Robust Wi-Fi fingerprinting-based positioning in the presence of lying identities

Wei-Chung Lu, Shih-Chun Yeh, Chung-Chih Chuang and Shih-Hau Fang

Yuan Ze University, Taiwan

*E-mail: a53051232000@yahoo.com.tw

Abstract. The lying identity of an access point (AP) is one of the most serious threat in Wi-Fi positioning because an adversary can easily acquire a valid address by monitoring the transmission and masquerade as another AP in the networks. This study proposes a robust Wi-Fi localization algorithm that can tolerate the liars instead of explicitly detecting them. The proposed algorithm considers all possible combinations of APs in an unionbased approach such that the adversaries cannot easily affect the positioning results by masquerading APs. On-site experimental results demonstrate that this approach apparently achieves more robust location estimation than the Bayesian approach and the cluster-based method in the presence of lying identities.

1. Introduction

Secure localization has gained considerable attention in the last several years because position estimations are often required for critical applications such as location-based authentication [1], [2], [3]. Today, many location systems utilize the base station's broadcasting power in wireless or cellular networks to infer the user's location due to the wide coverage and the widespread use of receivers [4], [5]. Among various approaches, fingerprinting is a popular location architecture, in which the user's location is estimated by comparing received signal strength (RSS) with the values pre-stored in a database [6], [7]. This architecture allows the location system to reuse existing wireless infrastructure [8], [9]. Recently, the massive deployment of Wi-Fi access points (APs), and the popular RSS sensing function on mobile devices make Wi-Fi a suitable technology for developing such location systems [10], [11]. However, due to the license-free spectrum (2.4GHz) and the shared nature of the wireless medium, the propagated radio waves in Wi-Fi networks are susceptible to attacks. Among various attacks, the lying identity of AP is definitely one of the most serious threat [12], [13].

In a typical 802.11 Wi-Fi environment, the MAC address is a unique identifier for an AP in Wi-Fi networks. This address is important for a location system because it indicates which AP transmits the signal. However, an adversary can easily acquire a valid MAC address and masquerade as another AP in the networks. Figure 1 depicts a typical scenario where the Wi-Fi network contains a lying identity. In this case, users in the Wi-Fi network cannot distinguish signals from AP1 and the lying AP. Such lying identities have a serious impact on a location system because the RSS value from AP1 mixes the lying data. That is, the adversary can easily fool the Wi-Fi location system by monitoring the transmission, masquerading APs, and changing the lying AP's broadcasting power.

The main contribution of this work is to use an enhanced probabilistic localization algorithm in [17] for building a robust Wi-Fi location system in the presence of lying identities. The proposed method of dealing with lying identities is to design dependable localization scheme that tolerates the liars instead of explicitly detecting them. Compared to the traditional probabilistic approaches which use the intersection of RSS from all APs, the proposed algorithm considers all possible AP combinations in an union-based approach such 2 that the adversaries cannot easily affect the positioning results by pretending APs. The main advantage of this approach is that a lying identity contributes little to the localization process, and thus, achieving more robust location estimations under security threats. `

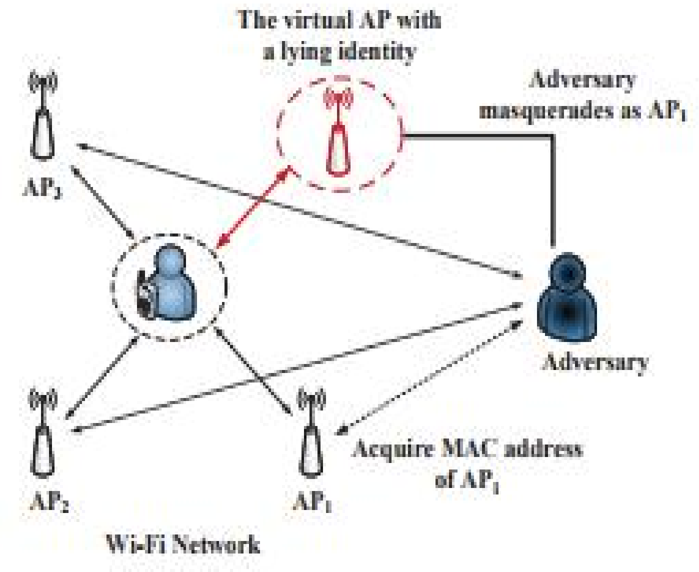

The virtual AP with a lying identity
Adversary masquerades as AP$_1$
AP$_3$
Adversary
AP$_2$
AP$_1$
Acquire MAC address of AP$_1$
Wi-Fi Network

Figure 1. A Wi-Fi network contains a lying AP, where the RSS values from AP1 mix the lying data.

