A Taxonomy for Radio Location Fingerprinting

Mikkel Baun Kjærgaard

Department of Computer Science, University of Aarhus,
IT-parken, Aabogade 34, DK-8200 Aarhus N, Denmark
mikkelbk@daimi.au.dk

Abstract. *Location Fingerprinting (LF)* is a promising location technique for many awareness applications in pervasive computing. However, as research on LF systems goes beyond *basic methods* there is an increasing need for better comparison of proposed LF systems. Developers of LF systems are also lacking good frameworks for understanding different options when building LF systems. This paper proposes a taxonomy to address both of these problems. The proposed taxonomy has been constructed from a literature study of 51 papers and articles about LF. For researchers the taxonomy can also be used as an aid when scoping out future research in the area of LF.

1 Introduction

A popular location technique is *Location Fingerprinting (LF)*, having the major advantage of exploiting already existing network infrastructures, like IEEE 802.11 or GSM, which avoids extra deployment costs and effort. Based on a database of pre-recorded measurements of network characteristics from different locations, denoted as *fingerprints*, a wireless client's location is estimated by inspecting currently measured network characteristics. Network characteristics are typically base station identifiers and the received signal strength.

LF is different by the use of fingerprints to other location techniques such as lateration, angulation, proximity detection and dead reckoning [1]. Lateration and angulation techniques estimate location from measurements to fixed points with known locations. A technology example is the *Global Positioning System (GPS)* which estimate a GPS client's location from measurements to GPS satellites with known locations. Proximity detection identifies the location of clients when in proximity of fixed points. A technology example is *Radio-Frequency IDentification (RFID)* where a passive RFID tag's location is known when in proximity of a RFID scanner. Dead reckoning estimates location by advancing previous estimates by known speed, elapsed time and direction. A technology example is dead reckoning based on accelerometer measurements.

Many different LF systems have been proposed. When surveying LF systems one has to answer many different questions. For instance, how do systems differ in scale; can they be deployed to cover a single building or an entire city? What network characteristics are measured? What are the roles of the wireless clients, base stations, and servers in the estimation process? Which estimation method

J. Hightower, B. Schiele, and T. Strang (Eds.): LoCA 2007, LNCS 4718, pp. 139–156, 2007.

is used? How are fingerprints collected and used? These questions are not only important for researchers surveying LF but also developers of LF systems who have to understand the different possibilities. We believe that a taxonomy will aid LF system developers and researchers better survey, compare, and design LF systems. Being able to better survey and compare existing work also makes it possible to use the taxonomy as an aid when scoping out future research. This is especially important as research more and more moves from understanding the basic mechanisms to optimizing existing methods for non-functional properties such as robustness and scalability. Existing taxonomies such as that proposed by Hightower et al. [2] cover location systems in general and are therefore not too much help when answering the many questions specific to LF.

The taxonomy we have chosen to propose has been constructed based on a literature study of 51 papers and articles. The 51 papers and article propose 30 different systems which have been analyzed and methods and techniques grouped to form taxons for the taxonomy. The analyses of four of the 30 systems are covered as case studies in Section 7. The analysis results for all of the 30 systems are available online at [3].

The structure of the paper is as follows. The taxons of the proposed taxonomy are discussed in Section 2. The individual taxons are then presented in Sections 3 to 6. Four case studies are afterwards presented in Section 7 and a discussion is given in Section 8. Finally, conclusions are given in Section 9. Due to the limited size of this paper, the presentation level is advanced; for introductions to LF refer to books such as Küpper [1] and papers such as Krishnakumar et al. [4].

2 Taxonomy

The proposed taxonomy is built around eleven taxons listed with definitions in Table 1. These were partly inspired by earlier work on taxonomies for location systems in general and from our literature study. The four taxons: *scale*, *output*, *measurements*, and *roles* describe general properties of LF systems. We mean by scale the size of the deployment area and by output the type of provided location information. Measurements means the types of measured network characteristics and roles means the division of responsibilities between wireless clients, base stations, and servers. Only these four of our eleven taxons are covered by existing taxonomies such as Hightower et al. [2]. Their concepts for these four taxons differ by output being split over the four concepts of physical, symbolic, absolute, and relative, measurements being indirectly described by their technique concept and roles being partly described by their concept of localized location computation.

Estimation method and *radio map* describe the location estimation process. Estimation method denote a method for predicting locations from a radio map and currently measured network characteristics and radio map a model of network characteristics in a deployment area. The division into estimation method and radio map is used by many papers about LF, for instance Youssef et al. [5].

Table 1. Taxon definitions

Taxon	*Definition*
Scale	Size of deployment area.
Output	Type of provided location information.
Measurements	Types of measured network characteristics.
Roles	Division of responsibilities between wireless clients, base stations, and servers.
Estimation Method	Method for predicting locations from a radio map and currently measured network characteristics.
Radio Map	Model of network characteristics in a deployment area.
Spatial Variations	Observed differences in network characteristics at different locations because of signal propagation characteristics.
Temporal Variations	Observed differences in network characteristics over time at a single location because of continuing changing signal propagation.
Sensor Variations	Observed differences in network characteristics between different types of wireless clients.
Collector	Who or what collect fingerprints.
Collection Method	Procedure used when collecting fingerprints.

However, some papers use a slightly different naming for instance Otsason et al. [6] use *localization algorithm* and *radio map*.

How changing network characteristics over time, space and sensors can be handled is described by *spatial, temporal and sensor variations*. The spatial and temporal dimensions were introduced by Youssef et al. [5]. The sensor dimension was introduced in our earlier work, Kjærgaard [7]. The taxons *collector* and *collection method* describe how fingerprints are collected. These two taxons have been introduced to characterize the assumptions systems put on fingerprint collection.

