot composite sensors, such as glance gesture, pick up gesture, significant motion, and wake up gesture. These sensors, also called trigger sensors, are used for end-user convenience, such as to briefly turn the screen on or to mimic press of the power button. Since these convenience features do not provide any clear benefits to LTSs, we did not implement support for them, but they may appear in the collected data if they are triggered. Android's sensor framework uses a standard 3-axis coordinate system depicted in Figure 2. It is important to note that the axes of this coordinate system do not change or swap if the orientation of the device changes. For example, the z-axis remains pointing

TABLE I: List of Android sensors [8]

Sensor	Description	Type	Sensor	Description	Type
AMBIENT_ TEMPERATURE	Ambient air temperature	Hardware	LIGHT	Luminance	Hardware
PRESSURE	Ambient air pressure	Hardware	RELATIVE_ HUMIDITY	Ambient relative humidity	Hardware
ACCELEROMETER	Acceleration force along the x,y,z axes	Hardware	GRAVITY	Force of gravity along the x,y,z axes	Hardware or Software
GYROSCOPE	Rate of rotation around the x,y,z axes	Hardware	GYROSCOPE_ UNCALIBRATED	Rate of rotation (without drift compensation) around the x,y,z axes	Software
LINEAR_ ACCELERATION	Acceleration force along the x,y,z axes (excluding gravity)	Hardware or Software	ROTATION_VECTOR	Rotation vector component along the x,y,z axes and Scalar component of the rotation vector	Hardware or Software
STEP_COUNTER	Number of steps taken by the user since the last reboot when the sensor was activated.	Software	GAME_ROTATION_ VECTOR	Rotation vector component along the x,y,z axes	Software
GEOMAGNETIC_ ROTATION_VECTOR	Rotation vector component along the x,y,z axes	Software	MAGNETIC_FIELD	Geomagnetic field strength along the x,y,z axes.	Hardware
MAGNETIC_FIELD_ UNCALIBRATED	Geomagnetic field strength (without hard iron calibration) and Iron bias estimations along the x,y,z axes	Software	ORIENTATION	Azimuth, Pitch and Roll	Software
PROXIMITY	Distance from object	Hardware			

outwards of the screen even when the person collecting data is crawling on the floor instead of walking. Each sensor component has a different reporting frequency. When the app registers to a sensor at the fastest rate, the reporting frequency can be as fast as 250 Hz. Unfortunately, asking for the fastest possible rate sometimes causes individual sensors to be deprioritized and starved. For example, the Samsung smartphone reported far fewer readings from the pressure sensor when all the sensors were registered at the fastest rate. To avoid loss of important data, we programmatically registered to the sensors at a 100 Hz rate, which is commonly selected in research on human motion recognition using motion sensors like accelerometers. This gave all sensors enough time to report their values. Temperature, light, proximity, and humidity sensors, and the step counter generate values only if their last measured values have changed.

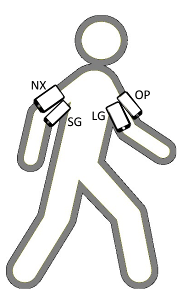


Figure 1: Positioning of the devices on the test subject's body

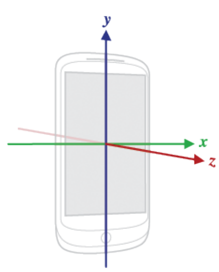


Figure 2: Device-relative coordinate system used by the Sensor API [9]

6) Metadata: The metadata includes the building ID, scenario ID, and measurement device's manufacturer, model, ID, brand, etc. It also includes the initial barometer value, if the smartphone has one, by averaging the first 5 pressure sensor measurements at the beginning of a data collection scenario. Knowing the initial pressure value can help the indoor localization app to detect elevation changes due to movement from one floor of the building to another. The list of sensors that a particular smartphone is equipped with along with their properties is also included in the metadata. For data collection in each building one protocol buffer message for the metadata in each smartphone is generated.

C. A Synopsis of the Collected Data

The number of scenarios used (the time it took to collect the data) in Buildings 1-4 were 11 ($\sim$4.1 hours), 10 ($\sim$5.3 hours), 9 ($\sim$4.4 hours), and 8 ($\sim$1.8 hours), respectively, resulting in a total of 38 scenarios and roughly 15.6 hours of raw data traces for each device. Tables II-V provide an overview of Wi-Fi, cellular, GPS, and sensor data traces, respectively, in two scenarios for each of the four buildings.