2. Background and related work

In this paper, we focus on the fingerprinting-based indoor location system that can effectively defend against lying identities in Wi-Fi networks. In a fingerprinting-based location system, the online stage is more vulnerable to attacks because the validation and detection schemes can be applied during the off-line stage. Moreover, the training data can be collected over various days to avoid corruption. Thus, this study assumes an attack-free condition in the off-line stage. Researchers have proposed many mechanisms to defend various attacks, including sybil [19], spoofing [15], [20], and transmit power attacks [13], [21].

Common approaches to securing wireless localization fall into two categories: attack detection and elimination, and robust localization schemes. The first method tries to detect the attacks and eliminate them from consideration during the positioning procedure. For instance, Liu et al. filtered out malicious beacon signals using the consistency among multiple beacon signals [22]. The second method of dealing with attacks is to design a robust localization scheme. For example, Li et al. developed two robust statistical methods, including a least median square position estimator for triangulation-based systems, and a median-based distance metric for fingerprinting location systems [23]. Liu et al. proposed a voting scheme to achieve attack-resistant location estimation [22].

3. Proposed algorithm

Assume a device senses N APs, then the observed RSS can be represented by a vector X with N elements as X=(x_1 , x_2 , · · · , x_N), where x_N means the observed RSS from the n-th AP. A fingerprinting location system regards X as an input and outputs the estimated location by a previously constructed RSS-location relationship. However, the presence of cheating APs may cause some xn totally unreliable and thus, result in a fake position of the user. The problem in robust localization is how to correctly locate the user given X which contains some lying x_N. In traditional probabilistic location systems, the joint probability of all sensed observations is estimated for all candidate locations to find the user's position. The main idea of such approaches is that each x_N is combined based on the "and" operator, denoted by ∧. This is because traditional methods assume all APs behave honestly. Assuming each x_N is independent, the joint probability of measuring X at a certain location L equals the product of the individual probability of x_N as:

$$\begin{aligned}\mathrm{P}(X|L) &= P(x_1 \wedge x_2 \wedge \cdots \wedge x_N|L)\\ &= P(x_1|L)P(x_2|L)\cdots P(x_N|L)\end{aligned} \tag{1}$$

Because of the multiplication, x_N's small probabilities dominate the overall probability in Eq.1. This property is desirable in a normal condition because the multiplication allows the location system to effectively discriminate the neighboring locations. However, such characteristic makes the positioning algorithm sensitive to lying identities. An adversary can easily affect the location estimation by controlling the lying AP to produce a fake signal strength x_N The fake xn produces a small P(xn|L) due to a mismatch between the lying data and the lying-free database. Unfortunately, this small value destroys the location system's ability of generating a large probability $\mathrm{P}(x_N|L)$ at a correct location. This explains why the method in Eq.1 is susceptible to lying APs. Here we assume that the database is lying-free because many detection schemes can be applied to the data during the offline stage whereas the online lying detection is difficult.

This study proposes a general probabilistic approach to consider possible lying APs. Unlike the model in Eq.1, the individual probability of x_N is combined via an "or" model. This model assumes that the RSSs from reliable APs may be any x_N or any combination among x_N belongs to the complete set. This can be formulated by the "or" operator ∨ as: $X_\vee = x_1 \vee x_2 \cdots \vee x_N$, where X∨ is a combined observation based on the "or", representing the possible true APs in X. In mathematics, the "or" is a logical operator that results in true whenever one or more of its arguments are true. Considering a three-AP case (N=3), X∨ = x_1 ∨ x_2 ∨ x_3. These combinations characterize that the observed RSS vector encounters, respectively, two-lying, one-lying and lying-free conditions. The probability of this or model can be obtained by the probabilistic rules of the union of two random events: P(A∨B)=P(A) +P(B)−P(A∧B). More specifically, the proposed algorithm computes P(X∨) based on a recursion as:

$$\begin{aligned}\mathrm{P}\big(\vee_{n=1}^{m} x_n\big|L\big) &= P\big(\vee_{n=1}^{m-1} x_n\big|L\big) + P(x_m|L)\\ &\quad - P\big(\langle\vee_{n=1}^{m-1} x_n\big|L\rangle \wedge \langle x_m|L\rangle\big)\\ &\quad \mathrm{m} = 2,\cdots,\mathrm{N}\end{aligned} \tag{2}$$

