Experimental Evaluation of RF-based Indoor Localization Algorithms Under RF Interference

Filip Lemic*, Vlado Handziski*, Adam Wolisz*, Timotheos Constambeys†, Christos Laoudias‡,
Stephan Adler§, Simon Schmitt§, Yuan Yang§
*Telecommunication Networks Group, Technische Universität Berlin
†Department of Computer Science, University of Cyprus
‡KIOS Research Center for Intelligent Systems and Networks, University of Cyprus
§Department of Mathematics and Computer Science, Freie Universität Berlin

***Abstract*—In the current practice, the performance evaluation of RF-based indoor localization solutions is typically realized in non-standardized environments and following ad-hoc procedures, which hampers objective comparison and does not provide clear insight into their intrinsic properties. Many evaluation procedures also neglect important environmental factors like RF interference, diminishing the real-world value of the obtained results. Localization competitions, in which different solutions are evaluated along a set of standardized metrics under unified and representative conditions, can play an important role in mitigating these problems, but their organization is cost and labor intensive. In this paper we report on the design, execution and results from an online localization competition in which several different RF-based indoor localization algorithms have been evaluated with the help of a remotely accessible and automated testbed infrastructure, that reduces these overheads. The competing algorithms have been evaluated following a combination of precision, latency and sensitivity metrics, under four different benchmarking scenarios, resulting in 28 different benchmarking experiments. The obtained results provide strong indication that specific types of RF-interference noticeably degrade the localization performance.**

I. Introduction

Device location is an essential context information for many applications, but is currently readily available only outdoors, thanks to the ubiquitous availability of Global Navigation Satellite System (GNSS). Despite intensive research and development efforts, no single indoor localization technology enjoys similar levels of deployment and acceptance. Selection of suitable technology for a given environment and application remains a hard problem, amplified by the scarcity of reliable data about their performance. Indoor localization solutions are typically evaluated in an ad-hoc fashion, in different environments and using different procedures, impeding objective comparison of their performance. Another problem is the lack of understanding how different environmental factors like Radio Frequency (RF) interference, which is unavoidable in real life, might influence the operation of a specific solution. To mitigate these shortcomings, evaluations need to be performed following standardized procedures and in environments that are close to the real deployment conditions, but still offer high levels of control in order to support different scenarios and repeatability of the results. Localization competitions are one model for achieving the above conditions: competitors are invited to physically deploy their localization solutions in a shared evaluation environment and are evaluated along a set of metrics reflecting different application and user requirements. The main benefit of the model is that the competitors can deploy custom hardware and are in full control of the localization systems, while the organizers are responsible only for the evaluation environment and procedure. Unfortunately, large localization competitions, like the one from Microsoft [1] with 21 teams from all around the world, are infrequent, mainly due to the high deployment and traveling costs. An important observation from this competition is that a high percentage of competing solutions utilize common hardware like Wireless Fidelity (WiFi) APs and smartphones. For this class of solutions a shared hardware base deployed in a selected location is very attractive option for their comparative evaluation. This hardware organized in a form of a testbed can also be made remotely available, thus enabling competitors to deploy their localization solutions just by code upload, without physical presence at the testbed premises.

This paper presents the design, execution and results from such a remote competition for RF-based indoor localization algorithms, in which participating teams competed in precise localization under the influence of different types of RF-interference. The competing localization algorithms were remotely deployed on top of hardware resources available at the experimental facilities of the Technical University of Berlin and including large number of IEEE 802.11g/n Access Points (APs) and IEEE 802.15.4 sensor nodes. They have been subsequently evaluated following a combination of accuracy, latency and sensitivity metrics, under four different RF-interference scenarios. The evaluation process has followed a methodology developed in the EVARILOS project [2], aligned with the upcoming ISO/IEC 18305 standard "Test and evaluation of localization and tracking systems".

The rest of the paper is structured as follows. In Section II we present the evaluation procedure, including the selected metrics and the leveraged methods for their calculation and fusion into final ranking scores. Section III describes the testbed environment and the benchmarking scenarios used in the evaluation of the competing localization algorithms. The execution of benchmarking experiments is described in Section IV, while Section V overviews the competing algorithms. The final results of the competition are summarized in Section VI and Section VII discusses some of the lessons learned from the unique viewpoints of the different parties involved in the competition. Section VIII briefly overviews the related work on indoor localization competitions and testbeds used for experimental benchmarking of indoor localization. Finally, Section IX concludes the paper and outlines directions for the future work.

II. Benchmarking Procedure

This section presents the procedure followed in the evaluation of RF-based indoor localization algorithms during the presented remote competition. The methodology used for the performance evaluation follows the guidelines established by the EVARILOS Benchmarking Handbook (EBH) [2], developed in the scope of the EVARILOS project. While the focus in this work has been on typical and widely deployed RF technologies, the leveraged evaluation methodology can also be applied to any other RF technology.

