Computer Networks 000 (2016) 1–9

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Sample Size Determination Algorithm for fingerprint-based indoor localization systems

Loizos Kanaris[a,*], Akis Kokkinis[a], Giancarlo Fortino[b], Antonio Liotta[a], Stavros Stavrou[c]

[a] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Department of Electronics, Informatics and Systems, University of Calabria, Italy
[c] Faculty of Pure and Applied Sciences, Open University of Cyprus, Nicosia, Cyprus

A R T I C L E I N F O

Article history:
Received 2 August 2015
Revised 12 November 2015
Accepted 14 December 2015
Available online xxx

Keywords:
Internet of Things
Sample size
Evaluation
Indoor positioning systems
Localization error
Positioning accuracy

A B S T R A C T

Provision of accurate location information is an important task in the Internet of Things (IoT) applications and scenarios. This need has boosted the research and development of fingerprint based, indoor localization systems, since GPS information is not available in indoor environments. Performance evaluation of such systems and their related localization algorithms, is usually based on sampling collection in predetermined test environments. The sample size determination and sampling methodology can significantly affect the reliability of the outcome. This work proposes an algorithm that calculates the minimum sample size of positioning data required for objective performance evaluation of fingerprint based localization systems. The use of a correct, independent, unbiased and representative sample size can speed up the training, evaluation and calibration procedures of a fingerprint based localization system, while ensuring that the system's true accuracy is achieved. The proposed Sample Size Determination Algorithm (SSDA) takes into consideration the desired confidence level, the resulting standard deviation of a small size preliminary sample as well as the error approximation with respect to the actual error of the system and proposes the final sample size for the evaluation and/or calibration and/or training of the utilized radio-maps. Additionally, the SSDA, assumes random sample allocation in the area of interest in order to avoid biased results. Risks arising from the selection of a sample of convenience are also investigated. Finally, the performance of the proposed algorithm is tested in both measured and simulated radio-maps.

1. Introduction

In the Internet of Things (IoT), several applications and scenarios envision the integration of a great variety of wireless technologies that will provide services based on the user behavior [15]. Such services often require the localization and tracking of the user in indoor environments of smart cites (such as malls, hospitals, underground stations) and smart houses [1,9,14,28]. Fingerprint-based positioning is one of the mostpopular indoor localization techniques implemented by Real Time Localization Systems (RTLS). This technique typically utilizes the Received Signal Strength (RSS) to perform positioning. Other radio parameters can be also used or combined, such as Power Delay Profile (PDP), Angle of Arrival (AOA) etc. RTLS may also utilize non-radio parameters, such as inertial measurements or prior knowledge of environment constraints, in an aim to improve accuracy [11]. Fingerprint-based positioning requires the generation of a dataset of measurements,

* Corresponding author.
E-mail addresses: l.kanaris@tue.nl (L. Kanaris), a.kokkinis@tue.nl (A. Kokkinis), g.fortino@unical.it (G. Fortino), a.liotta@tue.nl (A. Liotta), stavros.stavrou@ouc.ac.cy (S. Stavrou).

http://dx.doi.org/10.1016/j.comnet.2015.12.015
1389-1286/

usually the RSS, during an off-line phase. This dataset, called radio-map, requires calibration before being utilized for the estimation of the user location during the on-line phase. Calibration is important in order to train and configure the positioning algorithms to perform better for the specific radio-map [17]. The calibration techniques use a sample of measurements, which is taken in the area of the indoor environment. The sample size, as well as the allocation of the samples in the area of interest, influence the overall accuracy of the localization system. A similar sample of measurements is also used for the performance evaluation of the system. The calibration and evaluation procedures are two important steps that are influenced by the quality of the sample measurements. Selecting a small sample size or a biased sample can result in misleading calibration parameters and a degraded accuracy during location estimation. On the other hand, large sample sizes are more time consuming and more expensive to carry out. To the authors best knowledge, no previous work exists for selecting the aforementioned sample sizes in fingerprint-based indoor localization systems. The main goal of this paper is to develop and suggest an algorithm that will define the minimum sample size which will ensure correct calibration or training of the RTLS, and objective evaluation of the system's performance.

The rest of the paper is organized as follows: Section 2 presents related work on fingerprint-based methodologies and performance evaluation techniques. Section 3 introduces and analyses the proposed Sample Size Determination Algorithm (SSDA). Sections 4 and 5 describe the evaluation of SSDA based on an experimental indoor localization platform. Finally, Section 6 summarizes the conclusions.

2. Related work

Fingerprint-based positioning requires the construction of a fingerprint database during an off-line phase, in which a number of radio parameters are stored in the form of vectors. The aforesaid database is generated either by performing a measurement campaign or through simulation procedures. In the latter case, statistical, semi-deterministic or fully deterministic models are utilized [3,4,6,10,17,18,26]. A calibration/training procedure is then implemented, using a measurement sample, in order to identify the optimum configuration parameters for the positioning algorithms and minimize localization errors.

The location estimation is performed by the user during the on-line phase, by implementing various deterministic or probabilistic localization algorithms. Examples include the K-Nearest Neighbor (KNN) and the Weighted K-Nearest Neighbor (WKNN) as analyzed in [13], the Minimum Mean Square Error (MMSE) presented in [19] and the Maximum A Posteriori (MAP) [27].

The evaluation of the Fingerprint-based RTLS is also performed by retrieving a measurement sample. The importance of sampling is highlighted in [23], where the authors proposed the development of a benchmark standard. They specifically state that samples constitute the core of any benchmark for location systems, since they are used to compute the position estimates. They proposed that the benchmark specification should state the number of samples recorded per second and the duration of the measurements per location.

When trying to review common practices, literature suggests that researchers tend to utilize a diverse number of sample sizes for the purpose of evaluating their research work, without necessarily clarifying the rationale behind the sample selection. In [19], 40 observations were recorded for a set of 155 calibration points, that were used as training data to eliminate the randomness of human behavior. In [27], authors measured one sample per second, for a period of five minutes (300 samples total), while trying to investigate wireless channel changes over time.

