Location Cognition for Wireless Systems: Classification with Confidence

Stefan Aust[1], Tetsuya Ito[1], and Peter Davis[2]

[1] NEC Communication Systems, Ltd., Kanagawa, Japan
aust.st@ncos.nec.co.jp, ito.tts@ncos.nec.co.jp
[2] Telecognix Corporation, Kyoto, Japan
davis@telecognix.com

Abstract. Location cognition is a challenging task in cognitive wireless systems when there is no explicit location information system available, such as Global Positioning System (GPS) or dense wireless beacons. This paper decribes a simple-but-effective method of real-time location cognition which can be used by wireless devices in WLAN systems without depending on any location service infrastructure. The method is based on monitoring, learning and recognizing the statistics of received data traffic, with an awareness of the confidence in the recognition result. It uses the property that traffic statistics such as average and variance of throughput are correlated with the location of the transmission. Locations are recognized by comparing monitored statistics with a set of reference distributions and identifying the best match. A measure of the confidence in the location classification result is obtained by comparing matches with multiple candidate locations. It is demonstrated that the method can be implemented as middleware for use with WLAN devices and used to recognize multiple locations, indoor and outdoor. It is also demonstrated that the method can be used to detect the distance between a sender and receiver.

Keywords: monitoring, statistics, fingerprinting, classification, location cognition, confidence, cognitive radio systems.

1 Introduction

Acquiring location information in cognitive wireless systems is a required function in cognitive radio (CR) systems [1]. The FCC has recently regulated that TV white space CRs need to know their geo-location and avoid interfering as secondary user (SU) with the primary user (PU) by considering the coverage area of TV broadcast services. It is standardized using GPS service for positioning with an extension for systems which are not always able to acquire position information [2]. How a CR system should work when there is poor reception of GPS signals, for example at indoor locations, remains open. Indoor as well as outdoor location information is important for many other wireless functions, such as navigation, context-aware applications [3], pervasive computing [4] and movements of users, such as fire fighters [5]. GPS is widely used but can be

Y. Cai et al. (Eds.): Mobilware 2010, LNICST 48, pp. 144–157, 2010.

applied for outdoor applications only. Other methods exist which complement the use of GPS for providing location information for example, using beacons, triangulation or a centralized database (DB) of so-called fingerprints of signals from surrounding access points (APs).

We propose a novel approach for location cognition which can be independent of GPS, APs or any centralized DB. In particular, we propose a novel location cognition method for multiple locations based on monitoring, learning and recognizing the characteristics of data transmission between radio terminals. This is a form of so-called *fingerprinting*, where the statistics which are obtained during an off-line "learning" phase are used to identify a location. Our method includes an estimation of the confidence in the location cognition, which enhances the use of the method in practical scenarios. We show that location cognition can be done with confidence either inside or outside a building, using reference distributions held by each device and without using additional infrastructure. In particular, we show that the method can be implemented as wireless middleware, which we call a Location Cognition Engine (LOC), supporting WLAN IEEE 802.11 standards. We show that it can be used to classify multiple indoor and outdoor locations, and also recognize distances between sender and receiver terminals.

2 Related Work

With the use of the Global Positioning System (GPS) in outdoor environments the location problem can be easily solved. GPS is widely accepted as a useful positioning system and is applied in car navigation systems and military scenarios. However, even GPS or Galileo may have situations where shadowing or interference reduces the accuracy of these services [5]. Work has started that aims to extend GPS/Galileo to improve the positioning in outdoor and indoor environments [6]. The work aims to combine the GPS/Galileo system with other location information sources to fulfill location detection requirements efficiently. Other solutions and methods have been developed for indoor localization, using WLAN, infrared (IR), ultrasound, FM radio, fingerprinting, sensor networks, ultra-wideband (UWB), Bluetooth, magnetic signals, vision analysis and triangulation, etc. Each technology has unique advantages in performing location sensing, but also has some intrinsic limitations. Localization methods which combine one or more of these technologies to increase location performance have also been proposed [5].