The reader can extract similar information for the remaining 30 scenarios by downloading data traces from the PerfLoc web portal. Certain observations can be made about the data in Table II. First, the average durations of a Wi-Fi scan over the 8 scenarios for the LG G4, Motorola Nexus 6, OnePlus 2, and Samsung Galaxy S6 devices (ordered alphabetically) are 3238, 916, 2351, and 3351 ms, respectively. Second, the LG G4 device detects far fewer Wi-Fi APs than the other three devices that detect similar numbers of APs. This is due to the fact that the default configuration of the Wi-Fi scanning procedure in the LG G4 device does not report hidden Wi-Fi APs, i.e. those without the SSID parameter defined. Indeed, if we filter the hidden APs for the other devices for Scenario 1 in Building 1, the number of unique Wi-Fi APs detected would be 56 (unchanged), 61, 59, and 61, respectively. The same observation is made when hidden APs are filtered out in the other scenarios. Additional small variability in the number of detected APs is due to slight differences in the locations and movement patterns of devices, as worn by the person collecting the data, which results in different levels of signal attenuation and shadowing. This is consistent with a previous report of RSSI not being a fully stable feature of Wi-Fi signals [10].

As can be seen from Table III, the number of cellular scans is roughly the same as the duration of a scenario in seconds, except for Building 1, where cellular coverage was far worse than the other three buildings. This is consistent with the 1 Hz cellular scan rate mentioned in Subsection III.B. Note that we did not log empty cellular readings, which explains the slight differences in the duration of scenarios. In addition, the LG G4 and Samsung Galaxy S6 devices had SIM cards and subscriptions to CDMA (C) and LTE (L) data services. Consequently, the data traces for these devices consist of CDMA and LTE signal strengths, although WCDMA (W)

TABLE II: Selected features of the Wi-Fi data traces

Building/ Scenario	Number of Wi-Fi scans				Number of unique Wi-Fi APs				Duration [s]			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG
B1/S1	245	953	356	248	56	150	143	158	836	843	835	837
B1/S2	370	1397	524	358	24	70	70	67	1212	1219	1212	1212
B2/S1	741	2442	983	707	237	516	611	625	2349	2348	2352	2347
B2/S2	792	2667	1057	759	222	515	574	630	2518	2527	2519	2522
B3/S1	696	2545	985	677	70	166	178	190	2288	2298	2288	2291
B3/S2	460	1710	650	447	64	144	141	164	1515	1514	1518	1511
B4/S1	360	1283	496	347	39	104	118	105	1158	1164	1156	1159
B4/S2	207	731	285	198	30	80	77	96	660	661	664	658

TABLE III: Selected features of the cellular data traces

Building/ Scenario	Number of cellular scans				Number of unique cellular BSs				Duration [s]			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG
B1/S1	130	132	166	119	C:3,G:1,L:6	G:1,W:7	G:1,W:6	C:2,L:6	623	136	637	121
B1/S2	47	199	208	201	C:2,L:5	G:1,W:5	G:1,W:7	C:3,G:1,L:5	1216	1221	1215	1219
B2/S1	2311	2273	2327	2301	C:5	W:5	W:6	C:9,L:9	2352	2349	2354	2352
B2/S2	2455	2443	2491	2468	C:6	G:1,W:7	G:1,W:7	C:9,L:8	2523	2528	2522	2528
B3/S1	2108	2215	2263	2043	C:6,W:2	G:1,W:10	G:1,W:13	C:7,L:13,W:1	2293	2299	2292	2294
B3/S2	1413	1469	1505	1463	C:8,G:1,W:3	W:2	W:3	C:7,L:6	1518	1515	1521	1517
B4/S1	1149	1121	1146	1141	C:4,L:22	W:12	W:11	C:5,L:13	1161	1166	1159	1163
B4/S2	658	641	657	651	C:4,L:13	W:4	G:1,W:7	C:4,L:10	664	662	666	663

and GSM (G) readings occur also. For the other two devices, the stored signal strength data are of the WCDMA and GSM types only.