The focus of the proposed taxonomy is on methods for LF and therefore the taxonomy does not cover evaluation properties for LF systems. Evaluation properties for all kinds of location systems have for instance been suggested by Muthukrishnan et al. [8], who list: precision, accuracy, calibration, responsiveness, scalability, cost, and privacy. The taxonomy proposed by Hightower et al. [2] also lists several evaluation properties: precision, accuracy, scale, cost, and limitations. In our analysis we have included the following evaluation properties: precision, accuracy, evaluation setup, and limitations. These four were chosen because this information is available from most papers. Responsiveness and cost were not included because the first is only available from very few papers and the second from none. Calibration, privacy, scalability, and scale are partly covered by our taxons scale, roles and collection method. These four properties are also listed in our case studies in Section 7.

The taxonomy does not cover non-functional system properties, because work has not yet matured in these directions for LF systems. Non-functional properties of LF systems have been addressed by several recent papers, such as system robustness by Lorincz et al. [9], server scalability by Youssef et al. [5], and minimal communication by Kjærgaard et al. [10]. Also, the taxonomy does not cover the application of LF techniques to other types of sensor measurements such as sound and light.

3 General Taxons

The proposed general taxons for LF systems are: *scale*, *output*, *measurements* and *roles*. These taxons are shown including subtaxons in Figure 1. In this and the following sections when taxons are presented up to four references are given to papers or articles that propose systems that are grouped below the particular taxon. Therefore not all papers groupped under a taxon are listed, this type of information can be found online at [3].

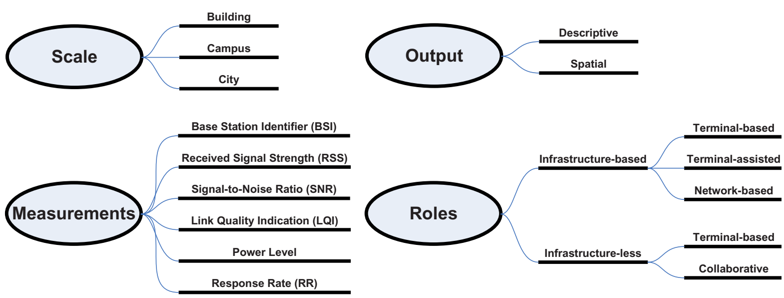

Fig. 1. Scale, output, measurements and roles

Scale describes a system's size of deployment. Scale is important because size of deployment impacts how fingerprints can be collected and some systems are limited in scale because of specific assumptions. Scale is proposed to be classified as *building*, *campus*, or *city*. Many LF systems have been proposed for a *building* scale of deployment [11,12,13,14]. Some systems are limited to this scale because they assume knowledge about the physical layout of buildings [15,16,17,18]; others because they assume the installation of a special infrastructure [19,20]. *Campus*-wide systems [21] scale by proposing more practical schemes for fingerprint collection. *City*-wide systems [22,23,24] scale even further by not assuming that a system is deployed by or for a single organization. City wide systems could scale to any area that is covered by base stations.

Output denotes the type of provided location information. The subtaxons for output are proposed to follow the notion introduced in Küpper [1] of dividing location information into *descriptive* and *spatial* information. Descriptive locations are described by names, identifiers or numbers assigned to natural geographic or man-made objects[1]. Spatial locations are described by a set of coordinates stated with respect to a spatial reference system. Many LF systems output *spatial* locations [11,14,24,25] but systems have also been proposed that output *descriptive*

[1] Some authors refer to this as symbolic locations.

locations [16,18,21]. However, a location outputted as either of the two types can be mapped to the other type given a suitable location model.

Measurements are the types of measured network characteristics. The following network characteristics have been used in existing systems: *Base Station Identifiers (BSI)*, *Received Signal Strength (RSS)*, *Signal-to-Noise Ratio (SNR)*, *Link Quality Indicator (LQI)*, *power level*, and *Response Rate (RR)*. BSI is a unique name assigned to a base station. RSS, SNR, and LQI are signal propagation metrics collected by radios for handling and optimizing communication. The power level is information from the signal sender about current sending power. The response rate is the frequency of received measurements over time from a specific base station. Many LF systems are based on BSI and RSS [11,14,18,25]; other systems have used RR in addition to RSS [15,17,24]. BSI and SNR have also been used [16] and the combination BSI, LQI, RSS, and Power level [9,26].

Roles denote the division of responsibilities between wireless clients, base stations, and servers. How roles are assigned impact both how systems are realized, but also important non-functional properties like privacy and scalability. The two main categories for roles are *infrastructure-based* and *infrastructure-less*. Infrastructure-based systems depend on a pre-installed powered infrastructure of base stations. Infrastructure-less systems consist of ad-hoc-installed battery-powered wireless clients where some of them act as "base stations". Infrastructure-based systems are following Küpper [1], being further divided into *terminal-based*, *terminal-assisted* and *network-based* systems. The infrastructure-less systems are divided into *terminal-based* and *collaborative* systems. The different types of systems differ in who sends out beacons, who makes measurements from the beacons and who stores the radio map and runs LF estimation, as shown in Figure 2. Most LF systems have been built as

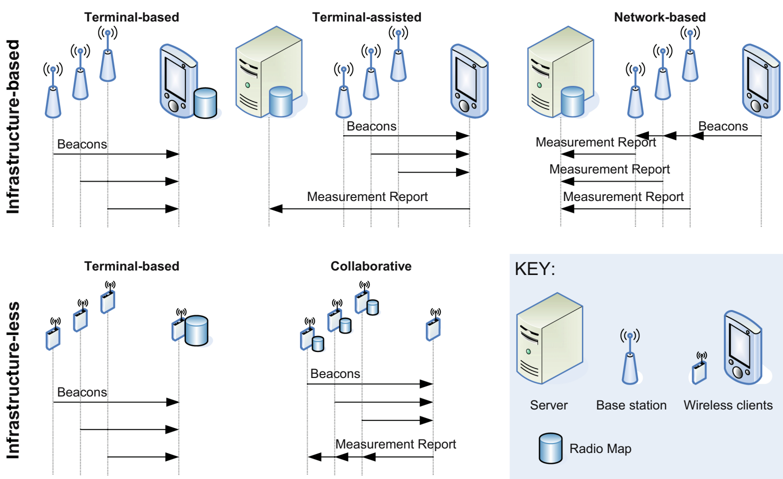

Fig. 2. Different assignments of responsabilities to wireless clients, base stations, and servers

infrastructure-based and terminal-based [5,12,24], which is attractive because this setup supports privacy. Terminal-assisted [16,21] and network-based systems [11,20] have also been built offering better support for resource-weak wireless clients[2]. Infrastructure-less LF-systems have to be optimized for the resource-weak wireless clients, which is addressed by the collaborative setup [9,26].