Comparing with outdoor, indoor location is usually more difficult due to the complexity of the surrounding physical environment, such as walls and doors, which influence the propagation of electromagnetic waves and can result in complex multi-path effects. A comprehensive study which is given in [7] shows details of indoor positioning systems (IPS) for wireless personal networks, addressing security, privacy, cost, performance, robustness, complexity and limitations. The authors in [7] pointed out that for indoor applications new challenges arise for IPS. IPS considers only indoor environments such as inside a building and extensions for location cognition in outdoor environments are not included. The authors classify four techniques for indoor positioning estimations, triangulation, fingerprinting, proximity and vision analysis, and show that each technique has its limitations, but combinations of positioning systems can significantly improve the quality of position estimates. In [8] an FM indoor positioning system (FINDR) was proposed. The power consumption of the proposed location method uses 15mW for FM transmission whereas for WLAN about 300mW is used which was the motivation for the authors to utilize FM as an energy efficient indoor positioning system. The location method uses the signal-to-noise ratio and the received signal strength of FM transmissions. FM transmitters have to be pre-installed and manually tuned to broadcast-free frequencies. A k-nearest neighbor classifier was applied for position classification. The accuracy of this system was reported to be 4.5m (at 95% confidence).

Fingerprinting methods for indoor positioning are widely applied [9], [10]. Fingerprinting positioning technique uses pre-measured location related data, including two phases, an off-line training phase and an on-line phase. The authors in [11] proposed an enhanced fingerprint-type technique based on trilateration through Received Signal Strength (RSS) values obtained in real-time in indoor locations. The method estimates the propagation models that best fit the propagation environments. The authors conducted a vast amount of measurements to obtain accurate values for their proposed RSS log-normal path-loss model that uses a constant value which depends on averaged fast and slow fading, gains and transmitted power. The authors concluded that this constant is known beforehand and should be valid for different environments. We conclude that none of the proposed location methods provides a solution for a combined location cognition in indoor and outdoor environments.

3 Proposal: Location Cognition Engine (LOC)

We aim to provide a versatile Location Cognition Engine (LOC) which is user friendly and reliable in the cognition of both indoor and outdoor locations. The original purpose of the Location Cognition Engine (LOC) was the use in cognitive wireless systems, to support autonomous selection of wireless channels avoiding interfered or busy channels. However, the LOC could also be integrated in various other kinds of services and applications which require location awareness. In order to achieve versatile, reliable and user-friendly location cognition, we propose a method based on monitoring, learning and recognizing the characteristics of data transmission between radio terminals, combined with estimates of the confidence in the learnt reference data and the result of the location recognition.

We show that a wireless device can recognize its location on-line by comparing the statistical distribution of its data transmissions with a set of previously (i.e., off-line) acquired reference distributions. This is a form of so-called *fingerprinting*, where the statistics which are obtained during an off-line "learning" phase are used to identify a location. In comparison to other fingerprinting methods which use RSS fingerprints of specific positions of multiple indoor WLAN APs to obtain a detailed signal strength map, our proposed location cognition is fully independent of any pre-installed WLAN APs. Another difference is that the fingerprints are based on high-level transmission characteristics, rather than physical layer information such as signal strength or delay. Moreover, we introduce parameters to indicate confidence in the set of fingerprints and the confidence in the best fit.

In [12] we argued that monitored traffic of constant data transmission show characteristic distributions. It was shown that it is possible to monitor data transmissions and obtain statistical quantities, such as mean, variance and standard deviation, which characterize distributions of received data at each different location. In particular, the received number of data packets (we called it *TxCount*) has been used as monitoring parameter. Further, we proposed in [12] the use of the Jeffrey Divergence (JD) with Gaussian approximation to calculate dissimilarities between monitored data and reference data. Figure 1 shows the idea of identifying a location by comparing an input distribution with a set of reference distributions. The reference distribution with the least dissimilarity is selected as the best fitting distribution, which identifies the location. The