A. Evaluation Metrics

For characterizing the performance of evaluated indoor localization Systems Under Test (SUTs) three primary and one secondary metric have been used, which are explained in the following. The primary performance metrics are extracted directly from the benchmarking experiments, like point and room level accuracy, and latency (response time). Subsequently, based on the primary performance metrics under different interference scenarios, the interference sensitivity of a SUT has been calculated as a secondary metric. In the following we define these metrics more formally.

a) Point Level Accuracy: Point level accuracy at one evaluation point is defined as the Euclidean distance between the ground truth location (x_{GT}, y_{GT}) and the location estimated by an indoor localization algorithm (x_{EST}, y_{EST}). The point accuracy of location estimation for one point is defined by the following equation:

$$PointAccuracy = \sqrt{(x_{GT} - x_{EST})^2 + (y_{GT} - y_{EST})^2}\ [m] \quad (1)$$

b) Room Level Accuracy: Room level accuracy of location estimation is a binary metric stating the correctness of the estimated room, defined by the following equation:

$$RoomAccuracy = \begin{cases} 1 & \text{if the estimated room is correct;} \\ 0 & \text{if the estimated room is not correct;} \end{cases} \quad (2)$$

c) Latency of Location Estimation: Latency relates to the time that a SUT needs to report the location estimate when requested. The time measured in the evaluation is the difference between the moment when the request for location estimate has been sent to a SUT ($t_{request}$) and the moment when the response arrived ($t_{response}$), as given in the equation:

$$Latency = t_{response} - t_{request}\ [s] \quad (3)$$

d) Interference Sensitivity: Interference sensitivity reflects the influence of different interference types on the performance of an indoor localization algorithm. It is the percentage of change in primary metrics in the scenarios with interference, in comparison to the performance in the scenario without interference (reference scenario). For the case of a generalized metric (M), the interference sensitivity is given according to the following equation:

$$InterferenceSensitivity = \frac{M_{reference} - M_{interference}}{M_{reference}} \cdot 100\,[\%] \quad (4)$$

where $M_{reference}$ is the value of a primary metric M in the reference scenario, while $M_{interference}$ is the value of a metric M in the scenario with interference. Note that if the performance of an algorithm for the performance metric M is better in the scenario with interference, in comparison to the reference scenario, then the interference sensitivity metric is set to 0 %.

B. Obtaining Evaluation Metrics

The evaluation procedure was organized in four benchmarking scenarios. In each scenario, for each of the 20 evaluation points, the set of metrics (point accuracy, room accuracy, latency) was obtained. For each set, the 75^{th} percentiles of point level accuracy and latency were calculated, together with the percentage of correctly estimated rooms. Interference sensitivity was calculated as the difference in each primary metric in each interference scenario, in comparison to the reference scenario, using Equation 4. The overall interference sensitivity was the averaged interference sensitivity over all interference scenarios and all performance metrics, given by the following equation:

$$\overline{M} = \frac{1}{9}\sum_{i=1}^{3}(M_1(i) + M_2(i) + M_3(i)) \quad (5)$$

In the equation the sum goes over all three interference scenarios ($i = 1, 2, 3$), and $M_1(i)$, $M_2(i)$ and $M_3(i)$ are interference sensitivity of 75^{th} percentile of point accuracy, interference sensitivity of room level accuracy and interference sensitivity of 75^{th} percentile of latency for an interference scenario i, respectively.

C. Calculation of Final Scores

Final scores were calculated according to the approach described in the EBH [2], where the score for each metric was calculated according to a linear function that is defined by specifying minimal and maximal acceptable values for each metric. Furthermore, EBH proposes the usage of weighting factors for defining the importance of each metric for a given Evaluation Scenario (ES). In general, the linear translation function for calculating scores for each particular metric is given in Equation 6, with scores ranging from 0 to 10.

$$Score = \max\left(0, \min\left(10, 10\frac{m - M_{acceptable}}{M_{acceptable} - M_{desired}}\right)\right) \quad (6)$$

Acceptable and desired values are defined with $M_{acceptable}$ and $M_{desired}$, respectively. Note that $M_{acceptable}$ can be bigger

TABLE I: Marginal values and weights in evaluation scenarios

Metric	$M_{acceptable}$	$M_{desired}$	ES1 Weight	ES2 Weight	ES3 Weight
Point accuracy	10 m	1 m	0.40	0.20	0.20
Room level accuracy	50 %	90 %	0.40	0.20	0.20
Latency	20 sec	1 sec	0.10	0.50	0.10
Interference sensitivity	50 %	10 %	0.10	0.10	0.50

than $M_{desired}$, e.g. in defining the acceptable point accuracy values one can discuss about acceptable localization error margins. Here $M_{acceptable}$ is the biggest acceptable, while $M_{desired}$ is the desired 75^{th} percentile localization error. The scores for each metric are weighted using predefined weighting factors and are summed together to produce the final score for a particular category. The winners of the competition were declared based on the final scores for the three different sets of marginal values and weights, as presented in Table I.

III. Environment, Infrastructure and Scenarios

This section shortly presents the testbed environment, hardware components and benchmarking scenarios used for the evaluation in the scope of this competition.