Assuming that the observations are independent with each other, Eq.2 is simplified as:

$$\begin{aligned}\mathrm{P}\big(\vee_{n=1}^{m} x_n\big|L\big) &= P\big(\vee_{n=1}^{m-1} x_n\big|L\big) + P(x_m|L)\\ &\quad - \mathrm{P}\big(\vee_{n=1}^{m-1} x_n\big|L\big)\mathrm{P}(x_m|L)\\ &\quad \mathrm{m} = 2,\cdots,\mathrm{N}\end{aligned} \tag{3}$$

Finally, the estimated location coordinate can be calculated as the average of all reference locations by adopting their new probability as weights:

$$\hat{L} = \textstyle\sum_{r=1}^{R} P(L_r|X_\vee) \cdot L_r \tag{4}$$

where L^ is the estimated location, R is the number of reference locations, L_r is the coordinate of the r-th reference location, and $\mathrm{P}(L_r|X_\vee)$ represents the union-based posterior probability of measuring X at the r-th reference location. The component of $\mathrm{P}(L_r|X_\vee)$, $\mathrm{P}(L_r|x_n)$ can be computed as

$$\mathrm{P}(L_r|x_n) = \frac{1}{\sqrt{2\pi}\sigma_{n,r}} exp\left(-\frac{(x_n - u_{n,r})}{2\sigma_{n,r}^2}\right) \tag{5}$$

where $u_{n,r}$ and $\sigma_{2n,r}$ represent the mean and variance of the n-th AP and at the r-th reference location. These parameters can be obtained by maximum likelihood estimation as

$$\widehat{\theta}_{n,r} = arg \max_{\widehat{\theta}_{n,r}} \prod_{k=1}^{K} P\big(x_n[k]\big|\theta_{n,r}\big) \quad (6)$$

where K is the number of training samples and $\theta_{n,\,r} = \{u_{n,\,r},\, \sigma_{n,\,r}\}$ are the unknown parameters of the n-th AP and at the r-th reference location. A necessary condition that $\theta_{n,\,r}$ must satisfy is the gradient of Eq.6 with respect to $\theta_{n,\,r}$ to be zero. These parameters are estimated by using K training samples during the offline stage.

4. *Experimental setup and results*

4.1. Experimental Setup

We placed three WiFi APs in the target area and collected realistic RSS data at 27 different reference locations to build a fingerprinting location system. We used an Asus laptop with Windows XP as the mobile node, and NetStumbler network software to gather RSS. The size of the test-bed was 70 m2 and RSS was measured 200 times for each reference location. We then selected 15 locations, separated by a distance of 1.5 to 2 meters, as the training data. To simulate the lying identities, we choose specified MAC addresses as lying APs. Furthermore, RSS from the lying AP is changed to an alternative AP's data. That is, the RSS samples from normal APs used in positioning mixed with different ratios of the lying RSS. The experiments compared with the Bayesian and the securing cluster-based scheme [15], [16].