TABLE IV: Selected features of the GPS data traces

Building/ Scenario	Number of GPS readings				Duration [s]			
	LG	NX	OP	SG	LG	NX	OP	SG
B1 / S1	20	12	25	10	18	85	23	13
B1 / S2	54	4	13	7	76	2	11	7
B2 / S1	619	65	179	242	1784	2076	2090	2335
B2 / S2	669	217	269	350	1915	1804	1843	2175
B3 / S1	585	187	180	345	2197	2077	928	2133
B3 / S2	33	76	219	140	31	191	228	195
B4 / S1	890	583	889	790	1150	841	1111	1158
B4 / S2	568	448	561	434	599	520	601	548

Table IV shows that the GPS signal was not always available during a scenario. Just as in the case of the cellular signal, the availability of the GPS signal was far sparser in Building 1 compared to the other 3 buildings. The data shows significant variability across the four devices in both the duration of time the GPS signal was available and the number of GPS readings. In addition, this variability appears to be random, and hence one cannot say that the GPS receiver in any device was better than those in the other devices.

Finally, the numbers of readings from select sensors deemed to be more useful for indoor localization purposes are provided in Table V. These sensors are light, pressure, accelerometer / gyroscope, magnetometer, and step detector (see [2] and the references therein). We have grouped the accelerometer and gyroscope together, because the numbers of readings from these sensors were always within one of each other. Note that the numbers of magnetometer readings for the OnePlus 2 and Samsung Galaxy S6 devices are always within one of the numbers of readings from the respective accelerometer / gyroscope. In addition, these numbers and the numbers of accelerometer / gyroscope readings for the LG G4 and Motoroal Nexus 6 devices are roughly hundred times the duration of the respective scenarios in seconds, which is consistent with the 100 Hz sensor sampling rate mentioned in Subsection III.B. The variation in the numbers of readings from a given sensor across the devices is primarily due to the differences in susceptibility thresholds of the sensors. The numbers of readings from the light and pressure sensors and the step detector are far fewer. Note that OnePlus 2 does not have a pressure sensor. Peculiarly, this device did not detect any steps in Scenarios 1 and 2 in Building 4!

IV. Data Validation

We took certain measures prior to the start of our extensive data collection campaign to ensure the data we were getting from the phones was sound and made sense. For example, we walked in a building following a predetermined course while holding a phone in hand horizontally and pointing to the direction of movement. We accelerated and decelerated in certain places, made turns, went up and down the stairs, violently shook the phone at certain times, turned the lights off and on in one corridor, and pressed the push-button switch whenever we went over a dot. We knew on which accelerometer and gyroscope axes to expect activity at what times, and we verified that that was indeed the case. The data from the pressure and light sensors made sense, and the number of timestamps generated by the switch was the same as the number of dots we visited. Due to lack of space, we do not present the details of that experiment.

However, the problem of validating the extensive data we collected with multiple devices is a lot more involved than that. A major challenge is to figure out what questions to ask and which tests to perform before even deciding whether the questions are adequately answered or the data passed the tests. We suspect more can be done than what is presented next in this section, but at least we address two issues. One is how periodic different sensor and RF signal strength data is. We did not have any influence on how the OS delivers the sensor data to the application level, and we could only suggest a delivery rate without any guarantee on the inter-sample time of individual sensors. This resulted in potentially irregularly sampled time series. This issue is important, because dealing with periodic data would be much easier for the researchers who would use our data to develop indoor localization apps than dealing with asynchronous data. Another question is

TABLE V: Number of sensor readings for selected sensors

Building/ Scenario	Light				Pressure				Accelerometer / gyroscope				Magnetometer				Step detector			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	OP	SG
B1/S1	3691	2285	3215	4708	64252	27276	0	4708	85294	83911	83284	84748	77467	63648	83285	84748	996	979	949	911
B1/S2	4760	2302	4198	6812	93371	39362	0	6812	123678	121265	120811	122620	112771	91821	120810	122620	788	712	714	196
B2/S1	14673	9318	14491	13135	181186	75854	0	13135	239020	233363	233834	236483	217952	176991	233835	236483	1915	2019	1660	1750
B2/S2	14957	9358	15277	14117	194411	81388	0	14117	256346	251116	250536	254168	234384	189892	250536	254168	1451	1349	1262	802
B3/S1	15515	9780	15966	12823	175718	74232	0	12823	232945	228348	227674	230841	211903	173211	227674	230841	2722	2841	2544	2711
B3/S2	9463	4826	9736	8478	116748	48902	0	8478	154316	150562	151077	152636	140812	114088	151078	152636	1030	1004	974	506
B4/S1	8493	4963	8654	6503	89041	37620	0	6503	118022	115789	115266	117063	107381	87780	115266	117063	1231	1155	0	988
B4/S2	4225	2473	5105	3710	51027	21362	0	3710	67613	65751	66230	66772	61696	49848	66229	66772	527	529	0	330