A. Benchmarking Environment

The Telecommunication Network Group (TKN) testbed, used as the evaluation environment for this set of benchmarking experiments, is located on the 2^{nd} floor of a campus office building. The environment can be characterized as office space with brick walls, i.e. more than 400 m^2 area with more than 10 rooms. The testbed is comprised of three types of rooms, namely small offices ($14m^2$), big offices ($28m^2$) and laboratories ($42m^2$), as shown in Figure 1. It offers realistic indoor conditions, with a number of people moving around the premises, resulting in small environmental changes like opening of doors or slight movements of infrastructure (chairs, tables, etc.). Most of the internal interference sources are controlled. However, uncontrolled wireless interference from external sources (such as WiFi APs from neighboring buildings) can not be fully excluded.

B. Benchmarking Infrastructure

This section shortly overviews the hardware components used in the experiments in the competition. A detailed description of the infrastructure can be found in [3].

a) Wireless Sensor Network Testbed: The leveraged infrastructure is a multiplatform, hierarchical testbed with 204 SUT sockets, currently populated with 102 eyesIFX and 102 Tmote Sky nodes. The nodes are deployed in a 3D grid spanning 3 floors of an office building, resulting in more than 1500 m^2 of total instrumented office space.

b) WiFi Access Points: The testbed is equipped with 18 dualband TP-link N750 APs (model TL-WDR4300). They run OpenWRT as operating system and cOntrol and Management Framework (OMF) as control and measurement plane. The positions of the WiFi APs in the 2^{nd} floor of the testbed are given in Figure 1.

c) Robotic Mobility Platform: The Turtlebot II robotic platform comprises of a mobile base called Kobuki, a laptop, a router and a Microsoft Kinect 3D camera sensor. The robotic platform is used to navigate in the testbed environment in a controlled and automated way.

d) Embedded PCs: The ALIX2D2 embedded Personal Computers (PCs) are equipped with Broadcom WL5011S 802.11b/g cards. In our infrastructure three ALIX2D2 PCs exist, as shown in Figure 1.

e) Spectrum Analyzers: The testbed infrastructure also comprises several WiSpy sensing devices. These are low-cost spectrum scanners that monitor activity in the 868 MHz, 2.4 and 5 GHz spectrum, and output the measured RF energy and the quality of the received signals.

f) Signal Generator: The Rhode&Schwarz SMBV100A is a flexible signal generator that uses various toolboxes, allowing generation of different standards conform signals like e.g., WiMAX, WiFi or LTE.

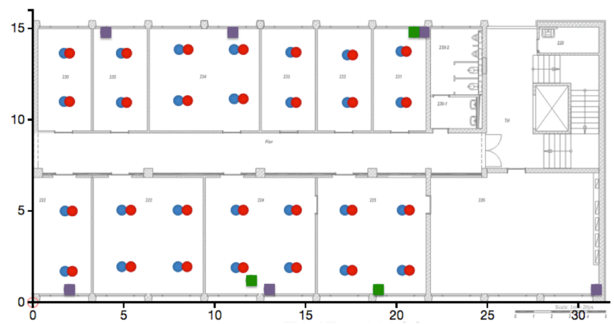

Fig. 1: Deployment locations of the devices in the testbed (red circle: eyesIFX, blue circle: Tmote sky, purple square: WiFi router, green square: AlixD2D PC)

C. Benchmarking Scenarios

The goal of the experimentation in the interference scenarios is to determine if and to which extend different types and magnitudes of RF interference can influence the performance of indoor localization algorithms. As mentioned, we considered four different benchmarking scenarios, in more details described in the technical report accompanying the paper [4].

g) Reference Scenario: The name reflects the fact that in this scenario no artificial interference was generated and the presence of uncontrolled interference was minimized. Thus, the performance of SUTs achieved in this scenario is used as the “reference” for evaluating the impact of interference in the remaining scenarios. The three interference scenarios have been designed to reflect different common coexistence scenarios: one one scenario in which there is a MAC protocol level mechanism supporting coexistence and 2 scenarios without coexistence support.

h) Interference Scenario 1: In the first interference scenario the interference was created using IEEE 802.15.4 Tmote Sky nodes. The interference type was jamming on one IEEE 802.15.4 channel with a constant transmit power equal to 0 dBm. Five of these jamming nodes were present in the testbed environment, as shown in Figure 2. The channel on which the jamming was performed was selected such as to overlap with the channel used by a particular SUT.

i) Interference Scenario 2: The second interference scenario was comprised of several interference sources that are typical for office or home environments. Interference was emulated using 4 WiFi embedded PCs (AlixD2D) having the roles of a server, access point, data client, and video client. The interference transmission streams are schematically depicted in Figure 2. During this scenario, the server acted as a gateway for the emulated services. The data client was represented by a TCP client continuously sending data over the AP to the server. Similarly, the video client was emulated as a continuous UDP stream source of 500 kbps with bandwidth of 50 Mbps. The AP was working on a WiFi channel overlapping with the SUT's operating channel and with the transmission power set to 20 dBm (100 mW).

j) Interference Scenario 3: For the third interference scenario, a signal generator, with location given in Figure 2, was used to generate synthetic interference with an envelope that reflects WiFi modulated signals, but without Carrier Sensing (CS). The transmission power was set to 20 dBm, while the wireless channel on which the interference was performed depended on a particular evaluated SUT.