4 Estimation Taxons

The following two taxons describe the location estimation process: *estimation method* and *radio map*. The two taxons are shown including subtaxons in Figure 3.

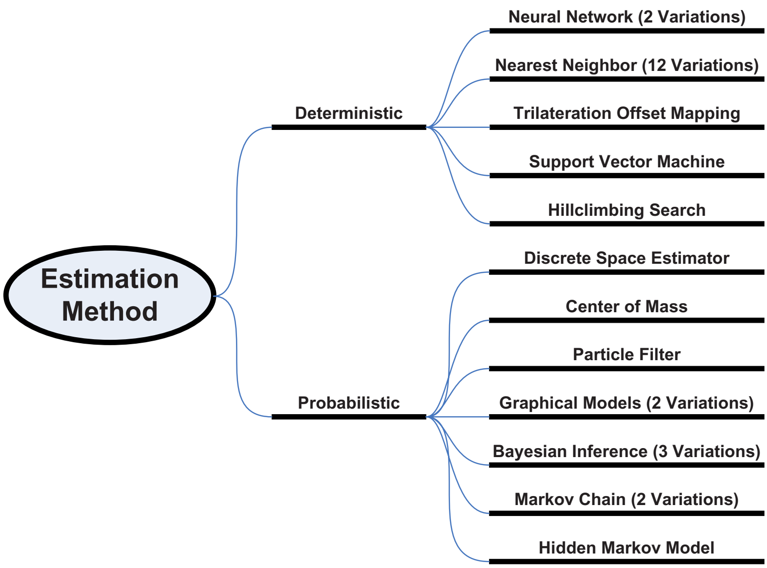

Fig. 3. Estimation method

A central part of a LF system is the *estimation method* used for predicting locations from a radio map and currently measured network characteristics. It would, however, be very challenging to taxonomize all possible methods because nearly all methods developed for machine learning (see Witten et al. [27] for a list of methods) or in the field of estimation (see Crassidis et al. [28] for a list of methods) are applicable to the problem of LF estimation. Here we follow Krishnakumar et al. [4] and divide methods only into deterministic and probabilistic methods. *Deterministic methods* estimate location by considering

[2] However, when only considering the basic method of each system, most can be realized in all of the three setups.

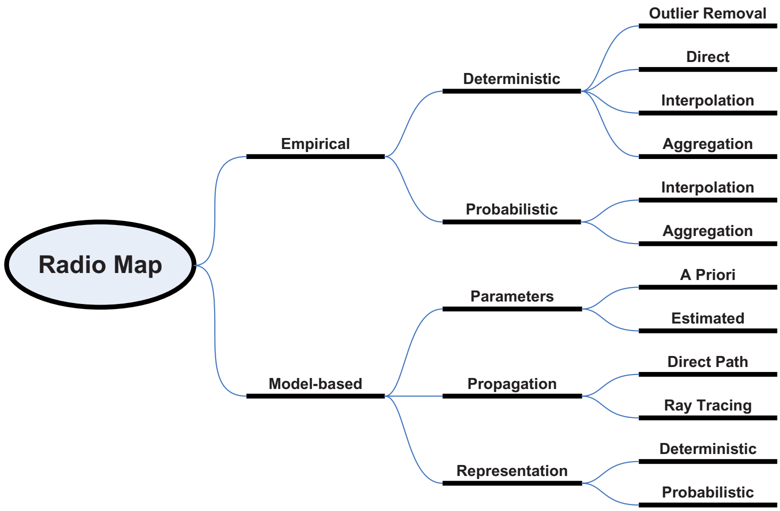

Fig. 4. Radio map

measurements only by their value [11,12,22,25]. *Probabilistic methods* estimate location considering measurements as part of a random process [5,15,16,18]. In Figure 3 examples of applied methods for LF are shown for each of the two categories, including number of identified varieties in our literature study[3]. For example, the classical deterministic technique of Nearest Neighbor was identified during the literature study in twelve different variations. A comment is that many of the studied LF systems use more than one of the listed methods.

A *radio map* provides a model of network characteristics in a deployment area. Radio maps can be constructed by methods which can be classified as either *empirical* or *model-based*. Empirical methods work with collected fingerprints to construct radio maps [5,11,15,18]. Model-based methods use a model parameterised for the LF-system covered area to construct radio maps [11,23,30,31].

