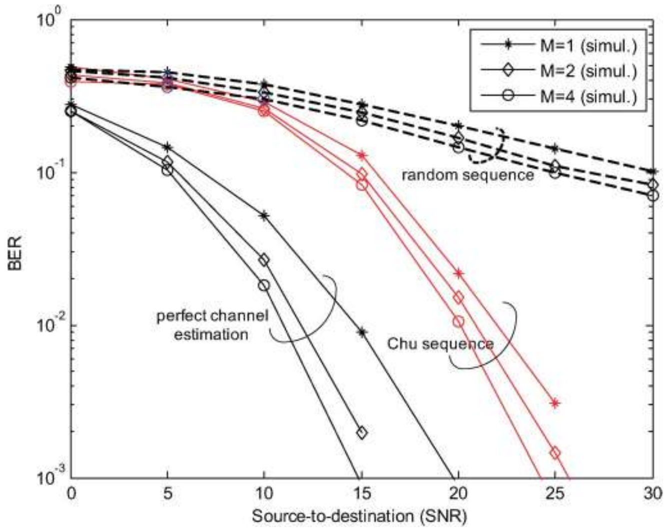

Fig. 6. BER performances according to $M = 1/2/4$ as a function of $\gamma_{SR} = \gamma_{SD}$ when $\omega = 1$.

VI. CONCLUSION

In this paper, we have presented the channel estimations for AF multiple-relay networks, particularly over frequency-selective fading channels. First, we have proposed the repeated known sequence consisting of Chu sequences, of which length depends on the number of relays and the maximum channel length of $\max(L_{h_0}, L_c)$. It was shown that the structure provides the complete precoding between relays without the IBI and the efficient frequency-domain channel estimation. Second, we have a simple signed precoding at relays by exploiting the extended orthogonal codes. It was shown that the precoding does not limit the number of relays and completely decouples the CIRs between links directly at D. Third, we have investigated various properties of the proposed channel estimations through the analysis of effective MSE in terms of link parameters, i.e., ω, γ_{SR}, and γ_{SD}. Moreover, it was verified that the combination of the repeated known sequence and precoding provides the minimum channel estimation performance (in the LS sense). Finally, the performance evaluations have been accomplished to verify the analysis results.

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[2] J. Choi, "MMSE-based distributed beamforming in cooperative relay networks," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1346–1356, May 2011.

[3] S. Zhang, X. Xia, and J. Wang, "Cooperative performance and diversity gain of wireless relay networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 9, pp. 1623–1632, Oct. 2012.

[4] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.

[5] J. Coon, M. Beach, and J. McGeehan, "Optimal training sequences for channel estimation in cyclic-prefix-based single-carrier systems with transmit diversity," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 729–733, Sep. 2004.

[6] M. G. Song, D. S. Kim, and G. H. Im, "Recursive channel estimation method for OFDM-based cooperative systems," *IEEE Commun. Lett.*, vol. 14, no. 11, pp. 1029–1031, Nov. 2010.

[7] A. S. Lalos, A. A. Rontogiannis, and K. Berberidis, "Frequency domain channel estimation for cooperative communication networks," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3400–3405, Jun. 2010.

[8] J. Coon, M. Beach, and J. McGeehan, "Channel and noise variance estimation and tracking algorithms for unique-word based single-carrier systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1488–1496, Jun. 2006.

[9] H. Mheidat, M. Uysal, and N. Al-Dhahir, "Equalization techniques for distributed space-time block codes with amplify-and-forward relaying," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1839–1852, May 2007.

[10] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 531–532, Jul. 1972.

[11] J. S. Baek and J. S. Seo, "Performance analysis of power control effects in channel estimation for AF relaying network," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 886–888, Feb. 2012.

RSSI-Fingerprinting-Based Mobile Phone Localization With Route Constraints

Sinem Coleri Ergen, Huseyin Serhat Tetikol, Mehmet Kontik, Raffi Sevlian, Ram Rajagopal, and Pravin Varaiya

***Abstract*—Accurate positioning of a moving vehicle along a route enables various applications, such as travel-time estimation, in transportation. Global Positioning System (GPS)-based localization algorithms suffer from low availability and high energy consumption. A received signal strength indicator (RSSI) measured in the course of the normal operation of Global System for Mobile Communications (GSM)-based mobile phones, on the other hand, consumes minimal energy in addition to the standard cell-phone operation with high availability but very low accuracy. In this paper, we incorporate the fact that the motion of vehicles satisfies route constraints to improve the accuracy of the RSSI-based localization by using a hidden Markov model (HMM), where the states are segments on the road, and the observation at each state is the RSSI vector containing the detected power levels of the pilot signals sent by the associated and neighboring cellular base stations. In contrast to prior HMM-based models, we train the HMM based on the statistics of the average driver's behavior on the road and the probabilistic distribution of the RSSI vectors observed in each road segment. We demonstrate that this training considerably improves the accuracy of the localization and provides localization performance robust over different road segment lengths by using extensive cellular data collected in Istanbul, Turkey; Berkeley, CA, USA; and New Delhi, India.**

***Index Terms*—Fingerprinting, localization, mobile phone, route constraints.**

I. INTRODUCTION

The accurate positioning of a moving vehicle along a route enables various applications, such as travel-time and pattern estimation, in transportation. Travel-time estimates are used to alleviate congestion by informing the drivers of the roads with large travel time, by finding better paths with smaller expected travel time in traffic-aware routing

Manuscript received September 27, 2012; revised May 20, 2013 and July 2, 2013; accepted July 17, 2013. Date of publication July 25, 2013; date of current version January 13, 2014. The review of this paper was coordinated by Prof. J. Chun.