IV. Execution of Benchmarking Experiments

This section describes how the previously described testbed infrastructure was used in the benchmarking experiments in the scope of the competition. All benchmarking scenarios were instantiated on the 2^{nd} floor of the testbed, and for all of them the same 20 evaluation points were defined, with their locations depicted in Figure 2. At each evaluation point, the indoor localization SUT was requested to estimate the location. The SUT device was positioned at each evaluation point using the robotic mobility platform, at the height of 45 cm above the floor. The navigation stack of the robotic platform gives an order of magnitude more accurate location estimation than the usual SUTs, i.e. less than 10 cm in average in the evaluation environment, and due to that the location obtained from the robotic platform was considered as ground truth for the evaluation [3]. The communication with the robotic platform is done in the 5 GHz Industrial, Scientific and Medical (ISM) band, in order to avoid interfering with the evaluated systems, which all performed in the 2.4 GHz ISM band. The experiments were performed during the weekends afternoons, minimizing the influence of uncontrolled interference. Furthermore, the wireless spectrum was monitored using a WiSpy device attached to the robotic platform and another one at a control point in the testbed, used for assessing the level of uncontrolled interference and validating the correctness of generated controlled interference.

a) SUT Mobile Nodes: Different devices were used as mobile parts of SUTs, i.e. TelosB sensor node, Apple MacBook Pro laptop, Nexus S Android smartphone and Nexus 7 Android tablet. Users were able to use Secure Shell (SSH) tunnels to the desired nodes to deploy their algorithms on a desired device.

b) SUT Infrastructure Nodes: As infrastructural parts of SUTs, nodes from the wireless sensor network testbed or WiFi APs were used, depending on the requirements of a particular algorithm. The locations of the available infrastructure nodes were communicated to the competitors in advance. Although in this work the focus is on the remote access to the shared hardware resources, the testbed also supports deployment of proprietary hardware (both infrastructural and mobile nodes), which broadens the scope of solutions that can be evaluated.

c) Autonomous Mobility: The mobility platform was accessible over a web interface where competitors were able to click on the location on which they wanted to position the robotic platform and their SUT. The competitors were able to send the platform to a location by setting the coordinates of a desired location. Also, it was possible to provide a set of way-points to the platform for a full automation of even the training phase of different algorithms. The platform was able to provide its current location or adequate messages if the desired location was not reachable.

d) Interference Generation: For training and parametrization purposes of their algorithms, competitors were also able to generate the interference scenarios using the above described devices. The code for generating three previously described interference scenarios was provided and the users were able to select the nodes on which the code should run.

e) Interference Monitoring: Competitors were also able to use different devices for monitoring interference levels. Moreover, they were able to obtain the dumps of wireless spectrum using the WiSpy device on the robot or at the fixed location given in Figure 2.

f) Interfacing with the SUT: All competitors had to deploy their algorithms on one of the devices intended for deploying SUTs. Furthermore, competitors had to provide an HTTP Uniform Resource Identifier (URI) on which their algorithm listens for location estimation requests. Upon a request, the algorithms needed to provide the location estimate as a JavaScript Object Notation (JSON) response in the following format:

```
{"coordinate_x": 'Estimated location: coordinate x',
 "coordinate_y": 'Estimated location: coordinate y',
 "coordinate_z": 'Estimated location: coordinate z',
 "room_label":   'Estimated location: room'}
```

JSON parameters *coordinate_x* and *coordinate_y*, expressed in meters, are required parameters and as such they had to be reported. Parameter *coordinate_z* is an optional parameter, due to the 2D evaluation environment. Finally, parameter *room_label* is an optional parameter that is either explicitly provided by the SUT, or automatically mapped from the estimated coordinates x and y. Coordinates (x, y) or (x, y, z) of the location estimates had to be calculated according to a predefined *zero-point* in the environment.

V. Evaluated Indoor Localization Algorithms

This section gives a short description of the indoor localization algorithms evaluated in the competition. The first two algorithms presented below are based on WiFi fingerprinting, followed by algorithms based on multilateration and proximity using low-power IEEE 802.15.4 nodes.

a) Quantile-based Indoor Fingerprinting using Dedicated APs: One of the most promising approaches in indoor localization is fingerprinting using WiFi infrastructure. This fingerprinting-based indoor localization algorithm [5] makes use of the Received Signal Strength Indicator (RSSI) values from WiFi beacon packets for estimating location. For generating fingerprints this algorithm uses the quantiles of RSSI values from beacon packets transmitted from various WiFi APs in the premises. Furthermore, the algorithm uses Pompeiu-Hausdorff distance for calculating the difference between training fingerprints and ones generated by user to be localized. Finally, the algorithm uses the k-Nearest Neighbours (kNN) procedure with the parameter k set to 3.