Empirical methods can be subdivided into *deterministic* and *probabilistic* methods in the same manner as estimation methods, depending on how they deal with fingerprint-collected measurements. Deterministic methods represent entries in a radio map as single values and probabilistic methods represent entries by probability distributions. Both of these can be further subcategorised into *aggregation* and *interpolation* methods. An aggregation method creates entries in a radio map by summarising fingerprint measurements from a single location [11,14,18,32]. Figure 5 illustrates two aggregation methods for five RSS measurements at two locations marked with a triangle and a square on the

[3] However, even this simple classification is fuzzy for instance when considering the machine learning technique of support vector machines (SVMs) as applied for LF [29]. Because SVMs are defined on a probabilistic foundation but when applied for LF SVMs only consider the actual values of measurements.

figure. The first aggregation method is a deterministic mean method which takes the five measurements and finds the mean and put this value as this location's entry in the radio map. The second aggregation method is a probabilistic Gaussian distribution method which takes the five measurements and fits them to a Gaussian distribution and puts the distribution as the location's entry in the radio map. An interpolation method generate entries in a radio map at unfingerprinted locations by interpolating from fingerprint measurements or radio map entries from nearby locations [15,20,24]. Figure 5 illustrates two interpolation methods at the location marked with a circle using the square-marked and triangle-marked locations as nearby locations. The first interpolation method is a deterministic mean interpolation which finds the mean of nearby radio-map entries and put this value as the entry in the radio map. The second interpolation method is a probabilistic mean method that finds the mean of nearby radio-map entries' gaussian distributions and put the mean distribution as the entry in the radio map. Two other deterministic methods are *outlier removal* filtering away outliers [33] and *direct* creating a radio map using a direct one-to-one mapping to measurements [6].

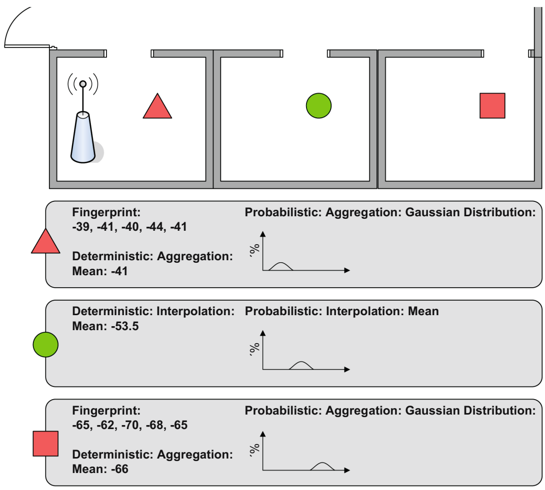

Fig. 5. Deterministic and probabilistic aggregation and interpolation

Model-based methods can be categorized based on how *parameters* for the model are specified, how signal *propagation* is modeled, and what type of *representation* is used by the generated radio map. Parameters can either be given *a priori* [11] or they can be *estimated* from a small set of parameter-estimation fingerprints [31]. Propagation can either be modeled by only considering the *direct path* between a location and a base station [11] or by considering multiple paths categorized as *ray tracing* [31]. The representation of the generated radio map can either be *deterministic* (using single values) [11] or *probabilistic* (using probability distributions) [34].

5 Variation Taxons

The three taxons for variations are: *spatial variations*, *temporal variations*, and *sensor variations*. The three taxons are shown including subtaxons in Figure 6.

Spatial variations are the observed differences in network characteristics at different locations because of signal propagation characteristics. Because of how signals propagate even small movements can create large variations in the measured network characteristics. The main method for addressing spatial variations is *tracking*: the use of constraints to optimize sequential location estimates. Tracking can be based on motion in terms of target *speed* [24,35], target being *still versus moving* [15], and knowledge about motion *patterns* [35]. Tracking can also be based on physical constraints such as how *connections* exist between locations [16] and the *distance* between them [15,19]. Tracking using one or several of the listed constraints is implemented using an estimation method (such as the ones listed in Section 4) that is able to encode the constraints. Spatial variations can also be addressed by *base station selection*, *fingerprint filtering*, and *sample perturbation*. Base station selection filters out measurements to base stations that are likely to decrease precision and accuracy [36,37]. Fingerprint filtering limits the set of used fingerprints to only those that are likely to optimize precision and accuracy [37]. *Sample perturbation* apply perturbation of measurements to mitigate spatial variations [5].

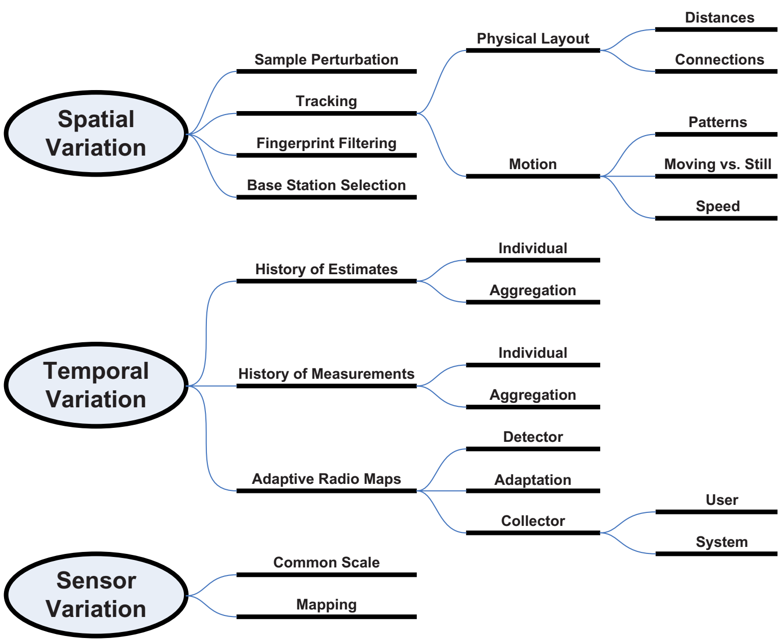

Fig. 6. Spatial variations, temporal variations, and sensor variations

Temporal variations are the observed differences in network characteristics over time at a single location because of continuing changing signal propagation. On a large-scale, temporal variations are the prolonged effects observed over larger periods of time such as day versus night. On a small-scale, temporal variations are the variations implied by quick transient effects, such as a person walking close to a client. Methods for handling temporal variations can be divided into methods that are based on a *history of estimates*, a *history of measurements*, or *adaptive radio maps*. A history of either measurements or estimates here denotes a set of estimates or measurements inside a defined time window. The alternative to a history is only to use the most recent estimate or measurements. The history of either measurements or estimates can either be used as *individual* [15,18] measurements or estimates or, using some *aggregation* [5,14], can be combined to one measurement or estimate. The adaptive radio map method introduces the idea of handling temporal variations by making the radio map adapt to the current temporal variations [19,20,32]. For this idea to work, some *collector* has to make measurements that can be used by a *detector* to control if some adaptation should be applied to the current radio map. The measurements can either be collected from the measurements a *user* collects [32] to run LF estimation on or it can be collected by some specially-installed *system* infrastructure [19,20].