Authors in [5] proposed a dynamic hybrid projection (DHP) technique for improved 802.11 localization. During their experiments they collected 802.11 RSS data at 27 different reference locations in the area of interest, on different days and at four different user orientations. Out of this sample they selected 15 locations with a step of 1.5–2 m, which they then used as training data.

A different sample size was used in [25], where different filtering strategies for real life indoor 802.11 positioning systems were analyzed and compared. The authors measured the radio distribution at 250 uniformly distributed grid points in an area of 15 m x 35 m.

In [7], the differences among the received signal strengths from a number of 802.11 adapters were investigated. The authors of this work conducted system validation using a total of 3120 positioning requests.

In another aspect of indoor positioning, authors of [12] introduced several fault models to capture the effect of failures in the wireless infrastructure. During the investigation of fault tolerance of positioning methods and evaluation in terms of their performance degradation, they contacted experiments based on a radio-map consisting of 107 distinct reference locations having a step of 2–3 m. A total of 3210 reference fingerprints, corresponding to 30 fingerprints per reference location, were collected at the rate of 1 sample/s. For testing purposes they collected fingerprints along a path, consisting of 192 locations. Authors of [24] proposed a novel indoor localization scheme based on sub-area fingerprint determination and surface fitting. During the performance evaluation of the proposed technique, they performed experiments in an area 16.2 m x 28.5 m by setting 25 reference points per room and randomly selecting 200 test points in the environment.

Finally, authors of [20,21], worked towards the challenge of deployment load reduction in RSS based indoor localization systems. In their work, they proposed an interesting scheme that combines the data retrieved from a ray tracing simulator with a limited number of measurements (15%–30% of the complete fingerprint dataset) and performs localization using manifold alignment. The aforementioned methodology leads to a significant load reduction but assumes that the utilized fingerprint datasets have stronger correlation among neighboring data points compared to other points. This assumption is not always valid in indoor environments with strong multipath effects in Rayleigh channels.

Summarizing, our literature review suggests that authors calibrate, train, test and evaluate indoor positioning systems by utilizing a variety of sample sizes and

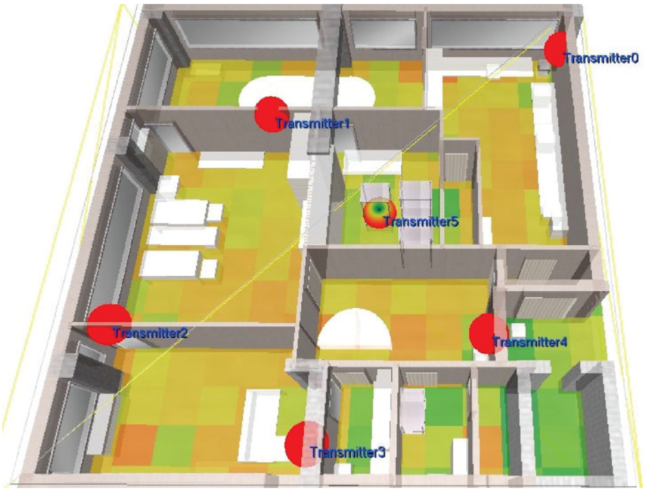


Fig. 1. Test environment with APs and radiomap as generated in TruNET simulator.

sample patterns. The authors of this paper aim to introduce a generic methodology to calculate a sample size that will capture a set of predefined accuracy criteria, within the desired confidence level, for any test environment. This methodology is expected to contribute towards the standardization of evaluation procedures for indoor fingerprint-based positioning systems.

3. Proposed approach

In this paper, the Sample Size Determination Algorithm (SSDA) is proposed, for calculating the minimum, independent, unbiased, representative sample size of positioning data, as this applies to fingerprint-based indoor localization systems. *Minimum* in the sense that, it is large enough to ensure that, the estimated mean positioning error of the system lies within the desired confidence level, hence the evaluation and/or calibration results are within the set reliability criteria. *Independent, unbiased and representative* in the sense that the sample positions need to be selected randomly within the area of interest, in order to avoid convenient patterns or specific samples that may cause biased results which can be questioned. The first step of the methodology requires the selection of a small preliminary sample of size n_{sps}, the identification of the desired confidence interval ci and the Error Bound EB. The *small size preliminary sample* is required in order to initially estimate a mean positioning error $\bar{x}_e$ for the positioning algorithm and the radio-map under examination. The *confidence interval ci*, indicates the probability that the calculated mean positioning error provided by the distribution of samples, describes the true radio-map positioning error and lies within the distribution spread. In other words, ci represents the desired confidence that the evaluator would like to have, and for this reason it is set by the evaluator himself at the beginning of the process. The same applies to the *Error Bound EB*, which refers to the acceptable – in the evaluators' opinion – error that may occur, relative to the calculated mean positioning error $\bar{x}_e$ of the *small size preliminary sample*, always within the selected ci. These parameters are used as initial input to the algorithm, assuming that the mean positioning error of the distribution of samples has a normal distribution (bell-curved). Based on the *Central Limit Theorem*, the aforementioned assumption is considered valid if the sample size is sufficiently large. The population of all possible sample means can then be considered approximately normally distributed, no matter what probability distribution describes the sampled population [16]. The SSDA algorithm is presented in the form of pseudo-code 3.1 and explained in more detail below.

Algorithm 3.1: SSDA(n_{sps}, ci, EB, A)

procedure INITIALPOSITIONING(Positioning Algo, n_{sps})
  for $i \leftarrow 1$ **to** n_{sps}
    do $\bar{x}_e, s_{x_e} \leftarrow$ Positioning Algo
  return $(\bar{x}_e, s_{x_e})$

main
  comment: Step 1: Calculate df, $t_{\frac{ci}{2}}$

  $df \leftarrow (n_{sps} - 1)$
  $t_{\frac{ci}{2}} \leftarrow$ t-table
comment: Step 2: Estimate preliminary values of $\bar{x}_e$,s_{x_e}
for $i \leftarrow 1$ **to** n_{sps}
  do INITIALPOSITIONING(WKNN, MMSE etc., n_{sps})
comment: Step 3: Calculate n_{SSS}

$$n_{SSS} \leftarrow \left(\frac{t_{\frac{ci}{2}} s_{x_e}}{EB}\right)^2$$

comment: Step 4: Calculate GS_{min}

$P \leftarrow (100 * n_{SSS})$
$GS_{min} \leftarrow \frac{\sqrt{A}}{\sqrt{P}+1}$
comment: Step 5: Select Random Samples
for $i \leftarrow 1$ **to** n_{SSS}
  do $Sample_i \leftarrow$ Random $x,y \in \mathcal{A}$, step GS_{min}

Initially, a small preliminary sample of size n_{sps} that is randomly selected, is used for the initial rough estimation

Table 1
Material constitutive parameters of the test environment.