b) Indoor Geolocation for Android Smartphones with Airplace: This indoor localization algorithm, named Airplace, is an indoor geolocation platform developed for Android tablets and smartphones [6]. Airplace exploits the available WiFi infrastructure and monitors RSSI values from the surrounding APs to determine the unknown user location. The system utilizes a number of RSSI fingerprints collected prior to localization and stored in the radiomap. Location is then estimated by finding the best match between the currently measured fingerprint and fingerprints in the radiomap. The algorithm relies on the Radial Basis Function Network (RBFN) algorithm, which is a neural networks based algorithm that addresses localization as a regression problem [7]. Essentially, it leverages the collected data in the RSS radiomap to build a mapping from the RSS fingerprint space to the 2D physical space. The algorithm has been instantiated in three different versions. In the first version the Android smartphone Nexus S is used as the mobile node for the deployment of the algorithm (Airplace 1). The fingerprinting procedure in this case uses only a set of dedicated APs. Similarly, the second version uses only dedicated APs, but here the mobile node for deploying the algorithm is an Android tablet Nexus 7 (Airplace 2). Finally, in the third version the fingerprinting procedure is done using all APs visible in the environment, while the algorithm is deployed on the same mobile node, Android tablet Nexus 7 (Airplace 3).

c) Geo-n Localization Algorithm: The Geo-n algorithm [8] is a highly precise, distance based, general purpose localization algorithm. It was developed based on evaluation of large experiments and extensive simulation of the impact of the spatial anchor distribution in an indoor localization setting. The algorithm uses multilateration and is able to deal with outliers as well as with heavily error prone distance measurements. Geo-n uses a two stage filtering technique. First, the most representative intersection points between every pair of circles induced by anchor coordinates and distance measurements are obtained. This is done by removing intersection points that do not contribute to the localization or are suspected to increase the positioning error. Geo-n then uses the residual intersections to estimate the position of the unlocalized node. The node types used for both mobile and infrastructural nodes are TelosB low-power sensor nodes.

d) RSS Range-based Positioning Using Grid-based Likelihood Estimation: Ranging using RSSI values from indoor networks is a convenient technique for positioning purposes. Thus, the algorithm [9] explores RSSI ranging from all reachable reference low-power sensor nodes (TelosB) in the environment. For the nonlinear and non-Gaussian positioning problem under highly imprecise RSSI ranging measurements, a simple grid-based likelihood estimation for obtaining location estimates is used. The probabilistic distribution of the target's position is represented by a grid obtained from the Bounding-box algorithm. Afterwards each grid cell is weighted by the residual of current observed ranging measurements and, at last, the state estimation is the likelihood expectation. This algorithm requires no history measurements and no assumption of the measurement model, which can be applied to other positioning scenarios. Furthermore, the grid size is very small, bounded to 36 grid cells for each positioning trial, resulting in a low computation and memory complexity.

e) 3CoM (3 Centers of Mass): For signal strength based indoor localization it is a well known approach to assume the position of the anchor node with the highest signal strength to be the location of the node to be localized. This approach works well in environments with a very dense anchor placement. The approach is extended in order to use the center of mass of the positions of the three strongest anchor nodes (TelosB) to estimate the location [10]. While this approach is far from optimal, this simple method can result in remarkably good results for a lot of testing positions.

VI. Results of the Performance Evaluation

This section presents results of the evaluation performed in the scope of the competition. We evaluated 7 different indoor localization algorithms in four benchmarking scenarios, resulting in 28 benchmarking experiments in total. We firstly present localization errors cumulatively over all competitors at each evaluation point, to show the spatial variability of localization errors due to a position in the environment and due to the RF interference generated in each interference scenario. The points are labeled from 1 to 20, as shown in Figure 2. Secondly, we present the ranking of evaluated algorithms for three different ESs, each emphasizing different indoor localization performance metric, according to Table I. Due to the lack of space, we only present, in our view, the most interesting results. A more extensive set, containing results for each competitor and detailed statistics for each metric, is given in the accompanying technical report [4].

A. Point Accuracies in Different Benchmarking Scenarios

Averaged localization errors per evaluation points for all benchmarking scenarios are given in Figure 3. As visible from