S. Coleri Ergen, H. S. Tetikol, and M. Kontik are with the Department of Electrical and Electronics Engineering, Koç University, Istanbul 34450, Turkey (e-mail: sergen@ku.edu.tr; htetikol@ku.edu.tr; mkontik@ku.edu.tr).

R. Sevlian and R. Rajagopal are with the Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: sevlian@stanford.edu; ram.rajagopal@stanford.edu).

P. Varaiya is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA (e-mail: varaiya@eecs.berkeley.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2013.2274646

0018-9545 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

algorithms, and by improving traffic operations such as traffic-light cycle control. Travel pattern estimation in the form of origin–destination matrices, on the other hand, is an essential component in enabling administrative authorities to plan and manage infrastructure improvements. Using vehicles as probes to collect these data, rather than deploying fixed sensors, such as wireless magnetic sensors on the road [1], or using household questionnaires, allows efficient coverage of large areas of the road network.

Current proposals for accurate positioning heavily rely on using GPS-enabled devices that are periodically recording the current time and position, and sending these data to a central server over a wireless network for further processing [2]. The GPS, however, is not available in a large percentage of today's mobile phones, particularly in developing regions, leaving their users out of the many location-based applications. Moreover, GPS has poor performance in urban areas near high-rise buildings and consumes much energy by quickly draining the battery of mobile devices, which may not necessarily be connected to a charger within the vehicle. Other alternatives, such as Wi-Fi [3] and augmented-sensor-based localization [4] in the cell phones, suffer from the low availability, which is similar to GPS. GSM on the other hand is available on all GSM-based cell phones and consumes minimal energy in addition to the standard cell-phone operation. Among the GSM-based localization algorithms, including time of arrival, time difference of arrival, angle of arrival, and received signal strength indicator (RSSI)-based methods, RSSI-based localization algorithms are mostly preferred since an RSSI is measured in the course of the normal operation of GSM-based mobile phones; therefore, they do not require any specialized hardware [5].

According to GSM standards, each cell-phone records a vector of the detected power levels of the pilot signals from at most seven cellular base stations, one of which the cell phone is associated with. Positioning algorithms using this RSSI vector follow two approaches: triangulation and fingerprinting. In triangulation, first, the RSSI reading from each base station is matched to a distance based on an experimentally determined path-loss model, and then, the position of the mobile phone is estimated based on the location of each base station and the distances to them. The accuracy of triangulation-based approaches, however, has been shown to be very low due to the inaccuracy of the path-loss models [6], [7]. In the fingerprinting approach, RSSI measurements from the cellular base stations are recorded along with the GPS measurements during the database construction phase. The position is then determined by comparing the measured RSSI vector to those in the database identifying the GPS location of the closest RSSI matches, which are determined by minimizing a cost function [8]. The accuracy of fingerprinting has been also shown to be very poor, requiring stochastic postprocessing such as Kalman filtering [9]. Recently proposed trajectory-based methods using the RSSI readings of the neighboring base stations apply hidden Markov models (HMMs) to improve the localization accuracy by taking into account route constraints; however, they employ a completely random behavior of the vehicles in each direction, such as an equal probability of moving to each of the adjacent cells in a grid-based map without any training [10]–[12]. None of the previously proposed HMM techniques developed for GSM RSSI-based localization derives the statistics of the driver's behavior on the road and analyzes the effect of including these statistics on the localization accuracy. Furthermore, these techniques mostly use deterministic approaches to estimate the location of the mobile phones without analyzing the probabilistic distribution of the observed RSSI vectors on the road segments as a function of the amount of the training data nor comparing this probabilistic-model-based approach to the commonly used cost functions in the literature [10], [11]. The only probabilistic fingerprinting approaches in the literature, on the other hand, either consider the associated base station only [12] or do not take the route constraints into account [13].

There has been an extensive amount of work on RSSI-based localization systems employing HMMs using indoor wireless local area network (WLAN) base stations [14], [15]. The movement models used for the indoor environments considering the building's floor plan and expected pedestrian speeds, however, are not suitable for modeling the movement of the vehicles. HMM- and RSSI-based localization algorithms using outdoor WLAN base stations [3] and sensor networks [16], again, cannot be used in GSM-based vehicle localization since they assume a completely random behavior of the nodes in the network, and the propagation characteristics of the short-range WLAN and sensor node communications are very different from those of the long-range GSM communications.