Material	El. perm. (F/m)	Loss tang.
Concrete	3.9	0.23
Wood	2	0.025
Brick	5.5	0.03
Metal	1	1,000,000
Plasterboard	3	0.067
Glass	4.5	0.007

Fig. 2. Positioning error vectors-sample of convenience $con5$ with $n_{con} = 10$ – WKNN.

of the mean positioning error. The table of the standard normal curve cannot apply due to inefficient sample size, hence the t-distribution is implemented, which requires the estimation of degrees of freedom df. From n_{sps}, df can be calculated using the Eq. (1):

$$df = n_{sps} - 1 \tag{1}$$

The t-distribution, given by Eq. (2), in combination with the desired ci are used to estimate the t-value $t_{\frac{ci}{2}}$ for any small preliminary sample:

$$f(t_c i) = \frac{\Gamma(\frac{df+1}{2})}{\sqrt{df\pi}\,\Gamma(\frac{df}{2})}\left(1 + \frac{t^2}{df}\right)^{-\frac{df+1}{2}} \tag{2}$$

where Γ is the gamma function.

A more convenient way is to directly use the t-distribution table, available in most probability and statistics books, based on which the t_{ci} value is given for several dfs. The next step, involves an initial position estimation procedure based on the preliminary sample data, in order to calculate the preliminary sample mean positioning error x_e, and the respective standard deviation s_{x_e}.

Assuming a normal distribution, as discussed earlier, the spread of the positioning error, for the desired ci, is given by the following formula:

$$x_e \pm t_{\frac{ci}{2}}\left(\frac{s_{x_e}}{\sqrt{n}}\right) \tag{3}$$

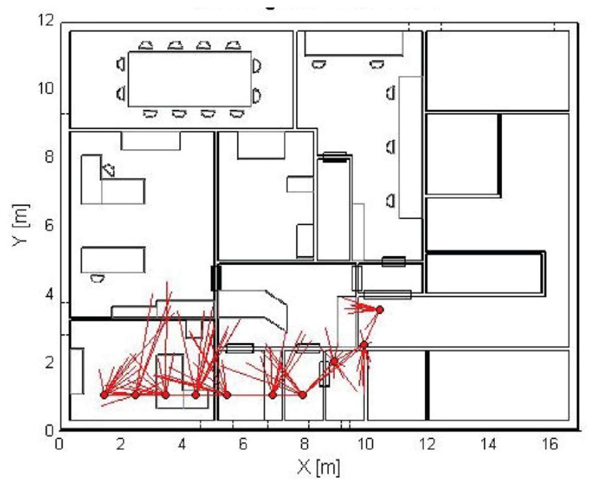


Fig. 3. Positioning error vectors-sample of convenience $con5$ with $n_{con} = 10$ – MMSE.

Table 2
Mean error and standard deviation of preliminary samples – radio-map generated through measurements.

Algorithm	Sample	$\bar{x}_e(m)$	s_{x_e}
WKNN	$sps1$	1.73	1.227
	$sps2$	1.78	1.268
	$sps3$	1.82	1.264
	$sps4$	1.59	1.078
	$con5$	1.49	0.763
MMSE	$sps1$	1.72	1.307
	$sps2$	1.68	1.371
	$sps3$	1.74	1.378
	$sps4$	1.54	1.178
	$con5$	1.50	0.798

where $t_{\frac{ci}{2}}(\frac{s_{x_e}}{\sqrt{n}})$ is the maximum deviation from the mean value, hence the error bound.

Having set the desired EB as the basic evaluation criterion, the suggested sample size (SSS) can be calculated by solving Eq. (3) with respect to n:

$$n_{SSS} = \left(\frac{t_{\frac{ci}{2}} s_{x_e}}{EB}\right)^2 \tag{4}$$

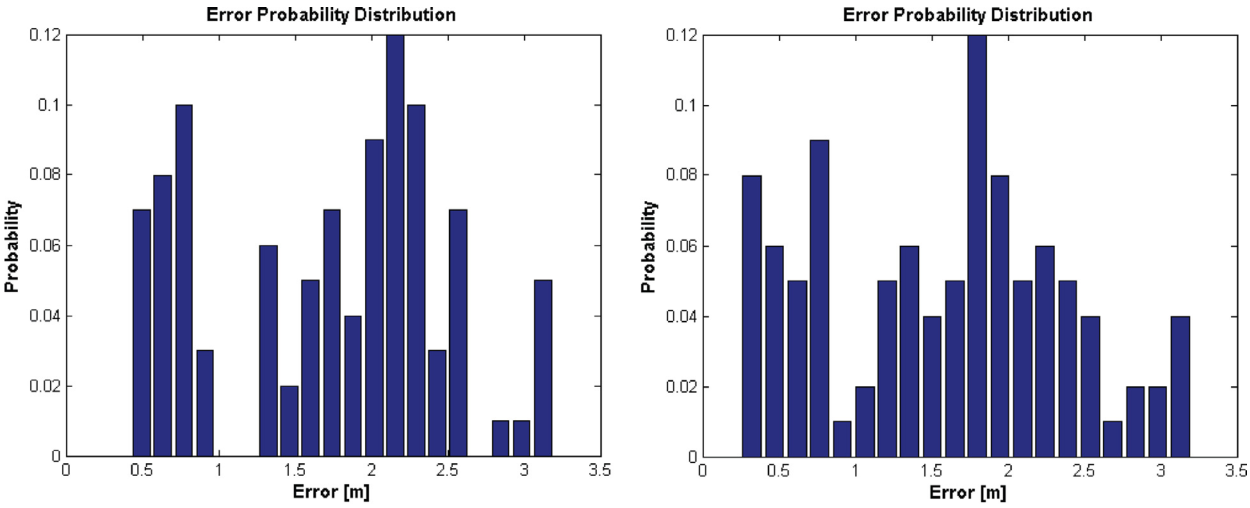
In order to select an independent random sample of n_{SSS} locations within the test environment, the population of the locations should be converted from an infinite number (continuous) to discrete, in such a way that it fulfills the statistical criteria that the population P should be at least 100 times more than the sample size [2]. Hence,