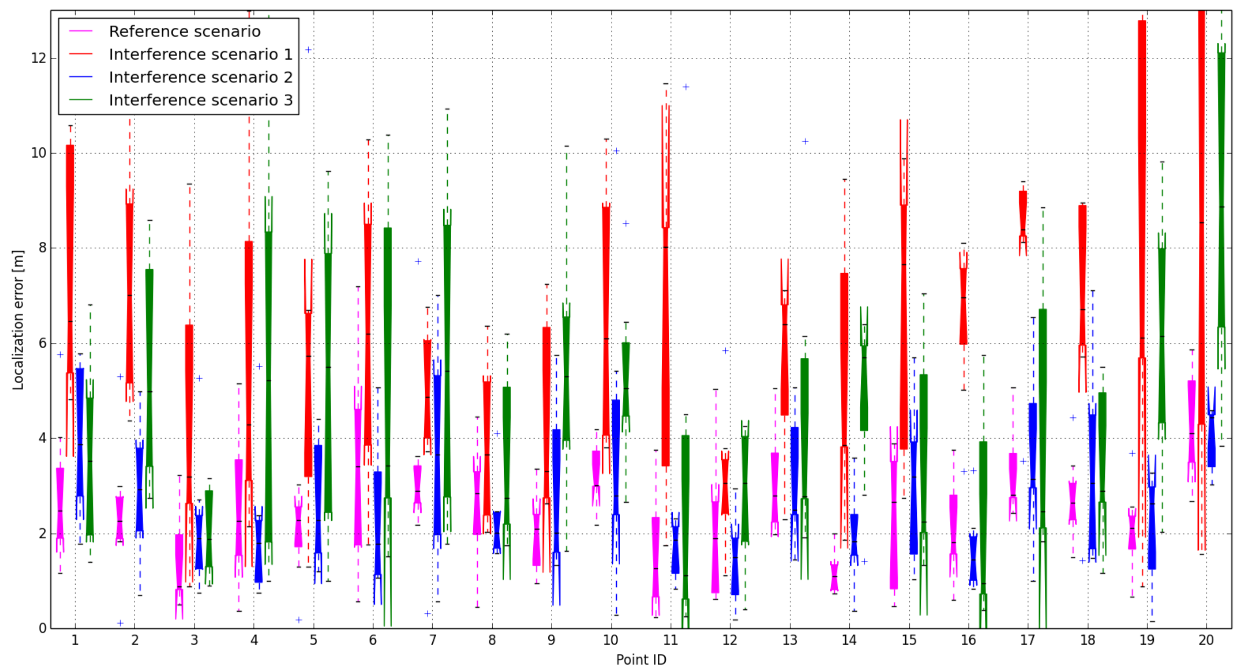

Fig. 3: Localization errors per evaluation points for all benchmarking scenarios

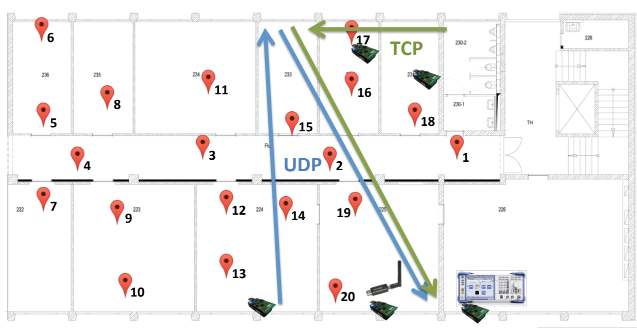

Fig. 2: Locations of interferers and evaluation points

the figure, in general, higher localization errors are achieved close to the borders of the environments, i.e. outside walls, for example in evaluation points 6, 10, 17 and 20. Consequently, in the center of the environment smaller localization errors are achieved, e.g. 2, 3, 9 and 14.

In the scenarios where controlled interference was generated in order to evaluate the influence of different RF interference patterns on the performance of indoor localization algorithms, higher localization errors are generally obtained. In the interference scenario 1, distributed jamming on one IEEE 802.15.4 channel, the achieved localization errors averaged over all competitors at each evaluation point are given in Figure 3. As it can be seen in figure, this interference scenario significantly degrades the averaged point accuracy of evaluated SUTs. Particularly interesting is to observe the tendency of achieving higher localization errors in the evaluation points close to the sources of interference, i.e. points 1, 16, 17, 18, 19 and 20.

Significantly smaller influence on the accuracy of localization is achieved in the interference scenario 2. However, the influence at some evaluation points is still visible, as shown in Figure 3. One reason for lower impact of this interference scenario on the averaged accuracy of localization is the fact that in this scenario WiFi traffic was generated, i.e. CS was enabled. In other words, the results achieved in the interference scenario 2 confirm the value of MAC coexistence as important mechanism for reducing the impact of interference, even for localization applications.

Finally, in the interference scenario 3, i.e. jamming on one IEEE 802.11 channel using signal generator, the averaged accuracy of all evaluated SUTs is again significantly degraded compared to the reference scenario, as shown in Figure 3. Similarly to the interference scenario 1, the averaged localization errors are higher in the evaluation points close to the interference source, e.g. 19 and 20.

B. Ranking of the Competing Algorithms

The ranking of the competing algorithms with respect to each other for the three ESs is given in Table II. Due to the space limitations in the table we only present the results for first three algorithms for each ES, while the full ranking can be found in the technical report [4]. First ES focuses on the point and room level accuracy of indoor localization by giving the highest scores to these metrics. The best performance is achieved by the fingerprinting-based algorithm “Indoor Geolocation for Android Smartphones with Airplace 1”, achieving average localization error of only 1.77 m and room accuracy of 80% in the reference scenario, and final score of 7.13/10. Note that the point accuracy in the table is the 75 percentile localization error, which was the metric used for calculating the scores.

In the ES 2, emphasizing the response time of evaluated algorithms, the best performance is again achieved by the

TABLE II: Summary results of the evaluation for all evaluation scenarios

SUT	Point acc. [m]	Room acc. [%]	Latency [s]	Interference sens. [%]	Point acc. score	Room acc. score	Latency score	Int. sens. score	Final score
Evaluation scenario 1									
Airplace 1	2.71	80.00	3.07	61.43	8.10	7.50	8.91	0.00	7.13
Quantile	3.87	70.00	20.11	28.09	6.81	5.00	0.00	5.48	5.27
Geo-n	2.64	60.00	0.48	62.50	8.18	2.50	10.00	0.00	5.27
Evaluation scenario 2									
Airplace 1	2.71	80.00	3.07	61.43	8.10	7.50	8.91	0.00	7.58
3CoM	2.85	55.00	0.01	37.30	7.94	1.25	10.00	3.17	7.16
Geo-n	2.64	60.00	0.48	62.50	8.18	2.50	10.00	0.00	7.14
Evaluation scenario 3									
Airplace 3	3.77	50.00	3.85	14.17	6.92	0.00	8.50	8.96	6.71
Quantile	3.87	70.00	20.11	28.09	6.81	5.00	0.00	5.48	5.10
3CoM	2.85	55.00	0.01	37.30	7.94	1.25	10.00	3.17	4.43