In this paper, we propose an HMM-based localization algorithm exploiting GSM RSSI data of the mobile phones located within the vehicles, which is trained by the statistics of the average driver's behavior on the road and the probabilistic distribution of the RSSI vectors in each road segment. The original contributions of this paper are as follows.

1) We derive the statistics of the average driver's behavior and the probabilistic distribution of the RSSI vectors in each road segment of urban and suburban roads in Istanbul, Turkey; Berkeley, CA, USA; and New Delhi, India, for the first time in the literature.
2) We demonstrate that the proposed HMM trained by these statistics significantly improves the accuracy of the previous HMM techniques and provides a localization performance robust over different road segment lengths based on extensive cellular data collected in Istanbul, Turkey; Berkeley, CA, USA; and New Delhi, India.

II. System Overview

A. Localization Using HMM

We assume that the vehicle follows the same road. We divide the road into segments representing the hidden states of the HMM. Transition among these hidden states is governed by transition probabilities. Given the segment of the road, the RSSI vector has a particular conditional probability distribution called the emission probability distribution of the observable events in the HMM. The HMM then traverses its states to produce its outputs, which are the observable events emitted at each state.

The algorithm has offline training and online tracking phases. In the offline training phase, both GPS measurements and RSSI vectors are periodically recorded while the vehicle is moving on the road. These data are then transmitted via any available wireless network to the central server. The collected data from multiple users are used to estimate the transition probability and emission probability of the HMM. During the online tracking phase, a sequence of periodically recorded RSSI vectors are sent to the central server and input to the HMM to estimate the most probable sequence of states of road segments.

B. Mathematical Model of HMM

The HMM can be mathematically represented as $\lambda = (S, V, A, B, \pi)$, with the following variables.

1) $S = \{s_1, s_2, \ldots, s_N\}$ is the set of states, with each state representing a road segment on the road and $N = |S|$ being the number of road segments on the road.
2) $V = \{v_1, v_2, \ldots, v_M\}$ is the set of observation vectors, where $M = |V|$ is the total number of possible observation vectors on the road. v_k is a $Z \times 1$ vector for $k \in [1, M]$ such that the ith element of v_k, which is denoted $v_{k,i}$, is the received power from the ith base station represented by an integer value in the range

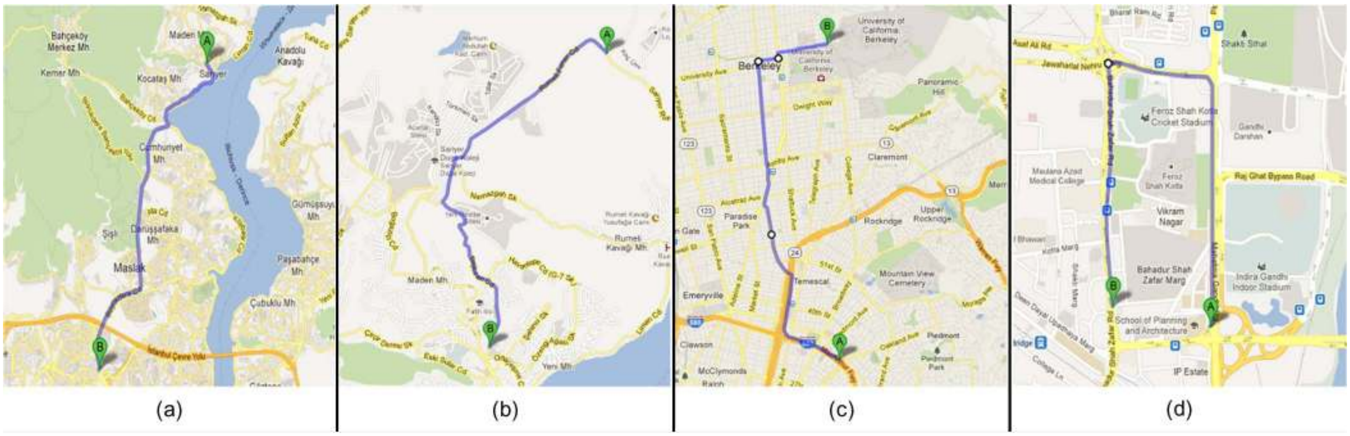

Fig. 1. (a) Urban road with a length of 12 km between Sariyer and 4. Levent, Istanbul, Turkey. (b) Suburban road with a length of 4 km between Koç University and Sariyer, Istanbul, Turkey. (c) Suburban road with a length of 8.2 km between the University of California, Berkeley, CA, USA and Temescal, Oakland, CA, USA, (d) Suburban road with a length of 2.9 km in New Delhi, India.

of $[-110, -40]$ dBm if the ith base station is one of the seven strongest base stations and 0, otherwise, where Z is the total number of base stations observed on the road.