Sensor variations are the observed differences in network characteristics between different types of wireless clients. On a large-scale, variations can be observed between clients from different manufactures. On a small-scale, variations can be observed between different examples of similar clients. One method for addressing sensor varations is to define a *common scale* and then, for each type of sensor, find out how this sensor's measurements can be converted to the common scale. A second approach is to use a single sensor to fingerprint with and then find a mapping from new sensors to the sensor that was used for fingerprinting [7,18].

6 Collection Taxons

The two taxons for fingerprint collection are *collector* and *collection method* as shown in Figure 7.

Collector describes who or what collect fingerprints. There are three categories: *user*, *administrator*, and *system*. A user is a person who is either tracked by or uses information from a LF system [21,24]. An administrator is a person who manages a LF system [11,18,38] and a system is a specially-installed infrastructure for collecting fingerprints [20].

The fingerprints are collected following some *collection method*. A collection method places assumptions on if fingerprints are collected on a *location* that is either *known* [6] or *unknown* [34,35]. If fingerprints are collected to match a *spatial property* such as: *orientation* [11], at a *point* [15], covering a *path* [24], or covering an *area* [18,36]. If the collected *number of measurements* for each fingerprint is *fixed* [5,14] or determined based on some *adaptive* strategy.

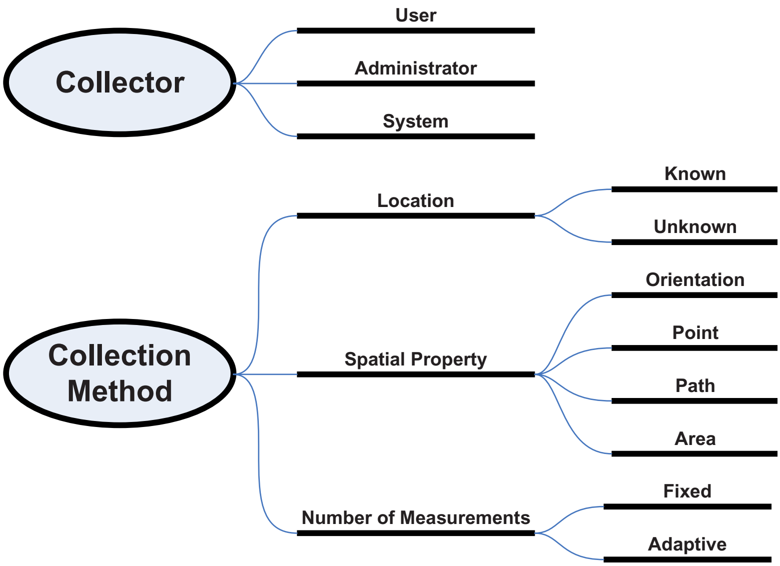

Fig. 7. Collector and collection method

7 Case Studies

To show the use of the proposed taxonomy, this section presents our analysis using the taxonomy on four of the 30 different systems identified in the literature study. Figure 8 shows the analysis results in a compact form. The four systems have been selected to highlight different parts of the taxonomy. As mentioned earlier, the analysis of the rest of the analyzed systems are available online at [3] in a similar format. In addition to the eleven taxons, four extra categories describe the systems from an evaluation perspective; these are: *accuracy, precision, evaluation setup and limitations*. The listed evaluation results have been taken from the original papers. Evaluation setup is grouped into *stationary* (meaning that the authors' test data was collected while keeping a wireless client at a static position) or *moving* (for which the wireless client was moved around mimicking normal use).

The RADAR system proposed by Bahl et al. [11] is aimed at a building scale of deployment and provides spatial locations as output. The system measures BSI, and RSS for the WaveLAN technology and roles are assigned as infrastructure-based: network. The estimation method is the deterministic k-nearest neighbor algorithm. They propose two setups, here named A and B. For A the radio map is constructed using deterministic aggregation using the mean from empirical-collected fingerprints. For B the radio map is deterministically constructed model-based considering the direct path of transmission using a priori parameters. For A an administrator will collect fingerprints at known locations standing at one point with different orientations collecting a fixed number of measurements and for B no fingerprints are collected. A limitation for setup B is that knowledge is needed of spatial locations of base stations and walls.