TABLE VI: Summary of inter-sample times [mean value±standard deviation]

Building/ Scenario	Wi-Fi [s]				Accelerometer / Gyroscope [ms]				Magnetometer [ms]		Cellular [s]			
	LG	NX	OP	SG	LG	NX	OP	SG	LG	NX	LG	NX	OP	SG
B1 / S1	3.3±0.2	0.9±0.0	2.4±0.5	3.4±0.1	9.8±3.7	10.1±6.1	10.1±3.8	9.9±4.5	10.8±4.6	13.3±6.2	4.8±43.2	1.0±0.0	3.9±36.5	1.0±0.0
B1 / S2	3.3±0.1	0.9±0.0	2.3±0.0	3.4±0.1	9.8±3.5	10.1±5.8	10.1±3.7	9.9±4.6	10.8±4.4	13.3±5.9	26.5±170.6	6.1±59.4	5.9±57.0	6.1±49.0
B2 / S1	3.2±0.1	1.0±0.1	2.4±0.4	3.3±0.1	9.8±3.6	10.1±7.3	10.1±4.2	9.9±5.4	10.8±4.5	13.3±7.7	1.0±0.1	1.0±0.0	1.0±0.0	1.0±0.0
B2 / S2	3.2±0.1	0.9±0.1	2.4±0.4	3.3±0.1	9.8±3.8	10.1±7.4	10.1±4.1	9.9±5.6	10.8±4.7	13.2±7.9	1.0±0.2	1.0±0.0	1.0±0.0	1.0±0.0
B3 / S1	3.3±0.1	0.9±0.1	2.3±0.0	3.4±0.1	9.8±4.1	10.1±6.1	10.1±4.0	9.9±5.1	10.8±4.9	13.3±6.2	1.1±1.3	1.0±0.0	1.0±0.0	1.1±1.6
B3 / S2	3.3±0.1	0.9±0.0	2.3±0.4	3.4±0.0	9.8±4.8	10.1±5.7	10.1±4.0	9.9±4.9	10.8±5.6	13.3±5.8	1.1±1.1	1.0±0.0	1.0±0.0	1.0±0.0
B4 / S1	3.2±0.0	0.9±0.0	2.3±0.0	3.4±0.0	9.8±3.6	10.1±6.3	10.1±4.1	9.9±4.6	10.8±4.5	13.3±6.1	1.0±0.0	1.0±0.0	1.0±0.0	1.0±0.0
B4 / S2	3.2±0.0	0.9±0.0	2.3±0.0	3.3±0.0	9.8±3.4	10.1±5.9	10.1±4.1	9.9±4.6	10.8±4.3	13.3±6.2	1.0±0.0	1.0±0.0	1.0±0.0	1.0±0.0

TABLE VII: Summary of Spearman's correlation coefficient ρ for selected sensor types

Building/ Scenario	Wi-Fi						Accelerometer					
	LG-NX	LG-OP	LG-SG	NX-OP	NX-SG	OP-SG	LG-NX	LG-OP	LG-SG	NX-OP	NX-SG	OP-SG
B1 / S1	0.876	0.883	0.884	0.834	0.904	0.835	0.854	0.875	0.845	0.834	0.836	0.804
B1 / S2	0.879	0.876	0.874	0.834	0.902	0.884	0.786	0.832	0.821	0.883	0.856	0.832
B2 / S1	0.856	0.885	0.898	0.874	0.901	0.834	0.864	0.823	0.882	0.843	0.884	0.832
B2 / S2	0.867	0.902	0.829	0.871	0.890	0.897	0.864	0.812	0.845	0.812	0.850	0.842
B3 / S1	0.902	0.875	0.908	0.843	0.876	0.865	0.874	0.831	0.810	0.831	0.845	0.791
B3 / S2	0.896	0.880	0.879	0.800	0.891	0.841	0.831	0.829	0.871	0.811	0.851	0.803
B4 / S1	0.852	0.847	0.902	0.886	0.904	0.906	0.811	0.832	0.814	0.824	0.823	0.841
B4 / S2	0.863	0.912	0.901	0.876	0.901	0.877	0.822	0.851	0.812	0.767	0.855	0.842