3) A is an $N \times N$ matrix such that the element in the ith row and the jth column of A, which is denoted a_{ij}, is the probability of the transition from state s_i to state s_j, i.e., $a_{ij} = P(q_{n+1} = s_j | q_n = s_i)$, where q_n is the state at time $n\tau$, and τ is the sampling period. First, we assume that the vehicle follows the same road and that the training of the HMM is performed by using the data collected from the vehicles following that same road. This assumption eliminates the requirement for considering the effect of the variation of the routes of different drivers to their destination. Second, we assume that the vehicles move with the same average speed in the same direction along the road used in the training and tracking phases, meaning that the transition probability a_{ij} is independent of the segment i but only depends on the difference between i and j, i.e., $a_{ij} = a_k$, where $k = j - i$ and $a_k = 0$ for $k < 0$. Matrix A needs to be derived for each road satisfying these conditions. This approach can be generalized to general vehicle tracking, where different drivers have different routes to their destination by dividing the road into multiple segments where these assumptions hold. The first novelty of this paper is the derivation of a_k values based on the training data summarizing the statistics of the average driver's behavior on the road. The previous HMM RSSI-based GSM localization algorithms allow the transitions only to the current or following road segments, i.e., $a_0 = 1/2$ and $a_1 = 1/2$ [10]–[12].

4) B is an $N \times M$ matrix such that the element in the jth row and the kth column of B, which is denoted b_{jk}, is the emission probability of v_k on the jth road segment corresponding to state s_j, i.e., $b_{jk} = P(v_k | s_j)$. The emission probability distributions derived for WLAN systems [3], [14], [15] and sensor networks [16] cannot be used for GSM systems due to their different propagation characteristics. The HMM techniques developed for GSM RSSI-based localization, on the other hand, use either a deterministic function of the Euclidean distance and the number of matching base stations between the observed vector and each RSSI vector in the training database as the cost of the observed RSSI vector [11] or the histogram of the RSSI readings of the associated base station in each road segment [12]. The second novelty of this paper is proving that the histogram of the RSSI readings of each base station within each road segment converges to the Gaussian distribution if enough RSSI samples are collected.

5) π is an $N \times 1$ vector such that the ith element of π, which is denoted π_i, is the probability of state s_i initially, i.e., $\pi_i = P(q_1 = s_i)$.

III. Data Collection

We collected data over four different roads shown in Fig. 1. The first road between Sariyer and 4. Levent, Istanbul, Turkey, with a length of 12 km, shown in Fig. 1(a), represents an urban road in Istanbul. The second road between Koç University and Sariyer, Istanbul, Turkey, with a length of 4 km, shown in Fig. 1(b), represents a suburban road in Istanbul. The third road between the University of California, Berkeley, CA, USA, and Temescal, Oakland, CA, USA, with a length of 8.2 km, shown in Fig. 1(c), represents a suburban road in the U.S. The fourth road in New Delhi, India, with a length of 2.9 km, shown in Fig. 1(d), represents a suburban road in India.

The monitoring device for collecting RSSI signatures was a standard Nokia handset operating in the 900-MHz band and connected to a laptop that runs network monitoring software. A sample of the RSSI signature containing the power levels and identifiers of the seven strongest base stations is reported once every 2 s along with the GPS latitude and longitude information, i.e., $\tau = 2$ s. Twenty test runs are performed at low-, medium-, and high-traffic conditions on the urban road in Istanbul, Turkey, shown in Fig. 1(a), whereas 20 test runs are performed on the suburban roads shown in Fig. 1(b)–(d) since the traffic conditions do not significantly change during the day, i.e., ten runs are used for training, and ten runs are used for tracking.

IV. Estimation and Usage of Hidden Markov Model Statistics

A. Training Phase

During the training phase, GPS and RSSI vector data observed from the cell phones are collected at the central server to create the database. The parameters of the HMM are then estimated using this database.

Figs. 2 and 3 show the probability of the transition to the following road segments, i.e., a_k, for the urban road and suburban roads of different segment lengths, respectively. L denotes the length of the segment. The calculation of a_k given the training database including the sequence of states recorded by GPS, i.e., $Q^r = (q_1^r, q_2^r, \ldots, q_{T_r}^r)$ for $r \in [1, R]$, where q_i^r is the ith state in the rth run, R is the number

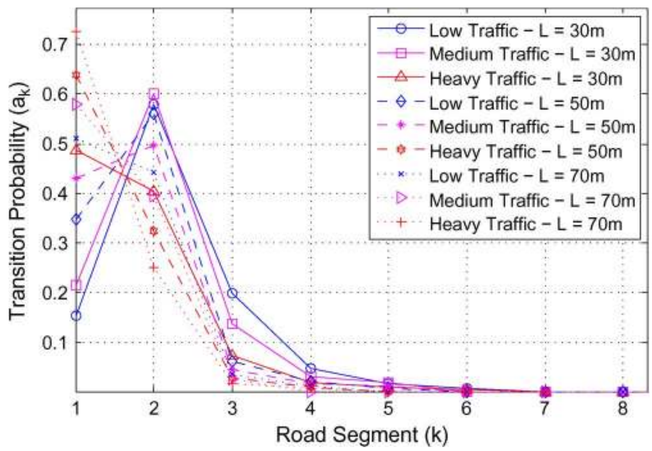

Fig. 2. Probability of the transition to the following road segments at low-, medium-, and high-traffic conditions for the urban road of different segment lengths.