4.2. Performance Evaluation

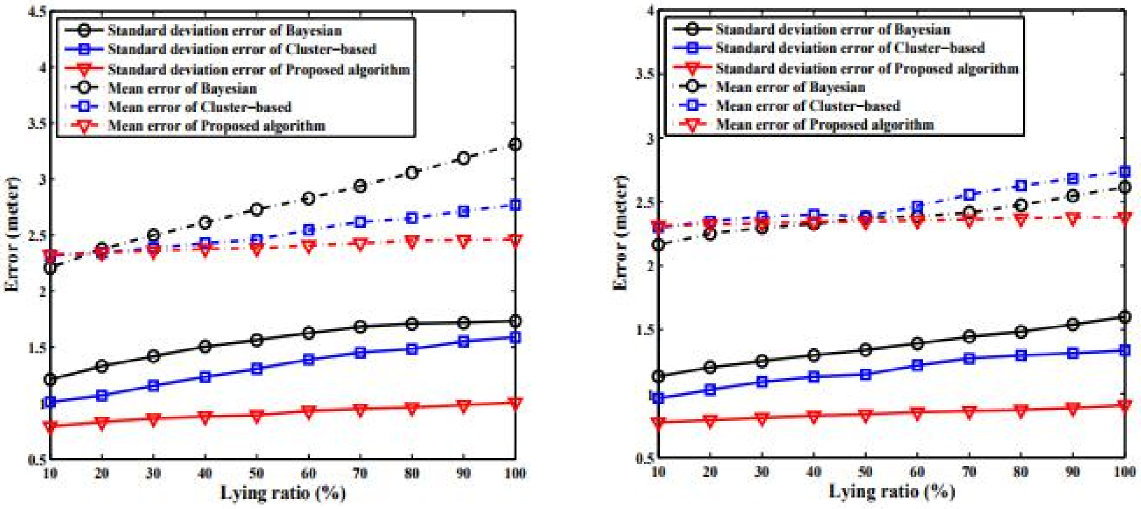
The first experiment evaluates the positioning performance of our algorithm when two lying identities exist. Table I shows that our approach apparently improves the robustness in the presence of lying identities. All statistical errors show a significant reduction based on our approach. This means that a malicious adversary cannot easily fool our location system by masquerading APs. Numerical results show that the proposed algorithm outperforms the Bayesian and the cluster-based method in reducing the 50-th and the 67-th percentile localization errors by 5.42-26.76%, and 18.89- 34.24%, respectively. Table I also indicates that the Bayesian approach is the most sensitive to lying identities, as its 90-th percentile localization errors increased to almost six meters. This is because the probabilistic model that the Bayesian approach adopts is susceptible to lying identities.

The next experiment investigates the impact of different mixture ratios of lying data. Figure 2(a) shows the standard deviation and mean positioning error versus the mixture ratios for different algorithms. The lying ratio represents the power of the lying identities, where 0% represents a normal condition and 100% indicates the total replacement of actual data. Figure 2(a) shows that the proposed model significantly improved robustness compared with other methods. As the lying ratios increased, both the mean and SD error of the proposed algorithm remained almost fixed, whereas that of traditional approaches clearly increased. This again indicates that our location system is more secure and robust to tolerate the lying identities.

The next experiment considers an alternative scenario in which only one lying AP exists in the environment. Figure 2(b) compares the positioning performance in this case. This figures is generally consistent with figure 2. The main difference is that the proposed algorithm shows a mean-error improvement when the lying ratio exceeds 50%. In addition, the performance is better while reducing one identity, as compared figure 2(a) with figure 2(b).

Table 1. Five statistical errors comparison when two lying identities exist with a mixture rate 100%.

	Median error	Mean error	SD error	90-th percentile error
Bayesian	3.2149	3.3103	1.7328	5.8423
Cluster-based	2.4897	2.7693	1.5865	5.4099
Proposed algorithm	2.3547	2.4585	1.0036	3.5978



(a) **(b)**

Figure 2. Standard deviation and mean positioning error versus the mixture ratio for different algorithms under two lying identities**(a)** and one lying identity**(b)**.

5. Conclusion

Secure wireless positioning is important because the location estimations are often input to some critical location-based services. However, due to the nature of the wireless medium, the broadcasting radio is vulnerable to a variety of malicious attacks such as the lying identity in Wi-Fi networks. This study proposes a robust fingerprinting localization algorithm that consider the existence of lying APs. By considering all possible AP combinations in an union-based approach, our location system allows a lying AP to contribute little to the localization process, and thus, achieving more robust location estimations under security threats. On-site experimental results demonstrate that our approach apparently improves the robustness in the presence of lying identities.

References

[1] C. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati,“An obfuscation-based approach for protecting location privacy,” IEEE Transactions on Dependable and Secure Computing, vol. **8**, no. 1, pp. 13–27, 2011.

[2] A. Boukerche, H. Oliveira, E. Nakamura, and A. Loureiro, “Secure localization algorithms for wireless sensor networks,” IEEE Communications Magazine, vol. **46**, no. 4, pp. 96 –101, 2008.

[3] Y. Zhang, W. Liu, Y. Fang, and D. Wu, “Secure localization and authentication in ultra-wideband sensor networks,” IEEE Journal on Selected Areas in Communications, vol. **24**, no. 4, pp. 829 – 835, 2006.

[4] M. Jadliwala, S. Zhong, S. Upadhyaya, C. Qiao, and J.-P. Hubaux,“Secure distance-based localization in the presence of cheating beacon nodes,” IEEE Transactions on Mobile Computing, vol. **9**, no. 6, pp. 810–23, 2010.

[5] R. Ouyang, A.-S. Wong, and C.-T. Lea, “Received signal strength-based wireless localization via semidefinite programming: Noncooperative and cooperative schemes,” IEEE Transactions on Vehicular Technology, vol. **59**, no. 3, pp. 1307 –1318, 2010.