$$P = 100 n_{SSS} \tag{5}$$

Using Eq. (5) and the dimensions of the test environment (width w x height h), the minimum grid size GS_{min} can be determined:

$$GS_{min} = \frac{\sqrt{A}}{\sqrt{P} + 1} \tag{6}$$

where A is the area of the test environment in m^2.



(a) WKNN (b) MMSE

Fig. 4. Positioning error distribution-sample of convenience *con*5 with $n_{sps} = 10$.

The proposed methodology concludes with the selection of n_{SSS} simple random locations within *A*, using the grid size GS_{min} as a step. The *simple random selection* is performed by a method in which each collection of n_{SSS} locations is equally likely to comprise the sample. In any other case, where the selection of the sample may involve a predefined pattern (*sample of convenience*), e.g. selection of all n_{SSS} samples from a specific room in the area of interest, the calculated mean positioning error, may differ systematically from the actual error of the RTLS [16].

SSDA implementation ensures that the selected sample will reflect the overall system performance and characteristics, within the desired confidence level and the predefined acceptable mean localization error bound, for the utilized positioning algorithm. In case of evaluating more than one positioning algorithms, then the SSDA should be applied to each one of them and the common suggested sample size n_{SSS} should be the largest extracted from all algorithms, in order to ensure consistency and reliability of the performance results for all testing scenarios.

4. Test environment

The proposed algorithm was tested using two RSS fingerprint data sets, also known as radio-maps. The first radio-map was generated through measurements and the second through simulation. The SSDA was implemented initially on the first radio-map and the performance was tested for several sample sizes, up to 40 samples. The second-simulated – radio-map – was generated and utilized extensively for much larger sample sizes (up to 500), in order to further investigate the reliability and convergence of the proposed algorithm. The test environment was an indoor area of 169 m^2. The wireless network consisted of 6 D-Link 802.11 APs, as shown in Fig. 1.

The first radio-map was generated through measurements using an Android-based MS device (HTC Desire HD). Fingerprints were collected at 110 locations at a step of 1 m, at a height of 90 cm. The device orientation was kept constant throughout the measurement positions. Additionally, during the measurement procedure no human or machine motion was allowed in the whole test environment, in order to minimize any dynamic fluctuations that may have affected the quality of the generated radio-map. At every measurement point, 30 district RSS samples (1 sample/s) were recorded and the mean RSS value for each location was extracted to formulate the radio-map. The RSS values in the radio-map ranged from −99 dBm to −34 dBm.

The second radio-map was created using *TruNET*, a 3D ray tracing polarimetric simulator, as shown in Fig. 1. The same building structure and large furniture were imported and configured using material constitutive parameters obtained from literature [22], as shown in Table 1. A calibration procedure was carried out as described in [8]. A high density receiver location layout was used in the simulated scenario, by defining the receiver step distance to 10 cm. This resulted in the generation of 16,900 fingerprints in the data-set, allowing the selection and testing of a wide range of sample sizes.

5. Performance evaluation

5.1. Implementing SSDA

A number of sample sizes was calculated by implementing SSDA in the test environment, for different acceptable Error Bounds, maintaining the desired confidence interval at 95% ($\frac{ci}{2} = 0.025$). Four small size preliminary samples of $n_{sps} = 10$ were randomly chosen within the test environment ($df = 9$, and $t_{0.025} = 2.262$). Additionally, a fifth *sample of convenience* of the same size ($n_{con} = 10$) was selected, in order to observe the probable differences and highlight the importance of selecting a *simple random* sample. This specific sample consisted of measurements taken explicitly from the southern area of the test

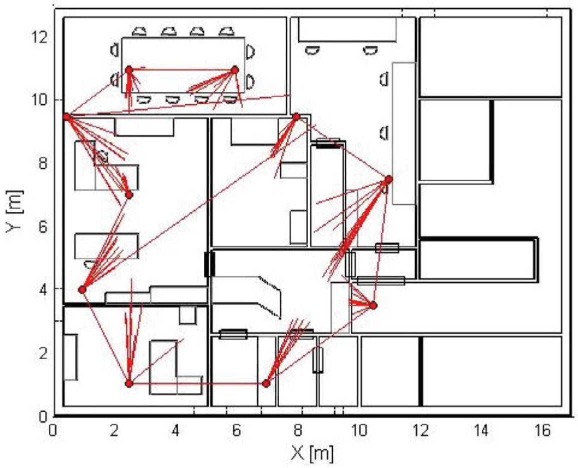


Fig. 5. Positioning error vectors-random sample *sps*3 with $n_{sps} = 10$ – WKNN.