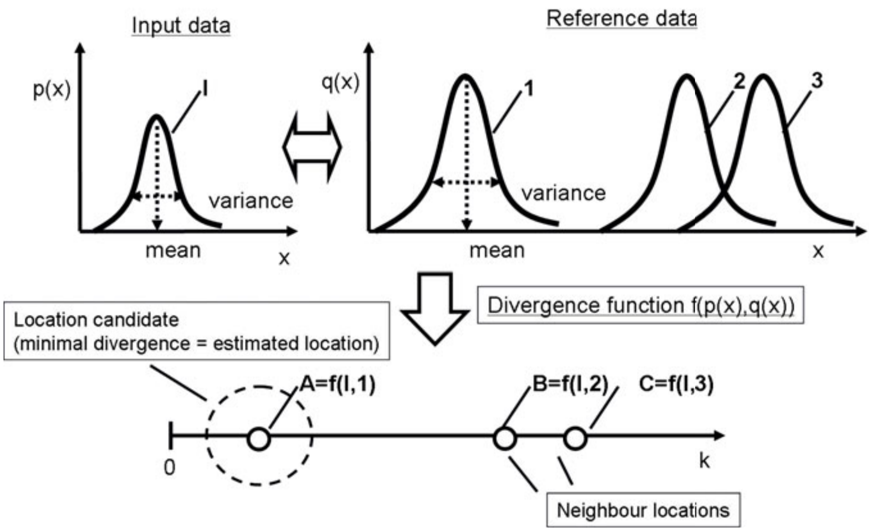


Fig. 1. Details of the proposed entropy-based location cognition method

reference distributions (RD) at different locations were obtained during the off-line phase, including data pre-processing, outlier-filter and removing transition states. The input data is frequently monitored during the on-line phase, for example once-per-second, and statistics of the input data are calculated on-line with a sliding window.

The location recognition procedure is conducted in two steps. First, the LOC executes the recognition of the generic location. In the second step, the LOC executes the recognition of the location and the distance between transmitter and receiver. One *generic RD set* is used for the estimation of the location and one *complete RD set* including distributions at different distances is used for the estimation of the distance. The generic RD at each location is obtained by selecting the distribution with the maximum static confidence value.

4 Estimation of Location Confidence

In addition to identifying a location by choosing the best fit which we introduced in [12], we introduce new parameters which describe the confidence of the training data sets and the LOC location classification. We contribute two new parameters called the *static confidence* and the *dynamic confidence* which have been developed to classify the confidence in the static set of training data or reference data and the confidence during the on-line classification of the monitored transmission to recognize the location of the wireless terminal. The proposed location method does not require an exhaustive learning at all points in the space. This is not desirable for an ad hoc user. However, the confidence can be used to measure the reliability of location classification over the entire space including intermediate points between the learnt points. The confidence shows the user when the location classification is reliable.

Figure 2 shows details of the implemented confidence estimator. The dynamic confidence is calculated using input data and the reference data in the on-line mode. The result is a time-variant value with a range of [0-100]. A confidence close to 100 indicates a large distance (large dissimilarities) between best fit and other templates, indicating high location confidence. A confidence close to 0 indicates a small distance (small dissimilarities), indicating low location confidence. In Fig. 2 an example is shown with a candidate (A) and two possible neighbor divergences (B), (C). Neighbor (B) has the smallest distance a to (A) and its distance contributes to the degree of confidence. If the distance a is close to the divergence of (A) the confidence will be low. If a is larger than the divergence of (A) the confidence increases. Neighbor (C) has the distance b to (A) which is larger compared to (B). The confidence increases when the distance of b increases. The confidence decreases when the number of neighbor candidates increases. Distances between all distributions are considered during the calculation of the dynamic confidence. The following algorithm has been implemented to calculate the dynamic confidence dc at sample k for n reference locations

$$a_k = \sum_{i=1}^{n} \frac{1}{|1 - JD_{i,k}/JD_{min,k}|} \tag{1}$$

with

$$dc_k = 100 \cdot \left(1 - e^{-\left(\frac{\alpha}{a_k}\right)}\right) . \tag{2}$$

The calculation of the distance between minimum divergence and neighbor divergence is executed in Eq. 1 and applied in Eq. 2. In Eq. 1 calculating distance between best fit $JD_{min,k}$ and neighbor candidate $JD_{i,k}$ is followed by normalizing each distance with JD_{min} to obtain the relative distance. The result in Eq. 1 increases when distances decrease (higher weight for small distances) and