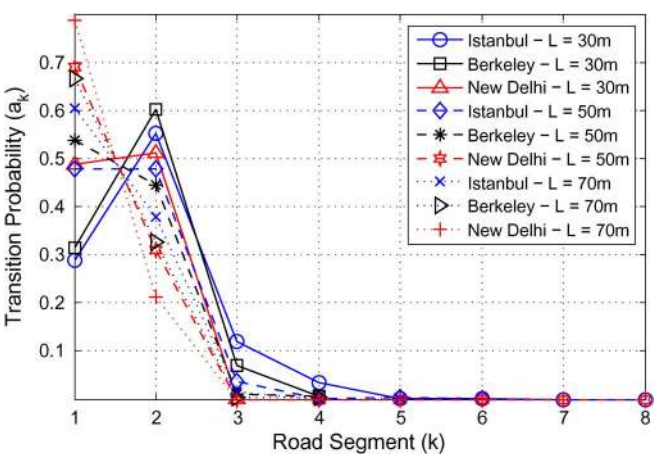

Fig. 3. Probability of the transition to the following road segments for the suburban roads of different segment lengths.

of runs in the training data, and T_r is the number of data points corresponding to the rth run in the training data, is given as

$$a_k = \frac{\sum_{r=1}^{R} \sum_{j=1}^{T_r-1} \delta_{q_{j+1}^r - q_j^r, k}}{\sum_{r=1}^{R} \sum_{j=1}^{T_r-1} 1} \tag{1}$$

where $\delta_{q_{j+1}^r - q_j^r, k}$ is the Kronecker delta function that takes the value of 1 if $q_{j+1}^r - q_j^r = k$ and 0, otherwise; the numerator is the number of transitions of magnitude k; and the denominator is the total number of transitions. We observe that the transition probability significantly varies depending on the location and traffic conditions of the roads, contradicting the common assumption of the HMM-based localization literature that the mobiles move only to the adjacent grid cells or adjacent road segments.

The analysis of many histograms of the observed RSSI values from many base stations shows that the histogram has a Gaussian shape if enough RSSI samples are collected from the base station in the corresponding road segment. The number of RSSI samples required for the Kolmogorov–Smirnov test, with a 95% confidence interval to decide for the Gaussian distribution of these RSSI samples, is less than 30 in all the segments of different lengths over all the roads used for the evaluation. (The Kolmogorov–Smirnov test is a nonparametric test for the equality of continuous 1-D probability distributions comparing a sample distribution with a reference probability distribution).The number of samples required to estimate the mean and variance of the Gaussian distribution of the RSSI samples with 90% accuracy is around 30. Given that the number of samples collected from the segment lengths that are greater than or equal to 10 m, which are used in the evaluation of the algorithm, in each run of the training data is more than 4, the total number of samples from ten runs of the training data, i.e., 40 samples, will be enough to estimate the mean and variance of the Gaussian distribution with 90% accuracy.

Let us denote the mean and variance of the best Gaussian fit to the RSSI histogram of the ith base station at the jth road segment corresponding to state s_j by $m_{j,i}$ and $\sigma_{j,i}^2$, respectively. Recall that b_{jk} is the emission probability of an observation vector v_k in state s_j. Let us define the cost of observing v_k in state s_j, which is denoted c_{jk}, as the normalized logarithm of b_{jk}. Based on the assumption that the samples from different base stations are independent, c_{jk} is given by

$$c_{jk} = \frac{1}{Z_p} \sum_{i=1, m_{j,i}<0, v_{k,i}>0}^{Z} \left(-\frac{1}{2}\ln(2\pi) - \ln(\sigma_{j,i}) - \frac{(v_{k,i} - m_{j,i})^2}{2\sigma_{j,i}^2} \right) \tag{2}$$

where Z_p is the number of base stations for which both $m_{j,i}$ and $v_{k,i}$ are negative. This is very similar to the previously proposed cost functions defined as the distance between RSSI vectors, as in [3] and [11], except the weighting factor as a function of the variance, giving less weight to the observation with high variance.

The sampling period τ, the segment length L, the road, and traffic conditions on the road affect the values of both the transition and emission probabilities, but the algorithm works as long as the same τ and L values, the same road, and the same traffic conditions are used in both the training and tracking phases.

B. *Tracking Phase*

During the tracking phase, the mobile phone only sends the RSSI vector information to the central server so that the server identifies the most likely sequence of the road segments traveled by the user, i.e., states $Q = (q_1, q_2, \ldots, q_T)$, given the sequence of the observed RSSI vectors, i.e., observations $O = (O_1, O_2, \ldots, O_T)$, where T is the number of observations. Since the parameters of the HMM are estimated during the training phase, the Viterbi algorithm is used to estimate the most likely sequence of the road segments [17].