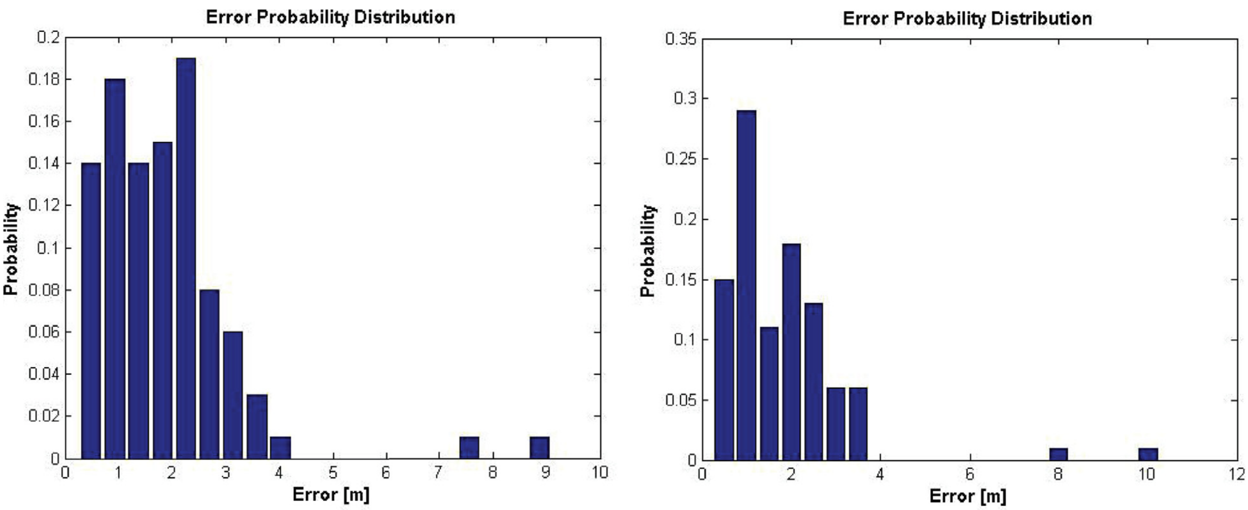
Fig. 6. Positioning error vectors-random sample *sps*3 with $n_{sps} = 10$ – MMSE.

environment as shown in Figs. 2 and 3. The southern area was selected based on the observation that the estimated local positioning error in this specific area was consistently less than the average of the test environment, offering the opportunity to highlight the risk of selecting such type of samples. All five small size preliminary samples were utilized, as defined by the SSDA algorithm, for an initial positioning procedure, using a deterministic (WKNN) and a statistical positioning algorithm (MMSE) on the generated through measurements radio-map. The results of $\bar{x}_e$ and s_{x_e}, which are necessary for the estimation of n_{SSS}, are depicted in Table 2. A representative example of the positioning error vectors for both WKNN and MMSE referring to simple random sample *sps*3 is shown in Figs. 5 and 6.

The respective positioning error distribution is presented in Fig. 7. The suggested sample size n_{SSS} and the recommended minimum grid size GS_{min} per case, provided by SSDA, are depicted in Table 4.

5.2. Simple random sample vs sample of convenience

Based on the test results presented in Table 2, it is observed that both $\bar{x}_e$ and s_{x_e} values of the sample of convenience differ noticeably from the respective values of all measured simple random samples. The mean positioning error recorded in the case of the sample of convenience is $\bar{x}_{e\,con} = 1.49$ m which is less than the minimum recorded positioning error of all other samples. Although the difference is relatively small, when combined with the calculated standard deviation $s_{x_{e\,con}} = 0.76$, may clearly mislead the evaluator to select smaller sample sizes (see Table 3), since it designates an RTLS with a small positioning error distribution. This observation is reflected in Fig. 4 vs Fig. 7. In the aforementioned figures, one can observe a positioning distribution error range between 0.5 m and 3.2 m in the case of the sample of convenience, while in the simple random sample case (*sps*3) the recorded range



(a) WKNN (b) MMSE

Fig. 7. Positioning error distribution-random sample *sps*3 with $n_{sps} = 10$.

Table 3
Suggested sample size for different desired *EB* values using sample of convenience – radio-map generated through measurements.

Algorithm	*EB* (m)	n_{SSS}	GS_{min} (m)
WKNN	±0.25	48	0.185
	±0.40	19	0.290
	±0.50	12	0.0.360
MMSE	±0.25	53	0.176
	±0.40	21	0.280
	±0.50	14	0.340

Table 4
Suggested sample size for different desired *EB* values using simple random *sps* – radio-map generated through measurements.

Algorithm	*EB* (m)	n_{SSS}	GS_{min} (m)
WKNN	±0.25	120	0.120
	±0.40	48	0.185
	±0.50	30	0.235
MMSE	±0.25	180	0.10
	±0.40	70	0.15
	±0.50	45	0.19

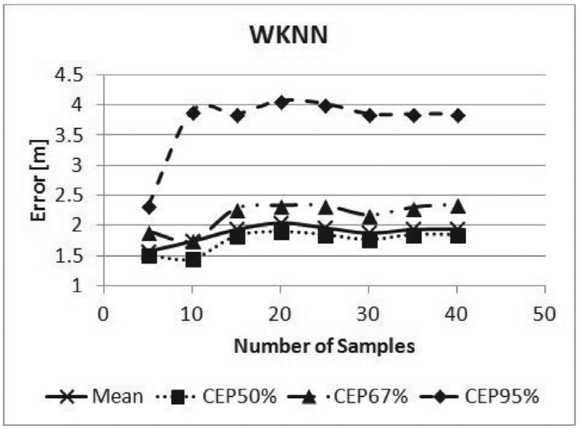


Fig. 8. Radio-map generated through measurements: Effect of sample size on positioning error using WKNN algorithm.

is approximately between 0.5and 4.0 m with some isolated position estimations reaching an error between 7.5 m and 10 m. From the above comparison it is understood that by not selecting an appropriate sample type and size, representative of the population, the outcome of the calibration or evaluation of the RTLS would not be objective.

5.3. Testing on a measured fingerprint data set

Initially, the performance of the SSDA was investigated for the radio-map generated through measurements. WKNN and MMSE positioning algorithms were implemented for different number of sample sizes ranging from $n = 5$ to $n = 40$. The behavior of the mean and Circular Error Probability (CEP50%, CEP67% and CEP95%) is shown in Fig. 8 and Fig. 9, respectively.