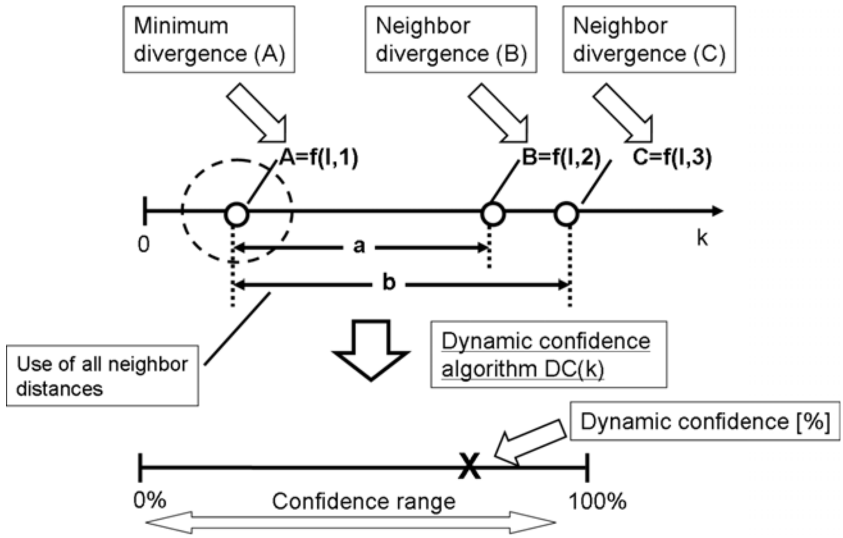


Fig. 2. Distance-based calculation of dynamic confidence

includes all distances. The sum is applied in Eq. 2 so that *dc* will decrease when number of distances increase. The algorithm uses a negative exponential function to converge the result toward 100 when distances increase (100% confidence) and toward 0 if the distances decrease (0% confidence would equal total similarity). The α-value was applied to the *dc* calculation for optimal scaling. Optimal scaling is achieved when *dc* increases towards 100 for large distances and reduces significantly if distances have been found small. Graphical analysis was performed prior to the implementation of the dynamic confidence algorithm identifying the optimal α-value (graphs excluded for reasons of brevity). The value of $\alpha = 0.5$ was found to allow an optimal recognition of dissimilarities.

In addition to the dynamic confidence which varies due to the time variant input data, the LOC has an intrinsic classification ability that depends on the selected distribution. In Eq. 3 the static confidence *sc* of the LOC is shown which can be estimated for different locations l selecting a candidate distribution D from the off-line learning phase. *sc* describes the ratio of true classifications T and the sum of true classifications and false classifications F for a given set of distributions D. The result is a value with a range of [0-1]. An increased number of true classifications and a reduced number of false classifications increase the *sc* value.

$$sc = \frac{\sum_l T_l|_D}{\sum_l T_l|_D + \sum_l F_l|_D}. \tag{3}$$

5 Implementation and Testing of a Prototype

In this section we describe the implementation and testing of a prototype of the location cognition (LOC). Acquisition of data during the off-line phase, selection of the reference distributions and then on-line location cognition are described. Transmission data was obtained by transmitting data between a pair of WLAN terminals at different locations under various conditions. The LOC engine was implemented in our wireless middleware in Linux, using kernel version 2.6, and includes a monitoring module, a location module that includes the LOC and a decision module for adaptive channel selection. We extended our WLAN IEEE 802.11 driver to obtain the TxCount value from the wireless device. The LOC uses the TxCount value which counts the number of received data packets. We have also implemented an outlier filter for data pre-processing. Transmission between transmitter and receiver was line-of-sight (LOS) in each case. The following five locations were used for testing of the LOC:

– Indoor locations

1. lab: an environment in a laboratory similar to an office environment (size: 10m x 20m x 3m).
2. corridor: a corridor inside the office building, consisting of doors, walls and windows (size: 3m x 30m x 3m).
3. entrance: an entrance hall in the office building with height of 10 m, mainly consisting of glass doors and large windows (size: 25m x 25m x 10m).

– Outdoor locations

1. building: a location beside the office building, 5m from the building.
2. road: a location at a road 50m from the office building.

At each location, transmission data were obtained for multiple distances in steps of 5m, namely 5m, 10m and 15m. Data were obtained for both short packets (200 bytes) and long packets (1500 bytes) capturing the effect of different packet sizes. A single UDP stream was sent continuously for the duration of 300s, increasing the transmission rate until maximum throughput is obtained. Transmission was monitored at the receiver in intervals of 1 sec at maximum throughput to obtain a reference distribution.

5.1 Selection of Reference Distributions

The mean and deviation of the reference distributions are shown in Fig. 3 and Fig. 4, respectively. The top graphs show the results for short packets obtaining max throughput at 7 Mbps before saturation occurred. The bottom graphs show the results for long packets at 28 Mbps before saturation occurred. Note that for all RDs the mean shows a higher number of successfully received data packets for outdoor locations, indicating a better quality link with less interference and multi-path effects. Regarding the indoor locations, the RDs for the entrance show the highest number of received data packets, followed by the RDs for lab and corridor which show the lowest number of received data packets. An explanation is that for the entrance less background wireless activities lead to a higher number of successfully received data packets. However, due to the exposed multi-path environment at the entrance (floor, doors, and windows) the number