whether the data collected with the four devices is "similar", with the caveat that for certain sensors one should not expect similarity due to the differences in where the devices were worn on the person who collected the data and the devices' movements. If we knew the ground truth for all the sensor and RF signal strength data, then we could check whether the data collected with any given device was sufficiently close to the ground truth. In the absence of such ground truth data, all we can do is to decide, for each sensor or RF signal strength, which devices generated similar data. Such tests would reveal, for example, whether three devices generated similar data but the fourth one was stuck at a value due to sensor malfunction or it generated unrealistic erratic or random data because it was not properly calibrated. There can also be programmatic errors in the code that collects and stores data. Even if none of these problems exists, the data collected by the four devices may not be as similar as desired due to the quality of the components used in the phones. For example, the price range for Inertial Measurement Units (IMUs) is from tens of dollars to thousands of dollars. While the more expensive IMUs make more precise measurements, there is much greater variation and uncertainty in the performance of low-end IMUs. These issues would certainly affect the performance of the indoor localization apps to be developed.

Table VI presents the statistics of inter-sample times for select sensor and RF signal strength data that were sampled at vastly different frequencies for various devices and two scenarios in each of the four buildings. Sensors such as light or step detector, which provide readings only when an event occurs, were not included in the table. We observed greater variation in Wi-Fi inter-sample times for the OnePlus 2 device, mostly due to outliers with scan durations of roughly 12 seconds that consistently occur for this device in all scenarios. Figure 3 depicts the relationship between Wi-Fi inter-sample times and the number of APs detected for Scenario 1 in Building 1. There is correlation between the number of APs detected and inter-sample time for the Nexus 6 and OnePlus 2 devices. Similar patterns are observed for other buildings and scenarios. Note that OnePlus 2 outliers are not shown in the figure.

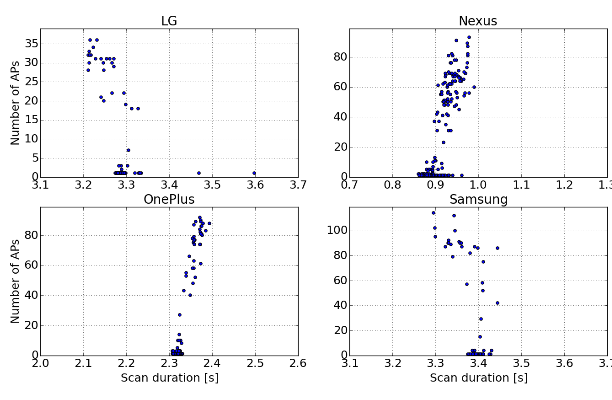
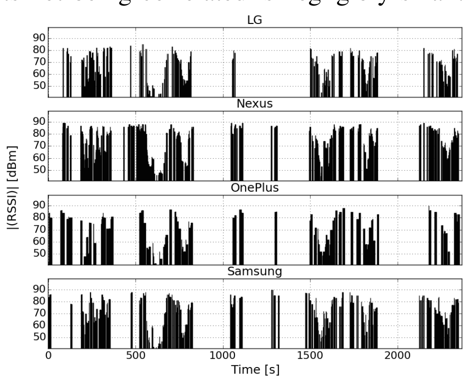


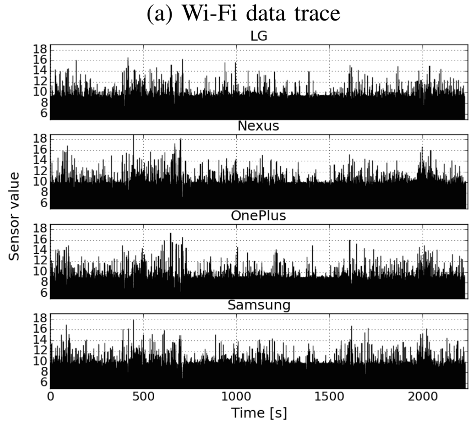
Figure 3: Number of detected Wi-Fi APs vs. scan duration

Table VI shows significant variation in inter-sample times for accelerometer / gyroscope and magnetometers. There are two reasons for this behavior. First, the intended sampling frequency of 100 Hz is not a hard requirement for devices' Operating Systems (OSs). Hence, certain variation in inter-sample time is tolerated by the OS. Second, the time to log the sensor readings, and hence the inter-sample time, is larger when many sensors report readings. Conversely, those times are shorter when fewer sensors need to report readings. The OS balances these times so that the 100 Hz sampling frequency is roughly realized. Anyone who uses our data to develop an indoor localization app should regard the sensor readings as averages over variable-length intervals of time. Except for most scenarios in Building 1 and some scenarios in Building 3, where cellular coverage was non-existent or weak, the inter-sample times for cellular readings show little deviation from the 1 s target.