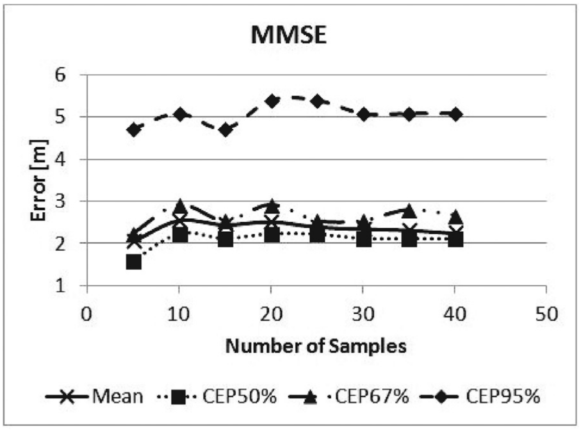


Fig. 9. Radio-map generated through measurements: Effect of sample size on positioning error using MMSE algorithm.

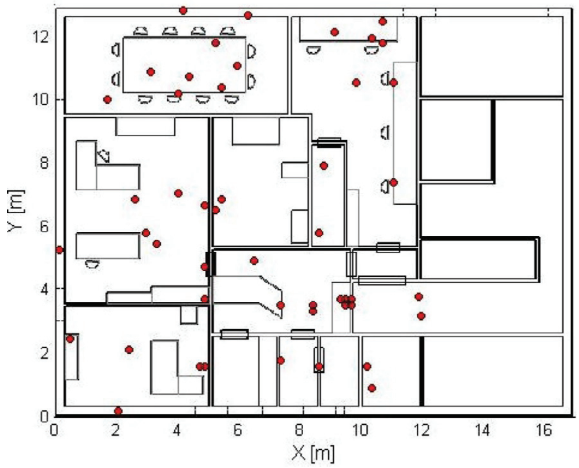


Fig. 10. Typical simple random sample selection of 48 locations for WKNN algorithm – small preliminary sample no 3.

It is observed that, for small sample sizes ($n = 5$ to $n = 30$), the positioning error fluctuates up to 0.5 m for the mean value, 0.7 m for the CEP95% and 1.15 m for the maximum error for this specific environment. The fluctuation tends to stabilize in a smaller range (< 0.4 m) for sample sizes above 35 in this specific radio-map. The observations agree with the calculated n_{SSS} which suggest a size of 48 samples in order to achieve an $EB = \pm 0.4$ m. The recommended n_{SSS} ensures that the performance results can be reliably verified, if a different sample, of the same or larger size, is chosen. A typical simple random selection for the aforementioned scenario is presented in Fig. 10.

5.4. Testing on a simulated fingerprint data set

In order to test SSDA with a larger range of sample sizes, a high resolution radio-map was generated by TruNET, a 3D ray tracing simulator. The grid size in this scenario was set at 10 cm, allowing the utilization of SSDA for *EB* values less that 25 cm as extracted from

Please cite this article as: L. Kanaris et al., Sample Size Determination Algorithm for fingerprint-based indoor localization systems, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2015.12.015

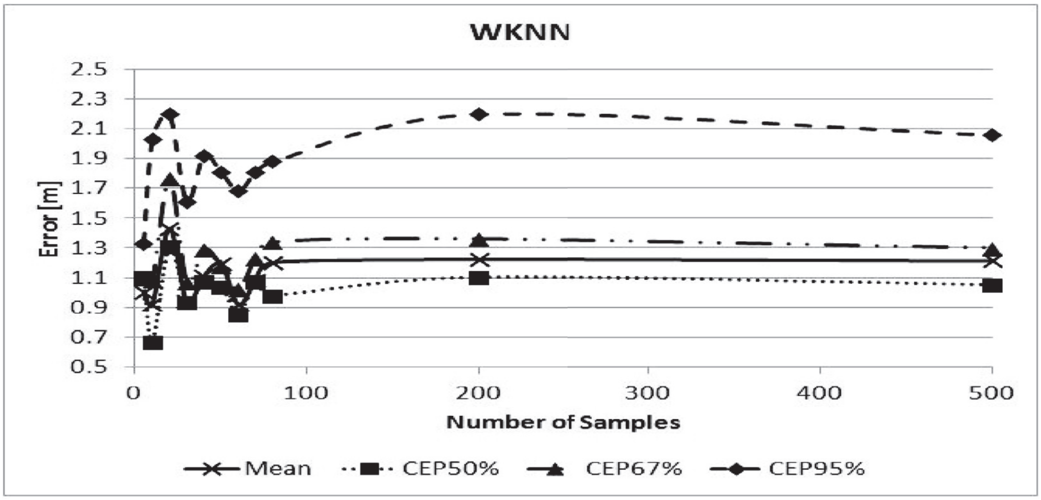


Fig. 11. Radio-map generated by a high-resolution simulation: Effect of sample size on positioning error using WKNN algorithm.

Table 4. A WKNN positioning algorithm was then implemented for sample sizes ranging from 5 to 500. The mean, CEP50%, CEP67% and CEP95% values are presented in Fig. 11.

It is noted that, depending on the sample size, the mean value may fluctuate up to 1.95 m and 2.62 m for CEP95%. It is also observed that the estimated mean stabilizes within a range of $\pm$ 25 cm for sample sizes near $n \simeq 150$. SSDA algorithm recommends the selection of 170 random locations based on a preliminary sample of only 10 points, while ensuring the reliability of any presented performance results or training procedure. It is also proved that any sample size greater than the one suggested, will not affect the outcome, hence such an action will only add unnecessary load.

6. Conclusion

In this paper an algorithm is presented (SSDA) that allows the calculation of a safe margin sample size to be used during the training, calibration and performance evaluation of fingerprint based localization systems. The proposed methodology suggests the utilization of an initial preliminary sample selection, the definition of an acceptable positioning error bound and a predetermined confidence interval. The suggested sample size is then extracted by converting the locations from infinite to discrete and by setting a minimum grid size for the area of interest. Additionally, the importance of selecting a simple random sample is highlighted and compared with a sample of convenience, demonstrating that in the latter case, the results can vary systematically, leading to unreliable conclusions. Finally, SSDA was tested in radio-maps that were generated through measurements and simulations. The outcome indicated that the estimated data sample size, objectively captures the actual system's positioning accuracy performance. The presented work contributes towards the standardization of RTLS evaluation procedures.