	Bahl et al. (2000): RADAR	**Youssef et al. (2003,…,2005): Horus**	**LaMarca et al. (2005): Place Lab**	**Lorincz et al. (2005): MoteTrack**
Scale	Building	Building	City	Building
Output	Spatial Locations	Spatial Locations	Spatial Locations	Spatial Locations
Measurements	BSI, RSS (WaveLan)	BSI, RSS (IEEE 802.11)	BSI, RSS, RR (IEEE 802.11 & GSM)	**A**: BSI, Power Level, RSS: (916 MHz FSK) **B**: BSI, LQI, RSS: (IEEE 802.15.4)
Roles	Infrastructure-based: Network	Infrastructure-based: Terminal	Infrastructure-based: Terminal	Infrastructure-less: Collaborate
Estimation Method	Deterministic: K-Nearest Neighbor	Probabilistic: [Discrete Space Estimator, Center of Mass]	Probabilistic: Particle Filter	Ratio-Nearest Neighbor (Manhattan Distance)
Radio Map	**A:** Empirical: Deterministic: Aggregation: Mean **B:** Model-based: [Parameters: A priori, Propagation: Direct Path: Transmission, Representation: Deterministic]	Empirical: Probabilistic: Aggregation: [Histogram Method, Kernel Distributions, Correlation Modeling]	Empirical: Deterministic: Interpolation: Mean, Probabilistic: Interpolation: Histogram Method	Empirical: Deterministic: Aggregation: Mean
Spatial Variation		Sample Perturbation	Tracking: Motion: Speed	
Temporal Variation	History of Measurements: Aggregation: Mean	History of Estimates: Aggregation: Mean History of Measurements: Aggregation: Mean		
Sensor Variation				
Collector	Administrator	Administrator	Users	Administrator
Collection Method	**A:** Location: Known, Spatial Property: [Point, Orientation], Number of Measurements: Fixed **B:** None	Location: Known, Spatial Property: Point, Number of Measurements: Fixed	Location: Known, Spatial Property: Path, Number of Measurements: Fixed	Location: Known, Spatial Property: Point, Number of Measurements: Fixed
Precision	**A:** 2.75m (k=5) **B:** 4.3m (k=1)	**Site 1**: 0.39m **Site 2**: 0.51m	**Urban**: 21.8m **Residential**: 13.4m **Suburban**: 31.3m	**A**: 2m **B**: 0.9m
Accuracy	50%	50%	50%	50%
Evaluation Setup	Stationary: See website for details	Stationary: See website for details	Moving: See website for details	Stationary: See website for details
Limitations	**B:** Spatial locations of base stations and walls		GPS (and car) for collecting fingerprints	Deployment of beacon nodes

Fig. 8. Analysis results for the four case studies

The Horus system proposed by Youssef et al. [5,39,40,41,42] also aims at a building scale of deployment and provide spatial locations as output. The system measures BSI, and RSS for the IEEE 802.11 technology and the assigned roles match infrastructure-based: terminal. The estimation method is a combination of two probabilistic techniques: discrete space estimator and center of mass. The radio-map is built using probabilistic aggregation, either based on a histogram method or on a kernel distribution method; in addition, a method for correlation modeling is also applied. To handle spatial variations sample perturbation is applied and temporal variations are handled by both mean aggregating measurements and estimates. An administrator collects fingerprints at known locations standing at one point collecting a fixed number of measurements.

The Place Lab system proposed by LaMarca et al. [24,43,44] aims at a city-wide deployment and provides spatial locations as output. The system measures BSI, RSS, and RR for both IEEE 802.11 and GSM and the assigned roles match infrastructure-based: terminal. The most advanced of the system's estimation methods uses a particle filter. The radio map is built in two steps, first applying deterministic interpolation based on means and then probabilistic interpolation based on the histogram method. Spatial variations are addressed by tracking based on motion by speed constraints. The fingerprints are user collected based on paths with known location with a fixed number of measurements. A limitation is that a GPS device (and a car) is needed to practically collect fingerprints.

The MoteTrack system proposed by Lorincz et al. [9,26] targeted for sensor networks aims at building-scale deployment and provides spatial locations as output. The system has been tested in two setups, here named A and B. Setup A measures BSI, Power level, and RSS for 916 MHz FSK communication and setup B measures BSI, LQI, and RSS for IEEE 802.15.4. The roles are assigned matching infrastructure-less: collaborate with beacon nodes taking the role as base stations. The estimation method is ratio-nearest neighbor with Manhattan distance to lower computational needs. The radio map is constructed using deterministic aggregation using the mean from empirical-collected fingerprints. An administrator collects fingerprints at known locations standing at one point collecting a fixed number of measurements. A limitation is the needed deployment and maintenance of beacon nodes.

8 Discussion

During the literature study both many similarities and differences were identified between studied systems. This can be seen from just the four included case studies in Section 7. For instance, the well-known nearest-neighbor estimation method were identified in many variations of the basic method. The differences were not only in terms of improvements to the basic estimation method but also how systems address spatial and temporal variations. One system use a history of measurements and mean-aggregate them before applying nearest neighbor [11]. Another system use the measurements directly and use a history of estimates and aggregate these instead [36]. By using the proposed taxonomy these differences become clear when classifying systems. Another example also for systems based on nearest neighbor is how the radio map is built. For instance Krishnan et al. [20] builds the radio map by applying advanced aggregation and interpolation methods where as the original system proposed by Bahl et al. [11] only use a simple aggregation based on mean values. The taxonomy also here creates a better starting point when comparing and evaluating systems.

To use the proposed taxonomy for comparison too a new system, one approach would be to, first, find classifications for compared-to existing systems. As mentioned earlier a starting point for finding such classifications is to look at our classifications online at [3]. Second, one would make a classification for the new system by classifying for each of the eleven taxons the new system's methods and assumptions according to the subtaxons. Third, one would make the comparison of the new and the existing systems. For evaluation of LF systems the taxonomy can also be used to highlight the evaluated system's assumptions and methods. This can be done by providing a classification for the evaluated system which makes it explicit what methods and assummptions are evaluated. For instance, as mentioned in the discussion above many systems have been evaluation in comparison to the nearest neighbor estimation method. But this estimation method has been implemented with many different choices when considering the used radio map and methods for addressing spatial and temporal variations. This means that it is not the same baseline method that is compared-to making results incomparable.

The taxonomy can also help scoping out future research by illustrating what research topics have not yet been covered. One way to analyse this is to group systems in terms of some of the taxons. A grouping for the taxons scale and radio map is shown in Table 2. The table shows that only one system aims at a campus-size scale was identified. The table also shows that generally systems either use empirical or model-based radio maps not a combination. So an open research topic is exploring the boundary between building and city-wide systems maybe by combining empirical and model-based radio maps. A grouping for the taxons spatial and temporal variations is also shown in Table 3. The table shows that for these taxons most systems only address one of the variations. Few systems combine them and several combinations of the different methods remain unexplored.