V. Performance Evaluation

A. *Localization Error Comparison at Different Traffic Conditions in Different Locations*

Figs. 4 and 5 compare the accuracy of the proposed HMM-based technique using the statistics of the driver and RSSI measurements at each road segment, which is denoted S-HMM, with the previously proposed HMM technique called Ctrack [11] for urban and suburban roads, respectively. Ctrack is actually developed as a two-pass HMM, which first determines a sequence of grid cells corresponding to an input sequence of GSM fingerprints and then matches the grid cells to a road. The cost of the observed RSSI vector is a weighted function of the Euclidean distance and the number of matching base stations between the observed vector and each RSSI vector in the training database. Ctrack assumes uniform transition probability from the current grid cell to each of the adjacent cells and is adapted here for the localization on the road. We observe that the median localization error for Ctrack is very large for small and large segment lengths, whereas S-HMM has a stable localization error due to the inclusion of the driver and RSSI statistics. Moreover, the median error of S-HMM is smaller than that of Ctrack for all segment lengths.

Fig. 6 shows the cumulative distribution function (cdf) of the localization error on the suburban road in Berkeley, CA, USA, for S-HMM,

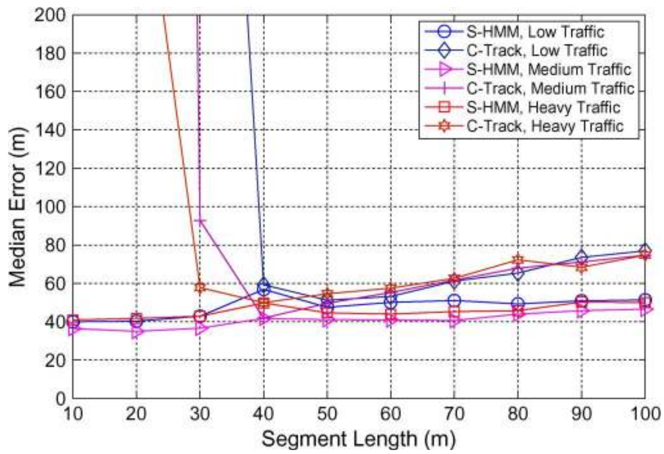

Fig. 4. Comparison of the median error of the proposed S-HMM algorithm with the previously proposed HMM technique called Ctrack at low-, medium-, and high-traffic conditions for the urban road.

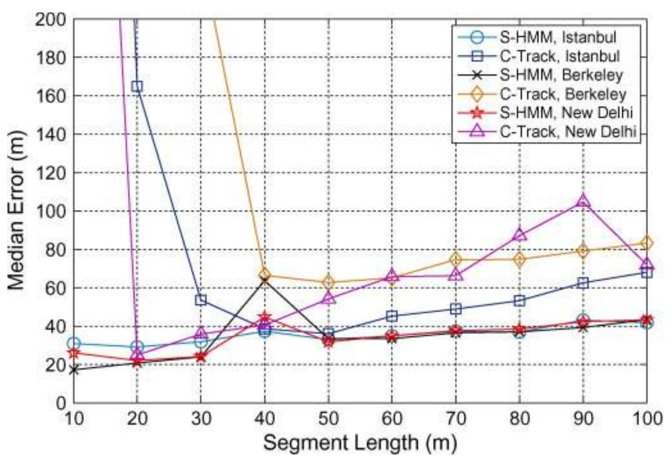

Fig. 5. Comparison of the median error of the proposed S-HMM algorithm with the previously proposed HMM technique called Ctrack for the suburban roads in different locations: Istanbul, Turkey, Berkeley, CA, USA, and New Delhi, India.

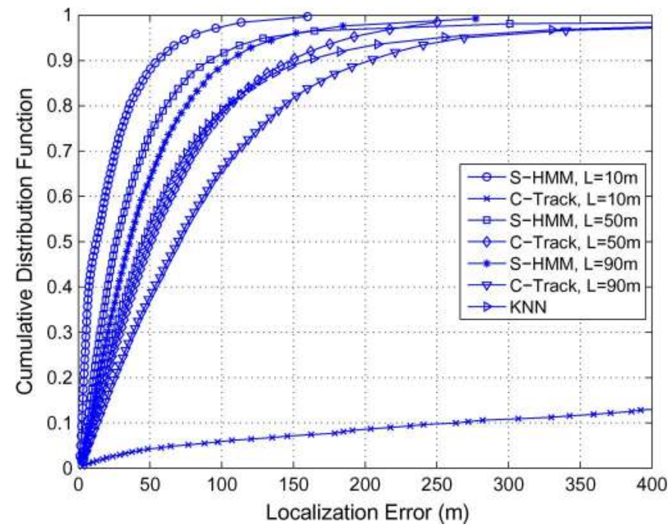

Fig. 6. CDF of the localization error of S-HMM, Ctrack, and KNN algorithms for the suburban road in Berkeley, CA, USA.