References

[1] M. Chen, Towards smart city: M2m communications with software agent intelligence, Multimed. Tools Appl. 67 (1) (2013) 167–178, doi:10.1007/s11042-012-1013-4.
[2] B.C. David S Moore, G.P. MCCabe, Introduction to the Practice of Statistics, seventh ed., W. H. Freeman, 2012.
[3] T. Deasy, W. Scanlon, Simulation or measurement: The effect of radio map creation on indoor WLAN-Based localisation accuracy, Wirel. Pers. Commun. 42 (4) (2007) 563–573.
[4] K. El-Kafrawy, M. Youssef, A. El-Keyi, A. Naguib, Propagation modeling for accurate indoor WLAN rss-based localization, in: Proceedings of the Seventy Second IEEE Vehicular Technology Conference in Fall (VTC 2010-Fall), 2010, pp. 1–5, doi:10.1109/VETECF.2010.5594108.
[5] S.-H. Fang, C.-H. Wang, A dynamic hybrid projection approach for improved wi-fi location fingerprinting, IEEE Trans. Veh. Technol. 60 (3) (2011) 1037–1044, doi:10.1109/TVT.2011.2107757.
[6] A. Hatami, K. Pahlavan, Comparative statistical analysis of indoor positioning using empirical data and indoor radio channel models, in: Proceedings of the Third IEEE Consumer Communications and Networking Conference (CCNC), 2, IEEE, 2006, pp. 1018–1022.
[7] S. Ito, N. Kawaguchi, Data correction method using ideal wireless LAN model in positioning system, in: Proceedings of the 2006 IEEE Seventeenth International Symposium on Personal, Indoor and Mobile Radio Communications, 2006, pp. 1–5, doi:10.1109/PIMRC.2006.254099.
[8] J. Jemai, R. Piesiewicz, T. Kurner, Calibration of an indoor radio propagation prediction model at 2.4 GHz by measurements of the IEEE 802.11b preamble, in: Proceedings of the Third IEEE Sixty First Vehicular Technology Conference (VTC 2005-Spring), 1, 2005, pp. 111–115.
[9] H. Ji, L. Xie, C. Wang, Y. Yin, S. Lu, Crowdsensing: A crowd-sourcing based indoor navigation using rfid-based delay tolerant network, J. Netw. Comput. Appl. 52 (0) (2015) 79–89, doi:10.1016/j.jnca.2015.02.010.
[10] L. Kanaris, A. Kokkinis, A. Liotta, M. Raspopoulos, S. Stavrou, A binomial distribution approach for the evaluation of indoor positioning systems, in: Proceedings of the Twentieth International Conference on Telecommunications (ICT 2013), Casablanca, Morocco, 2013.
[11] A. Kokkinis, M. Raspopoulos, L. Kanaris, A. Liotta, S. Stavrou, Map-aided fingerprint-based indoor positioning, in: Proceedings of the 2013 IEEE Twenty Fourth International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2013, pp. 270–274, doi:10.1109/PIMRC.2013.6666144.
[12] C. Laoudias, M. Michaelides, C. Panayiotou, Fault tolerant positioning using WLAN signal strength fingerprints, in: Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2010, pp. 1–8, doi:10.1109/IPIN.2010.5646216.
[13] B. Li, J. Salter, A.G. Dempster, C. Rizos, Indoor positioning techniques based on wireless lan, in: LAN, First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, 2006, pp. 13–16.

[14] D. Macagnano, G. Destino, G. Abreu, Indoor positioning: A key enabling technology for IoT applications, in: Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), 2014, pp. 117–118, doi:10.1109/WF-IoT.2014.6803131.

[15] D. Macagnano, G. Destino, G. Abreu, Localization with heterogeneous information, in: Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-IoT), 2014, pp. 124–129, doi:10.1109/WF-IoT.2014.6803133.

[16] W. Navidi, Statistics for Engineers and Scientists, McGraw-Hill, 2008.

[17] M. Raspopoulos, C. Laoudias, L. Kanaris, A. Kokkinis, C. Panayiotou, S. Stavrou, 3d ray tracing for device-independent fingerprint-based positioning in WLANs, in: Proceedings of the 2012 Ninth Workshop on Positioning Navigation and Communication (WPNC), 2012, pp. 109–113, doi:10.1109/WPNC.2012.6268748.

[18] M. Raspopoulos, C. Laoudias, L. Kanaris, A. Kokkinis, C. Panayiotou, S. Stavrou, Cross device fingerprint-based positioning using 3D ray tracing, in: Proceedings of the Mobile Computing Symposium (IWCMC2012-Mobile Computing), Limassol, Cyprus, 2012.

[19] T. Roos, P. Myllymäi, H. Tirri, P. Misikangas, J. Sievänen, A probabilistic approach to WLAN user location estimation, Int. J. Wirel. Inf. Netw. 9 (2002) 155–164, doi:10.1023/A:1016003126882.

[20] S. Sorour, Y. Lostanlen, S. Valaee, Rss based indoor localization with limited deployment load, in: Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM), 2012, pp. 303–308, doi:10.1109/GLOCOM.2012.6503130.

[21] S. Sorour, Y. Lostanlen, S. Valaee, K. Majeed, Joint indoor localization and radio map construction with limited deployment load, IEEE Trans. Mob. Comput. 14 (5) (2014) 1031–1043, doi:10.1109/TMC.2014.2343636.

[22] S. Stavrou, S. Saunders, Review of constitutive parameters of building materials, in: Proceedings of the Twelfth International Conference on Antennas and Propagation, 2003.(ICAP 2003), 1, IET, 2003, pp. 211–215.(Conf. Publ. No. 491)

[23] M. Wallbaum, S. Diepolder, Benchmarking wireless lan location systems wireless LAN location systems, in: Proceedings of the Second IEEE International Workshop on Mobile Commerce and Services, 2005. WMCS '05, 2005, pp. 42–51, doi:10.1109/WMCS.2005.7.