We do not expect that the proposed taxonomy is complete in its current form. Instead, it is intended to enable better and more complete understanding of LF and to evolve as that understanding improves. At the same time, we feel that our eleven main taxons and many of the subtaxons are fairly stable. During the process of creating the taxonomy, analyzing papers and classifying systems, we

Table 2. Grouping in terms of scale and radio map

	Empirical	*Model-based*
Building	[5, 6, 9, 11–13, 15–20, 23, 25, 29, 32, 33, 35–38, 45–47]	[11, 29–31, 34, 46]
Campus	[21]	
City	[22,24]	[14]

Table 3. Grouping in terms of spatial and temporal variations

	None	*History of Measurements*	*History of Estimates*	*Adaptive Radio Maps*
None	[6, 9, 12–14, 21, 29, 30, 34]	[11, 22, 25, 33, 46]	[23]	[20, 47]
Sample Perturbation		[5]	[5]	
Tracking	[16, 24, 31, 32, 38, 45]	[18,19,35]	[15,17,18]	[18,19]
Fingerprint Filtering	[37]			
Base Station Selection	[37]			

found that all 30 systems and their methods could be classified. On the other hand, some of the subtaxons are likely to evolve as our understanding of LF evolves. An area for which it would be interesting to extend the taxonomy is for non-functional properties as mentioned in Section 2. However, only a limited number of papers have so far been published in this direction [5,9,10].

9 Conclusion

This paper presented a taxonomy for location fingerprinting. The proposed taxonomy was constructed from a literature study of 51 papers and articles about LF. The taxonomy consists of the following eleven taxons: *scale*, *output*, *measurements*, *roles*, *estimation method*, *radio map*, *spatial variations*, *temporal variations*, *sensor variations*, *collector*, and *collection method*. The 51 analyzed papers described 30 LF systems of which four were presented as case studies.

Valuable taxonomies can account for everything that is known so far and can predict things to come, as variations of parameters accounted for and enumerated in the taxonomy. A taxonomy first and foremost shows the depth and the breadth of our understanding. We would like others to join and based on inputs from the community further improve the proposed taxonomy.

Acknowledgements

The author would like to thank Doina Bucur, Azadeh Kushki and the reviewers for their insightful comments on earlier drafts of this paper. The research reported in this paper was partially funded by the software part of the ISIS Katrinebjerg competency centre http://www.isis.alexandra.dk/software/.

References

1. Küpper, A.: Location–based Services — Fundamentals and Operation. John Wiley & Sons, Chichester (2005)
2. Hightower, J., Borriello, G.: Location Systems for Ubiquitous Computing. Computer 34(8), 57–66 (2001)

3. Website, http://wiki.daimi.au.dk/mikkelbk
4. Krishnakumar, A.S., Krishnan, P.: The Theory and Practice of Signal Strength-Based Location Estimation. In: Proceedings of the First International Conference on Collaborative Computing: Networking, Applications and Worksharing (2005)
5. Youssef, M., Agrawala, A.: The Horus WLAN Location Determination System. In: Proceedings of the Third International Conference on Mobile Systems, Applications, and Services (2005)
6. Otsason, V., Varshavsky, A., Marca, A.L., de Lara, E.: Accurate GSM Indoor Localization. In: Proceedings of the Seventh International Conference on Ubiquitous Computing (2005)
7. Kjærgaard, M.B.: Automatic Mitigation of Sensor Variations for Signal Strength Based Location Systems. In: Second International Workshop on Location- and Context-Awareness (2006)
8. Muthukrishnan, K., Lijding, M., Havinga, P.: Towards Smart Surroundings: Enabling Techniques and Technologies for Localization. In: Proceedings of the First International Workshop on Location- and Context-Awareness (2005)
9. Lorincz, K., Welsh, M.: MoteTrack: A Robust, Decentralized Approach to RF-Based Location Tracking. In: Proceedings of the First International Workshop on Location- and Context-Awareness (2005)
10. Kjærgaard, M., Treu, G., Linnhoff-Popien, C.: Zone–based RSS Reporting for Location Fingerprinting. In: Proceedings of the 5th International Conference on Pervasive Computing (2007)
11. Bahl, P., Padmanabhan, V.N.: RADAR: An In-Building RF-based User Location and Tracking System. In: Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, IEEE Computer Society Press, Los Alamitos (2000)
12. Prasithsangaree, P., Krishnamurthy, P., Chrysanthis, P.: On Indoor Position Location with Wireless LANs. In: Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE Computer Society Press, Los Alamitos (2002)
13. Battiti, R., Villani, A., Nhat, T.L.: Neural network models for intelligent networks: deriving the location from signal patterns. In: Proceedings of the First Annual Symposium on Autonomous Intelligent Networks and Systems (2002)
14. Roos, T., Myllymäki, P., Tirri, H.: A Statistical Modeling Approach to Location Estimation. IEEE Transactions on Mobile Computing 1(1), 59–69 (2002)
15. Krumm, J., Horvitz, E.: LOCADIO: Inferring Motion and Location from Wi-Fi Signal Strengths. In: Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (2004)
16. Castro, P., Chiu, P., Kremenek, T., Muntz, R.: A Probabilistic Room Location Service for Wireless Networked Environments. In: Proceedings of the Third International Conference on Ubiquitous Computing (2001)
17. Ladd, A.M., Bekris, K.E., Rudys, A., Marceau, G., Kavraki, L.E., Wallach, D.S.: Robotics-Based Location Sensing using Wireless Ethernet. In: Proceedings of the Eight ACM International Conference on Mobile Computing and Networking, ACM Press, New York (2002)
18. Haeberlen, A., Flannery, E., Ladd, A.M., Rudys, A., Wallach, D.S., Kavraki, L.E.: Practical Robust Localization over Large-Scale 802.11 Wireless Networks. In: Proceedings of the Tenth ACM International Conference on Mobile Computing and Networking, ACM Press, New York (2004)