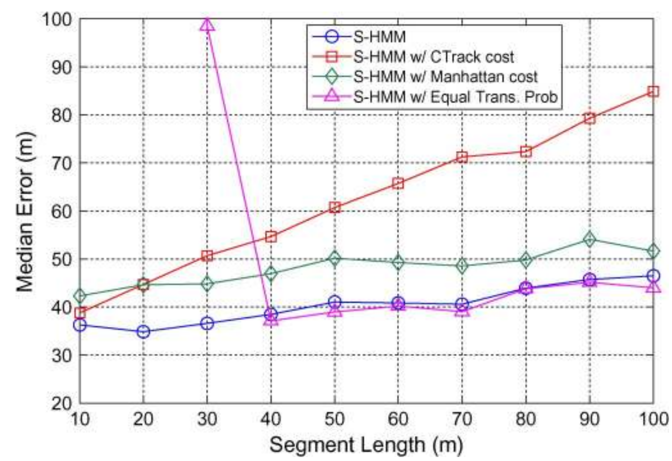

Fig. 7. Analysis of the effect of including the driver and RSSI statistics on the localization error for the urban road at medium traffic.

Ctrack, and a deterministic algorithm called K-nearest neighbor (KNN) [18]. In the KNN algorithm, all the collected RSSI vectors and the corresponding GPS locations are stored during the training phase. During the tracking phase, the received RSSI vector at an unknown location is compared with all the RSSI vectors stored in the database, and its location is estimated as the average of the GPS locations of the K-closest RSSI vectors in terms of Euclidean distance. We observe that the localization error for S-HMM is always to the left of the Ctrack and KNN algorithms for all the segment lengths. The cdf of the localization error for the urban road and suburban roads in Istanbul, Turkey, and in New Delhi, India exhibits a similar behavior.

B. Effect of Including Driver and RSSI Statistics

Fig. 7 shows the effect of including the driver and RSSI statistics on the localization error by comparing the performance of S-HMM to variations of S-HMM, including different transition probabilities and different cost functions for the urban road at medium traffic in Istanbul, Turkey. Since the behavior is similar for the other traffic conditions and locations, we did not include separate graphs for them. S-HMM uses driver and RSSI statistics to determine the transition probability matrix and emission probability of the HMM, respectively. "S-HMM w/ Ctrack cost" and "S-HMM w/ Manhattan cost" use driver statistics to determine the transition probability matrix but use the Ctrack cost function and the Manhattan cost for emission score, respectively. The Manhattan cost of the observed RSSI vector is defined as the Manhattan distance between the observed vector and each RSSI vector in the training database [8]. "S-HMM w/ equal trans.prob.," on the other hand, uses RSSI statistics to determine the emission probability but uses uniform transition probability to the adjacent segments. We observe that the Ctrack cost performs worst as the segment length increases, whereas the Manhattan cost is more stable but still performs worse than S-HMM. The cost function is the main reason for the increase in the localization error of Ctrack as the segment length increases. When we compare S-HMM with "S-HMM w/ equal trans.prob.," we see that the assumption of uniform distribution for the transition probability is the main reason for the steep increase in the localization error of Ctrack as the segment length decreases. However, interestingly, the localization error of "S-HMM w/ equal trans.prob." is smaller than that of S-HMM when the segment length is above 40 m. The main reason is that our implicit assumption of memoryless transitions between road segments in determining the statistics of the driver's behavior loses validity as the segment length increases. As the segment length increases, we may need to include the state duration density in the HMM [19].

C. Energy Consumption Comparison

The cell phone is expected to operate in the tracking phase after a short training period. The training can be also done by using some voluntary mobile phones only. Since cell phones continuously track base stations as part of their normal operation, the marginal energy cost of the tracking phase is driven by the CPU load to process base station signatures and to send this information to the central server. Processing a cell tower signature might require at most 100 000 instructions, which costs at most 30 μJ on an ARM Cortex-A8 processor. The energy required to send the packets depends on how frequently the data are uploaded to the central server. A negligible amount of additional power is consumed if the collected data are compressed using simple gzip compression and are sent every minute [11]. Acquiring cell-phone signatures with a period of 2 s therefore results in power consumption of 15 μW. The power consumption of GPS on the other hand is measured to be around 400 mW [11], which is several orders larger than that of the GSM-based localization.

VI. Conclusion

In this paper, we have proposed a statistical fingerprinting approach to position a mobile device that follows any predetermined route by using RSSI measurements recorded from at most seven cellular base stations in the course of the normal operation of GSM-based mobile phones with minimal energy consumption. We incorporate the route constraint for the motion of the vehicles by using an HMM. We demonstrate that training the HMM based on the average driver statistics on the road and the probabilistic distribution of the RSSI vectors in each road segment considerably improves the accuracy of the previously proposed HMM-based techniques and provides localization performance robust over different segment lengths by using extensive cellular data collected in Istanbul, Turkey; Berkeley, CA, USA; and New Delhi, India.

References

[1] S. Cheung, S. Coleri, B. Dundar, S. Ganesh, C. Tan, and P. Varaiya, “Traffic measurement and vehicle classification with a single magnetic sensor,” *J. Transp. Res. Rec.*, no. 1917, pp. 173–181, Feb. 2006.

[2] J. Herrera, D. Work, X. Ban, R. Herring, Q. Jacobson, and A. Bayen, “Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment,” *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 6, pp. 568–583, Aug. 2010.