fingerprinting-based algorithm "Indoor Geolocation for Android Smartphones with Airplace 1", as shown in Table II. Although this algorithm achieves the 75 percentile latency of 3.07 sec in the reference scenario, which is not the best latency result among the evaluated algorithms, due to the performance in other metrics this algorithm is a winning one with final score of 7.58/10. The best performance in terms of 75 percentile latency, namely only 0.01 sec, is achieved by the algorithm "Geo-n Localization", based on low-power sensor nodes.

Finally, the ranking of competitors in the ES 3, emphasizing the interference sensitivity in different interference scenarios, is given in Table II. The winning algorithm is the fingerprinting-based algorithm "Indoor Geolocation for Android Tablets with Airplace 3". Although the other performance metrics achieved by this algorithm in the reference scenario are average, compared to other evaluated algorithms, the change in the metrics due to different controlled RF interference patterns is really small, resulting in the best score (6.71/10) in this ES. In summary, the final scores of all competitors indicate that specific types of RF-interference noticeably degrade the performance of evaluated algorithms.

VII. Lessons Learned

This section shortly summarizes the lessons learned while performing the presented benchmarking experiments.

a) Remote Usage of the Testbed Infrastructure: It is clearly a benefit to use a remotely accessible testbed infrastructure for experimentation with RF-based indoor localization algorithms, since it reduces the costs of visiting the testbed environment and simplifies the execution of experiments. The simplification of the execution of benchmarking experiments is due to the developed support for generating and monitoring RF interference, and controlling the mobility platform, so the experimental overhead on the user is significantly reduced. Despite the offered simplifications, some users still required a fair amount of support, mostly due to not being physically present at the testbed premises, which increased the abstraction of using different hardware components. This indicates that some work on simplifying the usage of the benchmarking infrastructure is still necessary. Our further focus will be on designing and developing a high-level interface that wraps all functionalities of the remotely accessible testbed in one Application Programming Interface (API), thus further simplifying the usage of the infrastructure for the end-users.

b) Performing Automated Benchmarking Experiments: The presented benchmarking experiments were executed without the presence of a test-person, thus increasing the comparability of algorithms and objectiveness of achieved results. In our benchmarking experiments all SUT nodes were positioned with the same orientation and on the same height, with average error in positioning smaller than 10 cm [11], which would be hardly achieved by a test-person. Obviously, by not having a test-person we could not capture the effects that that person could have on the performance of the evaluated algorithms. Except for benefits in terms of increased comparability of benchmarking results, using the automated infrastructure for benchmarking significantly improves the time needed for performing a benchmarking experiment. Namely, one experiment as presented here, i.e. surveying 20 evaluation points, usually took 20 minutes, and during the execution results were automatically stored and metrics were calculated and presented. Only one testbed operator was necessary during the automated experimentation, and this only for the support purposes, e.g. robotic platform unable to avoid an obstacle, closed doors, drained battery of a SUT mobile node, etc.

c) Results of Benchmarking Experiments: The evaluation results show good performance of the evaluated algorithms, with the best performance being 1.77 m in average localization error, 80 % in room level accuracy and less then 1 sec in response time in the reference scenario. The spatial distributions of localization errors show that generally higher errors seem to be achieved in the margins of the environment, i.e. closer to the outside walls. This is most pronounced for the algorithms based on low-power sensor nodes, due to the larger number of reachable infrastructural nodes in the center of the evaluation environment. Presumably due to the higher transmission power of WiFi, this trend is not that emphasized in fingerprinting-based algorithms. Interestingly, due to the dense deployment of low-power sensor nodes, the proximity-based algorithm performed unexpectedly well. This indicates that, with enough dense deployment, similar technologies such as iBeacon based on Bluetooth LE can be adequate and simple to use for indoor localization purposes.

d) Influence of RF Interference on Benchmarking Results: Obtained results show that certain interference patterns highly influence the performance of the evaluated systems, which motivates further in-depth analysis on how to mitigate the observed negative impact. Specifically, jamming as type of interference has a strong impact on all algorithms, while the usual interference mitigation mechanisms help avoiding the effects of normal WiFi traffic as interference. Spatial distributions of localization errors in the interference scenarios indicate that higher errors usually occur on the locations closer to the sources of interference, which is specially emphasized for the algorithms based on the low-power sensor nodes. The evaluated algorithms already indicate some mechanisms that might be useful for reducing the impact of interference. For example, increasing the latency of producing an estimate, i.e. collecting more measurements ("Quantile-based Indoor Fingerprinting using Dedicated APs"), or using all APs in the environment instead of only a dedicated set of APs ("Indoor Geolocation for Android Smartphones with Airplace 3"), seem to achieve higher robustness to RF interference. This fact indicates the trade-off between latency and interference sensitivity in the first case, and number of used APs and interference sensitivity in the second one, both having a direct impact on the computation time to obtain the location estimate and consequently on the battery consumption of a mobile device.