19. Bahl, P., Padmanabhan, V.N., Balachandran, A.: A Software System for Locating Mobile Users: Design, Evaluation, and Lessons. Microsoft Research Technical Report MSR-TR-2000-12, Microsoft (2000)
20. Krishnan, P., Krishnakumar, A.S., Ju, W.H., Mallows, C., Ganu, S.: A System for LEASE: Location Estimation Assisted by Stationary Emitters for Indoor RF Wireless Networks. In: Proceedings of the 23th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, IEEE Computer Society Press, Los Alamitos (2004)
21. Bhasker, E.S., Brown, S.W., Griswold, W.G.: Employing User Feedback for Fast, Accurate, Low-Maintenance Geolocationing. In: Proceedings of the 2nd IEEE International Conference on Pervasive Computing and Communications, IEEE Computer Society Press, Los Alamitos (2004)
22. Laitinen, H., Lahteenmaki, J., Nordstrom, T.: Database Correlation Method for GSM Location. In: Proceedings of the 53rd Vehicular Technology Conference (2001)
23. Roos, T., Myllymäki, P., Tirri, H., Misikangas, P., Sievänen, J.: A Probabilistic Approach to WLAN User Location Estimation. International Journal of Wireless Information Networks 9(3), 155–164 (2002)
24. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., Schilit, B.: Place Lab: Device Positioning Using Radio Beacons in the Wild. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, Springer, Heidelberg (2005)
25. Smailagic, A., Kogan, D.: Location Sensing and Privacy in a Context-aware Computing Environment. IEEE Wireless Communications 9(5), 10–17 (2002)
26. Lorincz, K., Welsh, M.: MoteTrack: a robust, decentralized approach to RF-based location tracking. Personal and Ubiquitous Computing
27. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
28. Crassidis, J.L., Junkins, J.L.: Optimal Estimation of Dynamic Systems. Chapman & Hall/CRC Press, Boca Raton, FL (2004)
29. Brunato, M., Battiti, R.: Statistical learning theory for location fingerprinting in wireless LANs. Comput. Netw. ISDN Syst. 47(6), 825–845 (2005)
30. Wallbaum, M., Wasch, T.: Markov Localization of Wireless Local Area Network Clients. In: Proceedings of the First IFIP TC6 Working Conference on Wireless On-Demand Network Systems (2004)
31. Ji, Y., Biaz, S., Pandey, S., Agrawal, P.: ARIADNE: A Dynamic Indoor Signal Map Construction and Localization System. In: Proceedings of the 4th international conference on Mobile systems, applications and services (2006)
32. Berna, M., Sellner, B., Lisien, B., Thrun, S., Gordon, G., Pfenning, F.: A Learning Algorithm for Localizing People Based on Wireless Signal Strength that Uses Labeled and Unlabeled Data. In: Proceedings of the International Joint Conference on Artificial Intelligence (2003)
33. Saha, S., Chaudhuri, K., Sanghi, D., Bhagwat, P.: Location Determination of a Mobile Device using IEEE 802.11b Access Point Signals. In: Proceedings of the IEEE Wireless Communications and Networking Conference, IEEE Computer Society Press, Los Alamitos (2003)
34. Madigan, D., Einahrawy, E., Martin, R.P., Ju, W.H., Krishnan, P., Krishnakumar, A.S.: Bayesian Indoor Positioning Systems. In: Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, IEEE Computer Society Press, Los Alamitos (2005)

35. Chai, X., Yang, Q.: Reducing the Calibration Effort for Location Estimation Using Unlabeled Samples. In: Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications, IEEE Computer Society Press, Los Alamitos (2005)
36. Varshavsky, A., Lamarca, A., Hightower, J., de Lara, E.: The SkyLoc Floor Localization System. In: Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications, IEEE Computer Society Press, Los Alamitos (2007)
37. Kushki, A., Plataniotis, K.N., Venetsanopoulos, A.N.: Kernel-Based Positioning in Wireless Local Area Networks. Mobile Computing, IEEE Transactions on 6(6), 689–705 (2007)
38. Seshadri, V., Zaruba, G.V., Huber, M.: A Bayesian Sampling Approach to In-door Localization of Wireless Devices using Received Signal Strength Indication. In: Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications, IEEE Computer Society Press, Los Alamitos (2005)
39. Youssef, M., Agrawala, A.: Small-Scale Compensation for WLAN Location Determination Systems. In: Proceedings of the IEEE Conference Wireless Communications and Networking, IEEE Computer Society Press, Los Alamitos (2003)
40. Youssef, M.A., Agrawala, A., Shankar, U.A.: WLAN Location Determination via Clustering and Probability Distributions. In: Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, IEEE Computer Society Press, Los Alamitos (2003)
41. Youssef, M., Agrawala, A.: Handling Samples Correlation in the Horus System. In: Proceedings of the 23th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, IEEE Computer Society Press, Los Alamitos (2004)
42. Youssef, M., Abdallah, M., Agrawala, A.: Multivariate Analysis for Probabilistic WLAN Location Determination Systems. In: Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (2005)
43. Cheng, Y.C., Chawathe, Y., Lamarca, A., Krumm, J.: Accuracy Characterization for Metropolitan-scale Wi-Fi Localization. In: Proceedings of the 3rd international conference on Mobile systems, applications, and services (2005)
44. Hightower, J., Borriello, G.: Particle Filters for Location Estimation in Ubiquitous Computing: A Case Study. In: Davies, N., Mynatt, E.D., Siio, I. (eds.) UbiComp 2004. LNCS, vol. 3205, Springer, Heidelberg (2004)
45. Agiwal, A., Khandpur, P., Saran, H.: LOCATOR: location estimation system For wireless LANs. In: Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots, ACM Press, New York (2004)
46. Elnahrawy, E., Li, X., Martin, R.P.: The Limits of Localization Using Signal Strength: A Comparative Study. In: First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, IEEE Computer Society Press, Los Alamitos (2004)
47. Yin, J., Yang, Q., Ni, L.: Adaptive Temporal Radio Maps for Indoor Location Estimation. In: Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications, IEEE Computer Society Press, Los Alamitos (2005)