To study the similarity of readings from sensors of the same type, e.g. pressure sensor readings from the four devices, we compute Spearman's correlation coefficient ρ and corresponding p-values for all six possible pairs of devices. Due to lack of space, we show only the results for Wi-Fi signal strength and accelerometer data in Table VII. As visible in the table, all correlation coefficients are fairly high, i.e. close to the maximum value of 1. That means the data from pairs of devices are strongly correlated. The p-value indicates the significance of the Spearman's correlation coefficient. We did not explicitly report the p-values in the table, since they tend to zero, meaning that the null hypothesis of the combination of datasets not being correlated is negligibly small.



(a) Wi-Fi data trace



(b) Accelerometer data trace

Figure 4: Example data traces for all four devices

We had to deal with the problems of having slightly different start and end times for data collection in a scenario and different numbers of data samples for a given sensor or RF signal strength from various devices before we could compute any correlation coefficient. For both the RF signal strength and accelerometer readings we solved this problem by merging the sampling times from all phones to a full set of time instances. We then interpolated the missing values and used these newly generated time series for calculating correlation coefficients. Additionally, for the accelerometer, in order to reduce the impact of transients spikes, before the interpolation step, we averaged the data over a sliding window with duration of 100 ms and a sliding step of 20 ms. Example Wi-Fi and accelerometer data traces are given in Figure 4 as a pictorial evidence of data similarity. In the accelerometer case, the depicted data is the l_2-norm of the x, y, and z values of each reading. In the Wi-Fi case, we focus on the AP with the largest number of observations in a particular scenario, since such APs are, in contract to APs with less observations, more relevant for the majority of localization solutions that leverage Wi-Fi readings (e.g. [11], [12]).

V. Conclusions

The data we have collected in this project and are making available to the R&D community is truly unique in the world. We doubt many organizations would have the resources to instrument four large buildings, covering about 30,000 m^2 of space, with 900+ test points, have the locations of the test points professionally surveyed, and spend about 200 man-hours on data collection using four Android phones after months of preparation. We presented some analysis that speaks to the validity of the collected data. Researchers across the world will be able to use our annotated data to develop Android indoor localization apps. Our future plans include development and launch of a web portal for comprehensive performance evaluation of indoor localization apps based on the ISO/IEC 18305 standard. In addition, lessons learned from this and other planned indoor localization "system testing" activities will result in improvements to ISO/IEC 18305. Future part(s) of the paper will describe the performance evaluation web portal and present the results of evaluating a number of smartphone indoor localization apps through the portal.

References

[1] ISO/IEC DIS 18305, *Test and evaluation of localization and tracking systems*, (2015). [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:18305:dis:ed-1:v1:en.

[2] D. Lymberopoulos *et al.*, "A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned," in *Proc. IPSN*, 2015, pp. 178–189.

[3] P. Barsocchi *et al.*, "Evaluating ambient assisted living solutions: the localization competition," *IEEE Pervasive Computing*, vol. 12, pp. 72–79, Oct.-Dec. 2013.

[4] F. Lemic *et al.*, "Experimental evaluation of RF-based indoor localization algorithms under RF interference," in *Proc. ICL-GNSS*, 2015, pp. 1–8.

[5] S. Schmitt *et al.*, "A virtual indoor localization testbed for wireless sensor networks," in *Proc. IEEE SECON*, 2013, pp. 239–241.

[6] F. Lemic *et al.*, "Web-based platform for evaluation of RF-based indoor localization algorithms," in *Proc. IEEE ICCW*, 2015, pp. 834–840.

[7] Google, *Protocol buffers*, (2015). [Online]. Available: https://developers.google.com/protocol-buffers/.

[8] ——, *Sensors overview*, (2016). [Online]. Available: http://developer.android.com/guide/topics/sensors/sensors_overview.html.

[9] ——, *Sensors types*, (2016). [Online]. Available: https://source.android.com/devices/sensors/sensor-types.html.

[10] G. Lui *et al.*, "Differences in RSSI readings made by different Wi-Fi chipsets: a limitation of WLAN localization," in *Proc. ICL-GNSS)*, 2011, pp. 53–57.

[11] M. Azizyan *et al.*, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proc. ACM MobiCom*, 2009, pp. 261–272.

[12] F. Lemic *et al.*, "Experimental decomposition of the performance of fingerprinting-based localization algorithms," in *Proc. IPIN*, 2014, pp. 355–364.