VIII. Related Work

As stated previously, indoor localization competitions are rare due to their labor, time and cost intensity. In [1] the authors report on the experiences and lessons learned during the IPSN / Microsoft indoor localization competition. In contrast to the competition described in this work, the evaluation in [1] was done manually by carrying an indoor localization device to different evaluation points, while in this work an automated testbed infrastructure was leveraged. On the other side, the competition described in [1] evaluated a broader set of solutions, including magnetic, light and ultrasound-based systems, while in this work we considered only RF-based indoor localization. Also, while the evaluation in [1] was done on basis of one scenario, in this work we considered four scenarios focused on the interference effect on the performance of indoor localization algorithms. The set of EvAAL competitions is another popular line of competitions focused on indoor localization for the assisted living scenarios [12].

In terms of using a customized testbed infrastructure for the evaluation of indoor localization algorithms the VirTIL testbed offers [13] similar facilities like the one used in our case, but without the capability of creating controlled RF interference context. Finally, the the w-iLab.t II testbed [14] also offers similar functionality, but represents open-space instead of office environment.

IX. Conclusion and Future Work

In this paper we reported on the design, execution and results from an online localization competition in which several RF-based indoor localization algorithms have been evaluated using a remotely accessible and automated benchmarking infrastructure. The paper detailed on the testbed environment and infrastructure, and how it was remotely used in the evaluation of different algorithms. Furthermore, it presented scenarios in which different indoor localization algorithms were evaluated and the interference patterns generated in order to evaluate the impact of interference on the evaluated algorithms. The paper also presented the procedure followed in the evaluation of the performance, describing the used metrics, how they were calculated and how the final scores were obtained in three different evaluation scenarios. Finally, it presented the results of the evaluation and ranking of the evaluated algorithms in three different evaluation scenarios. While the focus of this paper was on the objective comparison of the performance of different algorithms, for the understanding of the reasons for e.g. sensitivity to RF interference, or for optimal parametrization of different algorithms, the decomposition and more detailed testing is necessary, like for example the one presented in [15]. Future work will be oriented on the characterization of environments for the needs of extrapolation of results / ranking of indoor localization algorithms from one environment to another.

Acknowledgments

This work has been partially funded by the European Commission (FP7-ICT-FIRE) within the project EVARILOS (grant No. 317989). The author Filip Lemic was partially supported by DAAD (German Academic Exchange Service).

References

[1] D. Lymberopoulos *et al.*, "A Realistic Evaluation and Comparison of Indoor Location Technologies: Experiences and Lessons Learned," in *IPSN'15*, 2015.

[2] T. Van Haute *et al.*, "The EVARILOS Benchmarking Handbook: Evaluation of RF-based Indoor Localization Solutions," in *MERMAT'13*, 2013.

[3] F. Lemic, J. Büsch, M. Chwalisz, V. Handziski, and A. Wolisz, "Infrastructure for Benchmarking RF-based Indoor Localization under Controlled Interference," in *UPINLBS'14*, 2014.

[4] F. Lemic *et al.*, "EVARILOS Open Challenge: Track 3," Technische Universität Berlin, Tech. Rep. TKN-14-005, 2014.

[5] F. Lemic, "Benchmarking of Quantile-based Indoor Fingerprinting Algorithm," Telecommunication Networks Group, Technische Universität Berlin, TKN Technical Report Series TKN-14-001, 2014.

[6] T. Constambeys *et al.*, "Indoor Geolocation for Android Smartphones with Airplace," EVARILOS Open Challenge, Short paper, 2014.

[7] C. Laoudias *et al.*, "Localization using radial basis function networks and signal strength fingerprints in wlan," in *GLOBECOM'09*, 2009.

[8] H. Will, T. Hillebrandt, and M. Kyas, "The Geo-n localization algorithm," in *Indoor Positioning and Indoor Navigation (IPIN)*, 2012.

[9] Y. Yang, Y. Zhao, and M. Kyas, "RBGF: Recursively Bounded Grid-based Filter for Indoor Position Tracking using Wireless Networks," *IEEE Communications Letters*, 2014.

[10] S. Adler *et al.*, "3CoM (3 Centers of Mass) Indoor Localization Algorithm," EVARILOS Open Challenge, Short paper, 2014.

[11] W. Liu *et al.*, "Demo Abstract: A Proof of Concept Implementation for Cognitive Wireless Sensor Network on a Large-scale Wireless Testbed," in *EWSN'13*, 2013.

[12] P. Barsocchi *et al.*, "Evaluating AAL Solutions through Competitive Benchmarking: The Localization Competition," *IEEE Pervasive Computing Magazine*, 2013.

[13] S. Schmitt *et al.*, "A Virtual Indoor Localization Testbed for Wireless Sensor Networks," in *SECON'13*, 2013.

[14] S. Bouckaert *et al.*, "The w-iLab. t Testbed," in *Testbeds and Research Infrastructures*. Springer, 2011.

[15] F. Lemic, A. Behboodi, V. Handziski, and A. Wolisz, "Experimental Decomposition of the Performance of Fingerprinting-based Localization Algorithms," in *IPIN 2014*, 2014.