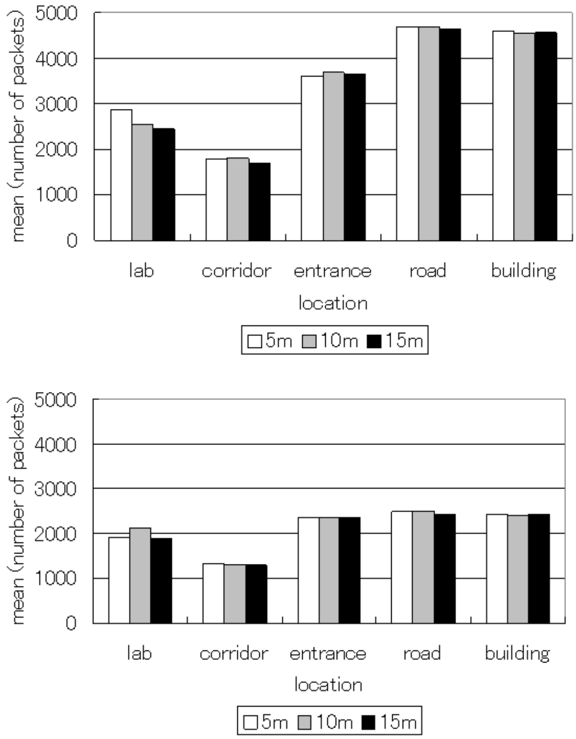
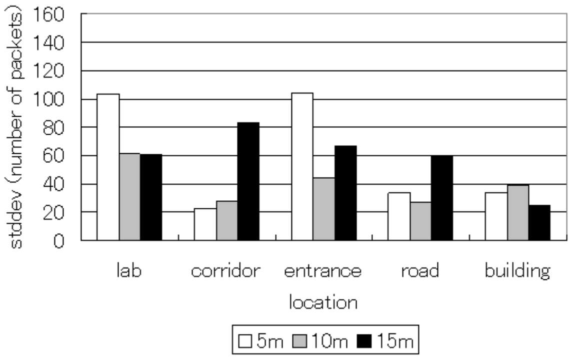


Fig. 3. Mean (number of packets) for short packets (top graph) and long packets (bottom graph)

of received packets is less than outdoors. Multi-path fading and increased wireless activities result in lower throughput in particular for the lab and the corridor as shown in Fig. 3. The characteristics for receiving data of long data packets are similar for short packets, except that the maximum value for outdoor is reduced for long data packets.

In Fig. 4 the standard deviation of all locations and different packet sizes are shown. It shows the highest value for the lab location for short packet lengths and decreases at the locations corridor and entrance. For the outdoor locations the standard deviation decreases significantly, and the location building shows a higher standard deviation than road. A conclusion is that multi-path fading at the long side of the building lead to higher standard deviation. This conclusion is valid for short packets including all distances. For the corridor location a lower standard deviation can be observed for short distances. For long packet length (1500 byte) the results show a different characteristic. The standard deviation is significantly reduced for all locations, expect for the lab location. Our explanation for this is that it is due to the significantly larger background wireless activity inside the lab.



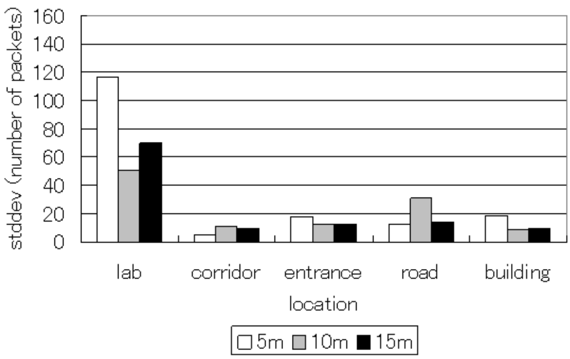


Fig. 4. Standard deviation (number of packets) for short packets (top graph) and long packets (bottom graph)

In Fig. 5 and Fig. 6 the relation between mean and standard deviation of the RDs are presented. Fig. 5 shows the results for the selected generic RD at each location, i.e., the RD at each location with a maximum static confidence value. Fig. 6 shows the results for the complete set of RD at each location and distance. Lab and corridor show the largest dissimilarities whereas entrance and building shows the smallest dissimilarities. Using Eq. 3 the result of static confidence for RDs in Fig. 5 was sc=1.0 for small packet length and sc=0.93 for long packet length. For RDs in Fig. 6 it was sc=0.76 for short packet length and sc=0.73 for long packet length.