[6] M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen, “Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on WiMAX networks,” IEEE Transactions on Vehicular Technology, vol. **59**, no. 1, pp. 283–

294, 2010.

[7] Y. Jin, W.-S. Soh, and W.-C. Wong, “Indoor localization with channel impulse response based fingerprint and nonparametric regression,” IEEE Transactions on Wireless Communications, vol. **9**, no. 3, pp. 1120 –1127, 2010.

[8] S.-H. Fang and T.-N. Lin, “A dynamic system approach for radio location fingerprinting in wireless local area networks,” IEEE Transactions on Communications, vol. **58**, no. 4, pp. 1020–25, 2010.

[9] E. Chan, G. Baciu, and S. Mak, “Using the newton trust-region method to localize in WLAN environment,” Wireless and Mobile Computing, Networking and Communications, pp. 363 –69, 2009.

[10] A. Ghosh, V. Kaul, and D. Famolari, “An approach to secure localization in WLANs,” Wireless Communications and Networking Conference, pp. 3145 –50, 2008.

[11] S.-H. Fang and T.-N. Lin, “Projection-based location system via multiple discriminant analysis in wireless local area networks,” IEEE Transactions on Vehicular Technology, vol. **58**, no. 9, pp. 5009 –19, 2009.

[12] Y. Chen, J. Yang, W. Trappe, and R. Martin, “Detecting and localizing identity-based attacks in wireless and sensor networks,” IEEE Transactions on Vehicular Technology, vol. **59**, no. 5, pp. 1–1, 2010.

[13] K. Bauer, D. McCoy, E. Anderson, M. Breitenbach, G. Grudic, D. Grunwald, and D. Sicker, “The directional attack on wireless localization: how to spoof your location with a tin can,” IEEE conference on Global telecommunications, pp. 4125–30, 2009.

[14] J. Joseph, B.-S. Lee, A. Das, and B.-C. Seet, “Cross-layer detection of sinking behavior in wireless ad hoc networks using SVM and FDA,” IEEE Transactions on Dependable and Secure Computing, vol. **8**, no. 2, pp. 233 –45, 2011.

[15] J. Yang, Y. Chen, W. Trappe, and J. Cheng, “Determining the number of attackers and localizing multiple adversaries in wireless spoofing attacks,” IEEE INFOCOM 2009, pp. 666 –74, 2009.

[16] Y. Chen, W. Trappe, and R. Martin, “Detecting and localizing wireless spoofing attacks,” Sensor, Mesh and Ad Hoc Communications and Networks, pp. 193 –202, 2007.

[17] S.-H. Fang, C.-C. Chuang, and C. Wang, “Attack-resistant wireless localization using an inclusive disjunction model,” IEEE Transactions on Communications, vol. **60**, no. 5, 2012.

[18] X. Li, Y. Chen, J. Yang, and X. Zheng, “Designing localization algorithms robust to signal strength attacks,” in Proceedings IEEE INFOCOM, pp. 341 –45, 2011.

[19] J. Yang, Y. Chen, and W. Trappe, “Detecting sybil attacks in wireless and sensor networks using cluster analysis,” Mobile Ad Hoc and Sensor Systems, pp. 834 –39, 2008.

[20] H. Chen, W. Lou, J. Ma, and Z. Wang, “TSCD: A novel secure localization approach for wireless sensor networks,” Sensor Technologies and Applications, pp. 661 –666, 2008.

[21] A. Kushki, K. Plataniotis, and A. Venetsanopoulos, “Sensor selection for mitigation of RSS-based attacks in wireless local area network positioning,” IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2065–68, 2008.

[22] D. Liu, P. Ning, A. Liu, C. Wang, and W. K. Du, “Attack-resistant location estimation in wireless sensor networks,” ACM Transactions on Information System Security, vol. **11**, no. 4, pp. 1–39, 2008.

[23] Z. Li, W. Trappe, Y. Zhang, and B. Nath, “Robust statistical methods for securing wireless localization in sensor networks,” Information processing in sensor networks, pp. 91–98, 2005.

[24] Y. Chen, K. Kleisouris, X. Li, W. Trappe, and R. P. Martin, “A security and robustness performance analysis of localization algorithms to signal strength attacks,” ACM Transactions on Sensor Networks, vol. **5**, no. 1, pp. 1–37, 2009.