[3] A. Thiagarajan, L. Sivalingam, K. LaCurts, S. Toledo, J. Eriksson, S. Madden, and H. Balakrishnan, “Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones,” in *Proc. ACM Sensys*, Nov. 2009, pp. 85–98.

[4] I. Constandache, R. R. Choudhury, and I. Rhee, “Towards mobile phone localization without war-driving,” in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.

[5] M. Porretta, P. Nepa, G. Manara, and F. Giannetti, “Location, location, location,” *IEEE Veh. Technol. Mag.*, vol. 3, no. 2, pp. 20–29, Jun. 2008.

[6] A. Weiss, “On the accuracy of a cellular location system based on RSS measurements,” *IEEE Trans. Veh. Technol.*, vol. 52, no. 6, pp. 1508–1518, Nov. 2003.

[7] N. Ghaboosi and A. Jamalipour, “The geometry of overhearing and its application for location estimation in cellular networks,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3324–3331, Sep. 2011.

[8] A. Arya, P. Godlewski, and P. Melle, “Performance analysis of outdoor localization systems based on rss fingerprinting,” in *Proc. 6th ISWCS*, Sep. 2009, pp. 378–382.

[9] C. Takenga, T. Peng, and K. Kyamakya, “Post-processing of fingerprint localization using Kalman filter and map-matching techniques,” in *Proc. 9th Int. Conf. Adv. Commun. Technol.*, Feb. 2007, pp. 2029–2034.

[10] T. Stamoulakatos, A. Markopoulos, M. Anagnostou, and M. Theologou, “Vehicle velocity estimation based on rss measurements,” *Wireless Pers. Commun.*, vol. 40, no. 4, pp. 523–538, Mar. 2007.

[11] A. Thiagarajan, L. S. Ravindranath, H. Balakrishnan, S. Madden, and L. Girod, “Accurate, low-energy trajectory mapping for mobile devices,” in *Proc. 8th USENIX Symp. NSDI*, Mar. 2011, p. 20.

[12] M. Ibrahim and M. Youssef, “A hidden Markov model for localization using low-end GSM cell phones,” in *Proc. IEEE ICC*, Jun. 2011, pp. 1–5.

[13] M. Ibrahim and M. Youssef, “Cellsense: An accurate energy-efficient GSM positioning system,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 286–296, Jan. 2012.

[14] J. Krumm and E. Horvitz, “LOCADIO: Inferring motion and location from Wi-Fi signal strengths,” in *Proc. 1st Annu. Int. Conf. MOBIQUITOUS Syst., Netw. Services*, Aug. 2004, pp. 4–13.

[15] L. Kung-Chung and L. Lampe, “Indoor cell-level localization based on RSSI classification,” in *Proc. 24th CCECE*, May 2011, pp. 21–26.

[16] R. Arthi and K. Murugan, “Localization in wireless sensor networks by hidden Markov model,” in *Proc. 2nd ICoAC*, Dec. 2010, pp. 14–18.

[17] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.

[18] M. Y. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. LaMarca1, F. Potter, I. Smith, and A. Varshavsky, “Practical metropolitan-scale positioning for GSM phones,” in *Proc. Int. Conf. UbiComp*, Sep. 2006, pp. 225–242.

[19] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

Energy-Efficient Water-Filling with Order Statistics

Ismael Gomez-Miguelez, Vuk Marojevic, and Antoni Gelonch

***Abstract*—This paper proposes a methodology based on order statistics to study the energy-efficient (EE) power allocation for wireless communication transmitters. Water-filling power allocation maximizes the EE when transmitting and processing power consumption is considered. Solutions are typically presented in an iterative form, which complicates analysis and adaptive implementation. Time-consuming simulations are rather required to assess their performance. Order statistics allows analyzing the solution without knowledge of the channel realization. We analytically show how the EE depends on the system parameters. The computing efficiency of our proposal, which is executing one or two orders of magnitude faster than the state-of-the-art algorithms with a performance loss of less than 1 dB, facilitates its applicability in vehicular environments.**

***Index Terms*—Energy efficiency (EE), link adaptation, ordered statistics, water-filling.**

I. Introduction

The energy consumption of wireless communications networks contributes to the escalation of the greenhouse effect, which has been recognized as a major threat to environmental protection and sustainable development. Sophisticated data services and the always-on concept of battery-operated systems impose significant energy management challenges. Energy-efficient (EE) transmission architectures and technologies are therefore needed for mobile terminals and base stations alike.

Manuscript received February 9, 2013; revised April 24, 2013; accepted May 28, 2013. Date of publication June 28, 2013; date of current version January 13, 2014. This work was supported by the Spanish Ministry of Economy and Competitiveness under Project TEC2011-29126-C03-02. The review of this paper was coordinated by Dr. C. Yuen.

The authors are with the Department of Signal Theory and Communications, Technical University of Catalonia, Barcelona 08034, Spain (e-mail: ismael.gomez@tsc.upc.edu; marojevic@tsc.upc.edu; antoni@tsc.upc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2013.2271647

0018-9545 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.