5.2 Emulation of Location Cognition

During an emulation phase the LOC prototype was used to test the accuracy of the location cognition. In order to obtain quantitative results, emulations were run based on 3 trials at the 5 locations, lab, corridor, entrance, building and road at 3 different distances. The LOC used the pre-selected RDs of each location with a maximum of static confidence. Moreover, the LOC used an extended set of RDs which are used to identify the distance between the sender and

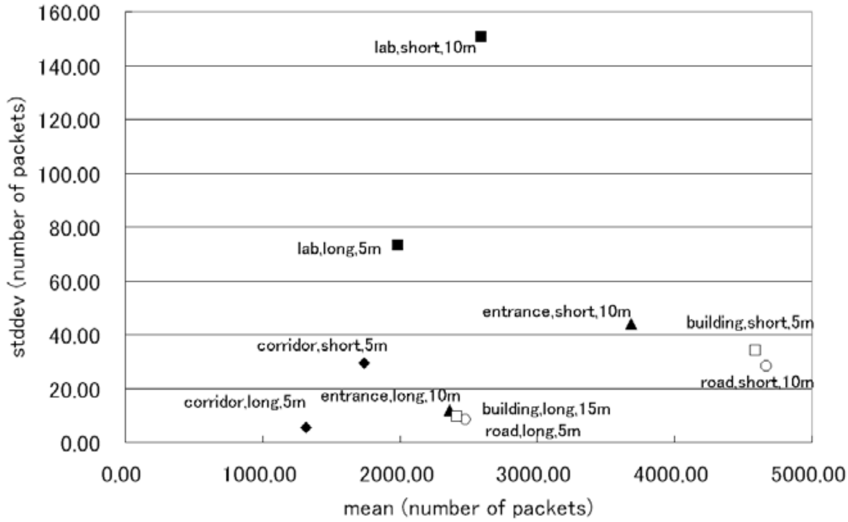


Fig. 5. Mean and standard deviation (stddev) of all RDs selected for location cognition including short and long packet sizes at 5 different locations (sc=1.0 for short packets, sc=0.93 for long packets)

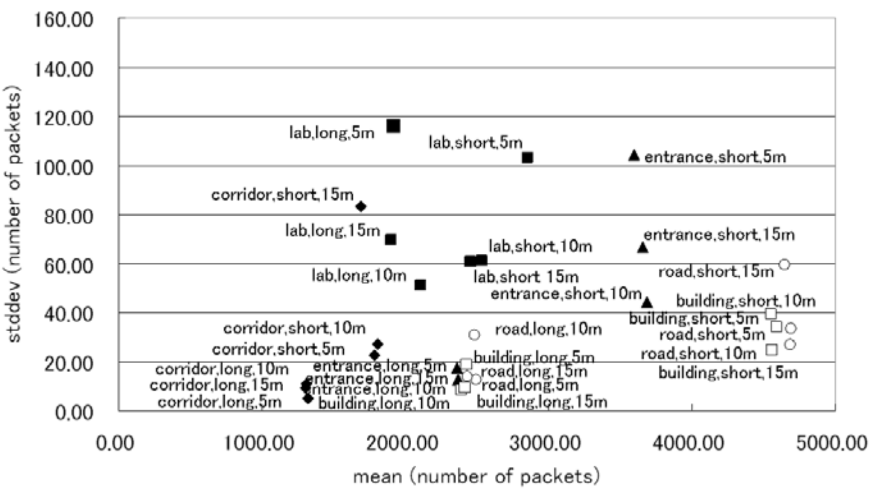


Fig. 6. Mean and standard deviation (stddev) of all RDs selected for location and distance cognition including short and long packet sizes at 5 different locations including the cognition of 3 different distances (sc=0.76 for short packets, sc=0.73 for long packets)

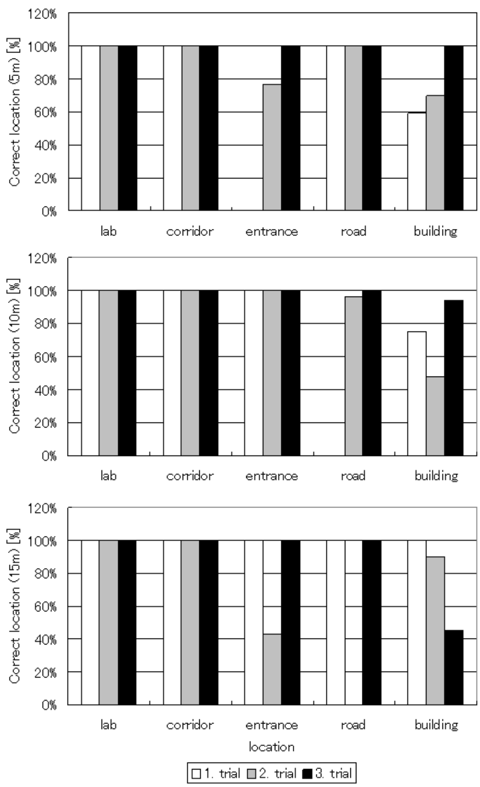


Fig. 7. Correct location cognition for all locations and distance (1500 byte, 3 trials)