[24] B. Wang, S. Zhou, L.T. Yang, Y. Mo, Indoor positioning via subarea fingerprinting and surface fitting with received signal strength, Pervasive Mob. Comput. 23 (C) (2015) 43–58, doi:10.1016/j.pmcj.2015.06.011.

[25] H. Wang, A. Szabo, J. Bamberger, D. Brunn, U. Hanebeck, Performance comparison of nonlinear filters for indoor WLAN positioning, in: Proceedings of the 2008 Eleventh International Conference on Information Fusion, 2008, pp. 1–7.

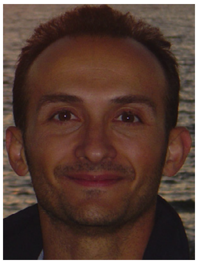
[26] Widyawan, M. Klepal, D. Pesch, Influence of predicted and measured fingerprint on the accuracy of rssi-based indoor location systems, in: Proceedings of the Fourth Workshop Positioning, Navigation and Communication WPNC '07, 2007, pp. 145–151, doi:10.1109/WPNC.2007.353626.

[27] M. Youssef, A. Agrawala, The Horus WLAN location determination system, in: Proceedings of the Third International Conference on Mobile Systems, Applications, and Services, in: MobiSys '05, ACM, New York, NY, USA, 2005, pp. 205–218, doi:10.1145/1067170.1067193.

[28] D. Zhang, L. Yang, M. Chen, S. Zhao, M. Guo, Y. Zhang, Real-time locating systems using active RFID for internet of things, Syst. J. IEEE PP (99) (2014) 1–10, doi:10.1109/JSYST.2014.2346625.

Loizos Kanaris received a B.Sc in Aviation Science at the Hellenic Airforce Academy in Athens, with a diploma thesis in applied aerodynamics ("Computational calculation of 3D flow occurring under rotary wing aircrafts"). After obtaining an MBA, he worked in several research projects of the Cyprus MoD and represented the Republic of Cyprus in several technical telecom workgroups of international organizations, such as the COSPAS-SARSAT and the IMO (COMSAR meetings). He is a member of the R&D department of SIGINT, a telecom research performing organization, which supports the development of several platforms, such as TruNET, a 3D, polarimetric, multithreading Ray Tracing Simulator. He has actively participated in various EU and National research projects, such as WHERE (FP7), WHERE2 (FP7) and LOCME (RPF) investigating wireless positioning aspects. He is currently studying towards a Ph.D (TU/e) concentrating on wireless positioning techniques and the evaluation of fingerprint databases.

Akis Kokkinis received a B.Sc in Aviation Science from the Hellenic Airforce Academy in Athens, with a diploma thesis in computational aerodynamics ("Operational Navigation Flight Planning, including 3D Radar coverage, Radar Prediction and Visualization Features") in 1998. From 2001, he is a member of the R&D department of SIGINT, as a Software Architect and lead developer for several Telecommunication platforms, such as TruNET, a 3D Ray Tracing Simulator. Additionally, he is heavily involved in FP6, FP7 and national ICT funded projects where Sigint Solutions Ltd carries out research work on multiple layers of the OSI protocol stack. Examples of these projects include UNITE (FP6), FUTON (FP7), WHERE (FP7) and COGEU (FP7). He is currently studying towards a Ph.D (TU/e) concentrating on the investigation of enhanced hybrid indoor positioning algorithms.



Giancarlo Fortino received the Laurea (B.S and M.S) and Ph.D in computer engineering from the University of Calabria (Unical), Italy, in 1995 and 2000, respectively. He is currently an Associate Professor of Computer Engineering (since 2006) with the Department of Informatics, Modeling, Electronics and Systems (DIMES) at Unical. He holds the scientific Italian national habilitation for Full Professorship (since 2013) and, since 2015, is Adjunct Professor of Computer Engineering at the Wuhan University of Technology (China) and Adjunct Senior Research Fellow at the Italian National Research Council. He authored more than 275 publications in journals, conferences, and books. His research interests include distributed computing, wireless sensor networks, software agents, cloud computing, and Internet of Things. He is currently co-editor-in-chief of the Springer series on Internet of Things and an associate editor of IEEE Transactions on Affective Computing, IEEE Transactions on Human-Machine Systems, Information Fusion, Engineering Application of Artificial Intelligence, Journal of Network and Computer Applications. He is co-founder and CEO of Sensyscal S.r.l., a Unical spin-off, engaged in advanced applied research and development of IoT-based systems. He is an IEEE senior member.

Antonio Liotta holds the Chair of Communication Network Protocols at the Eindhoven University of Technology (NL), where he leads the Smart Networks team. He is Editor-in-Chief of the book series Internet of Things: Technology, Communications and Computing (Springer) and associate editor of the Journal of Network and System Management (Springer) and of the International Journal of Network Management (Wiley). He investigates topical issues in the area of computer and multimedia networking, with emphasis on cognitive systems for wireless and sensor networks. He is the author of Networks for Pervasive Services: six ways to upgrade the Internet (Springer). His work has recently been featured in IEEE Spectrum in the cognitive Net is coming.

Stavros Stavrou holds an Electrical and Electronics diploma, a B.Eng Degree in Computers and Communications from the University of Essex (U.K.), and a Ph.D degree in Electronic Engineering (Telecommunications) from the University of Surrey (U.K). Between 1997 and 2009 he was a faculty member of the University of Surrey at the Centre for Communication Systems Research (CCSR), formerly Institute for Communication Systems (ICS), where he pursued and led with his research team a number of research activities related to Terrestrial and Satellite telecommunication systems, MIMO and UWB communications. In 2009 he joined the Open University of Cyprus where he currently leads the Telecommunication Systems Research Lab (TSRL). His current research interests span through different topics between layer 1 and 3 of the OSI model, including localization, wireless sensor networks, smart cities and IoT networks. He is also a Fellow of the Higher Education Academy U.K.

Please cite this article as: L. Kanaris et al., Sample Size Determination Algorithm for fingerprint-based indoor localization systems, Computer Networks (2016), http://dx.doi.org/10.1016/j.comnet.2015.12.015