receiver. The extended RD set contains additional distributions of each location for the distances 5m, 10m and 15m. We evaluated the accuracy of the location cognition by counting the number of correct classifications for multiple tests at the same location. Location cognition tests have been conducted including short packet length and long packet length for 3 trials (graphs excluded for reasons of brevity). The ratio of correct location cognition when transmitting short packets have shown 100% correct location for all 3 trials, expect for lab 15m (90%, 2. trial), road 15m (90%, 3. trial), and building (91%, 3. trial). We show the results for location cognition of long packets in Fig. 7. From the graph it can be observed that the location lab and corridor are judged correctly (100%), where as location entrance for 5m (0%, 77%) and 15m (43%) shows a reduced location cognition performance. We conclude that the entrance and building RD show similar statistics for long packets which lead into false detection.

In Fig. 7 the location road at 10m (1. trial) and 15m (2. trial) was judged as building, which can be counted as successful outdoor location cognition. Next, we discuss the accuracy of the combined classification of location and distance, in particular for the location building and entrance and for long packets (again,

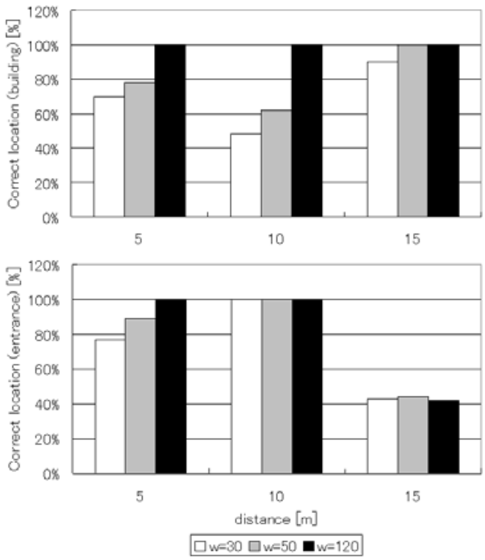


Fig. 8. Results of increased window size to improve combined location and distance for long packets (2. trial)

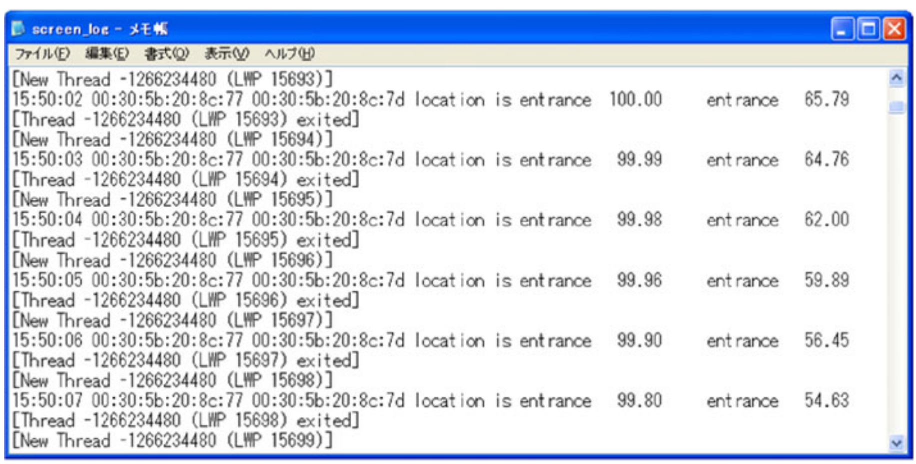


Fig. 9. On-line display of the LOC engine showing time-stamp, sender and receiver MAC address, location cognition, and dynamic confidence

graphs for short packets are excluded for reasons of brevity). These two locations are difficult to judge for the LOC.

In Fig. 8 it can be observed that the cognition is almost higher than 60% at 5m distance for both locations, having a window size of 30. The distance classification can be improved when the window size was increased to 50. Up to 80% successful location and distance classification can be observed. Similar improvements can be observed at 10 and 15, except the classification for the entrance at 15m, which remains unchanged (only observed for this particular trial). We conclude that the distance cognition has high success for a large variation of locations and distances.

Finally, Fig. 9 shows a screen shot of our implemented LOC engine including confidence values. Each line shows the results for a single cognition event. The first item is a time-stamp, the second and third items are MAC addresses of sender and receiver, and the following items show location classification results followed by confidence values. Two location cognition results can be observed, one for the generic location classification, obtained using the generic set of reference distributions, and the other for the extended location classification obtained using the extended set of reference distributions including distance.

6 Conclusions

Location cognition in wireless systems is an important problem and a challenging task. Various methods have been proposed to recognize the location of a terminal in specific environments. However, multi-location identification for indoor and outdoor is needed for future wireless networks. We proposed a simple-but-effective solution for recognition of multiple locations in indoor and outdoor environments, which does not need any infrastructure. The method is based on learning and recognizing the location dependent differences in wireless transmission characteristics based on a fingerprinting like method. We implemented a location cognition engine and demonstrated it is able to recognize different indoor/outdoor locations. We also implemented a novel distance cognition function to indicate the distance between sender and receiver. Finally, we proposed novel system parameters which report the confidence of the location cognition process.

The location cognition engine as it is described in this paper has been fully implemented in our Linux wireless middleware supporting the current IEEE 802.11a/b/g/n standards. The LOC uses a novel monitoring parameter, the number of received data packets, instead of RSS values. We have implemented an outlier filter for data pre-processing, and fast online location cognition. The proposed LOC is highly versatile and can be applied to detect indoor/outdoor environments for cognitive radios or location based services. All performance tests have been conducted in real WLAN environments including typical dynamics such as changes of wireless activities or surrounding environments. It can be concluded that multi-location cognition has been tested successfully in scenarios including both indoor and outdoor locations, and has the potential to be an integral part of intelligent WLAN systems in the future.

References

1. Arslan, H.: Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems. Springer, Heidelberg (2007)
2. Shellhammer, S.J., Sadek, A.K., Zhang, W.: Technical Challenges for Cognitive Radio in the TV White Space Spectrum. In: Information Theory and Applications Workshop, San Diego, pp. 323–333 (2009)
3. Kim, H.H., Ha, K.N., Lee, K.C.: Resident Location-Recognition Algorithm using a Bayesian Classifier in the PIR Sensor-Based Indoor Location-Aware System. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 39(2), 240–245 (2009)
4. Pan, J.J., Kwok, J.T., Yang, Q., Chen, Y.: Multidimensional Vector Regression for Accurate and Low-Cost Location Estimation in Pervasive Computing. IEEE Transactions on Knowledge and Data Engineering 18(9), 1181–1193 (2006)
5. Fischer, C., Gellersen, H.: Location and Navigation Support for Emergency Responders: A Survey. IEEE Pervasive Computing 9(1), 38–47 (2010)
6. Indoor Galileo/GPS Indoor Navigation & Positionierung (in German), funded by the Deutsches Zentrum fuer Luft & Raumfahrt (DLR) and the German Ministry, `http://www.indoor-navigation.de`
7. Gu, Y., Lo, A., Niemegeers, I.: A Survey of Indoor Positioning Systems for Wireless Personal Networks. IEEE Communications Surveys & Tutorials 11(1), 13–32 (2009)
8. Papliateseyeu, A., Kotilainen, N., Mayora, O., Osmani, V.: FINDR: Low-cost Indoor Positioning Using FM Radio. LNCS, Social Informatics and Telecommunications Engineering, vol. 7. Springer, Heidelberg (2009)
9. Kjaergaard, M.B., Treu, G., Ruppel, P., Kuepper, A.: Efficient Indoor Proximity and Separation Detection for Location Fingerprinting. In: 1st International Conference on Mobile Wireless Middleware, Operating Systems, and Applications, vol. 278, Innsbruck (2008)
10. Honkavirta, V., Perala, T., Ali-Loevtty, S., Piche, R.: A Comperative Survey of WLAN Location Fingerprinting Methods. In: 6th Workshop on Positioning, Navigation and Communication, Hannover (2009)
11. Mazuelas, S., Bahillo, A., Lorenzo, R.M., Fernandez, P., Lago, F.A., Garcia, E., Blas, J., Abril, E.J.: Robust Indoor Positioning Provided by Real-Time RSSI Values in Unmodified WLAN Networks. IEEE Journal of Selected Topics in Signal Processing 3(5), 821–831 (2009)
12. Aust, S., Matsumoto, A., Ito, T., Davis, P.: Supervised Classification Using Jeffrey Divergence for Location Cognition in Cognitive Radio. In: 12th International Symposium on Wireless Personal Multimedia Communications, Sendai (2009)