

The Digital Marauder’s Map: A New Threat to Location Privacy

Xinwen Fu
UMass Lowell

Nan Zhang & Aniket Pingley
George Washington University

Wei Yu
Cisco Systems, Inc.

Jie Wang
UMass Lowell

Wei Zhao
University of Macau

Abstract

“*The Marauder’s Map*” is a magical map in J. K. Rowling’s fantasy series, “*Harry Potter and the Prisoner of Azkaban*”. It shows all moving objects within the “*Hogwarts School of Witchcraft and Wizardry*”. This paper introduces a similar attack to location privacy in wireless networks. Our system, namely the digital Marauder’s map, can reveal the locations of WiFi-enabled mobile devices within the coverage area of a single high-gain antenna. The digital Marauder’s map is built solely with off-the-shelf wireless equipments, and features a mobile design that can be quickly deployed to a new location and instantly used without training. We present a comprehensive set of theoretical analysis and experimental results which demonstrate the coverage and localization accuracy of the digital Marauder’s map.

1. Introduction

In this paper, we study a class of novel localization attacks to compromise the location privacy of mobile devices in WiFi networks. Location privacy is the ability of a person to prevent others from learning his/her current or past locations. The widespread use of WiFi and cellular networks has led to increasing concerns about location privacy for mobile device users [1]. It is critical to study the threats from localization attacks and evaluate their impact on location privacy.

We present the *Digital Marauder’s Map*, a system that reveals the locations of WiFi-enabled devices in the coverage area of a specialized sniffing system. Our proposed techniques are principled on two novel ideas: First, we design a single-antenna-based system which monitors 802.11 *probing traffic* to determine the set of APs communicable with each mobile device in the covered area. We propose two types of attacks, passive and active, to monitor a wide variety of mobile devices. To maximize the coverage area and the number of covered channels, we propose to combine the usage of high-gain antennas with a low noise amplifier (LNA) and a signal splitter, in order to collect as much wireless traffic as possible.

Second, we propose three localization algorithms, M-Loc, AP-Rad, and AP-Loc, for an attacker to accurately position a mobile device based on the set of APs communicable with it. M-Loc and AP-Rad exploit the external (spatial) knowledge of APs available at many wireless geographic

logging websites such as WiGLE [2]. In particular, M-Loc locates mobile devices when the locations and maximum transmission distances of APs are available through external knowledge, while AP-Rad only requires the locations to be available. AP-Loc addresses the scenario where no AP information is available through external knowledge. In this case, AP-Loc first uses wardriving or warwalking techniques [3] to collect a minimal number of training data tuples, then locates APs based on the training data, and finally calls AP-Loc to compromise the location of mobile devices.

Our contributions can be summarized as follows: (i) To the best of our knowledge, our work is the first to study attacks against location privacy through a full-fledged malicious tracking system utilizing high-gain antennas and LNA in WiFi networks. (ii) We propose a novel mechanism for localization attacks in WiFi networks which uses passive or active techniques to collect probing traffic generated by a mobile device, in order to determine the set of APs communicable with each mobile device in the covered area. (iii) We present M-Loc, AP-Rad, and AP-Loc, three localization algorithms for a third-party attacker to accurately locate all monitored mobile devices, without utilizing signal strength information. (iv) Our contribution also includes a thorough experimental study which demonstrates the coverage and localization accuracy of the digital Marauder’s map as well as the superiority of our localization algorithms over the existing applicable efforts.

The digital Marauder’s map can be used for tracking mobiles with static MAC addresses, which are common in reality. Researchers have proposed pseudonym based schemes to hide MAC addresses. However, Pang *et al.* [4] found that many implicit identifiers such as network names in probing traffic may break those pseudonyms. Combined with their schemes, the digital Marauder’s map can also track a victim in case pseudo-mac addresses are used.

The rest of the paper is organized as follows. Section 2 introduces the digital Marauder’s map and discusses its components and related issues. The attack theory and algorithms will be given in Section 3. We evaluate the system in Section 4. We review the related work in Section 5 and conclude this paper and discuss future work in Section 6.

2. System

This section introduces the main ideas behind the digital marauder’s map, our system for malicious wireless tracking.

2.1. Basic Idea

The basic idea of localization attack is to sniff the interaction between mobile devices and access points (APs) and utilize the AP spatial information (e.g., location and/or maximum transmission distance) to pinpoint the mobile device. There are two phases in a full cycle of the attack: an optional training phase and a (main) attacking phase.

In the (optional) *training phase*, an adversary derives the AP locations through wardriving or warwalking [3]. This stage is optional because such training data is often available through wireless geographic logging websites. For example, the location of 15 million APs is available at WiGLE [2]. When such information is not available through external knowledge, an adversary initiates the training phase by equipping its mobile device with GPS and wireless sniffing tools such as Netstumbler or Kismet. Then, the adversary travels through the target area when the sniffing tools constantly probe APs and record training data.

In the *attacking phase*, an adversary can derive the location of a wireless device in two steps: First, it identifies a set of APs communicating with the device. Then, it derives the wireless device's location based on the AP locations available either through the external knowledge or from the training phase. To accomplish the first step, a high-gain antenna is set up to collect *probing* traffic transmitted between the victim and APs on all available wireless channels. We will introduce a passive mechanisms and an active one for probing traffic collection over mobile devices of different operating systems. In the second step, we will introduce two algorithms *M-Loc* and *AP-Rad* to pinpoint a mobile device based on its set of communicable APs and the AP locations and/or maximum transmission distances.

2.2. System Overview

Figure 1 depicts the architecture of the *Digital Marauder's Map*, our system for malicious wireless tracking. This system is used in the attack phase to locate a victim mobile based on AP spatial information from external knowledge or the training phase. It consists of four major components:

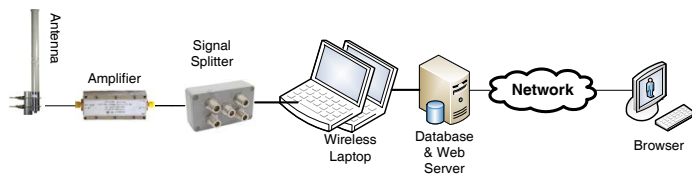


Figure 1: Framework of the Digital Marauder's Map

(i) *The wireless receiver chain*. It includes a number of high gain antennas, low noise amplifiers (LNAs), signal splitters, and wireless cards connected through connectors. The high gain antenna boosts the received signal strength. The signal passes through a powered low noise amplifier, which amplifies the signal power while improving the receiver chain's signal to noise ratio (SNR). The amplified

signal passes a signal splitter, which splits the input signal into multiple threads, which are in turn fed into a number of corresponding wireless cards.

(ii) *Wireless traffic capture*. Each thread of wireless signal is captured by a wireless card, which processes and extracts information such as SSIDs and AP MAC addresses from packets. In an ideal case, the adversary should be able to capture wireless traffic through all channels, including 11 channels for 802.11 b/g and 12 channels for 802.11a. The extracted information is then stored in a database.

(iii) *Malicious localization*. The trained or preknown AP information is stored in another database. For each AP, the information includes its SSID, MAC address, spatial location and (optionally) maximum transmission distance. Based on this information, the adversary uses our proposed *M-Loc* and *AP-Rad* algorithm to locate the victim mobile device according to its set of communicable APs.

(iv) *Digital Marauder's map display*. A simple web interface is then used to display the locations of all mobile devices in the monitored area. In particular, we use Google maps [5] to overlay the location on top of topology map.

2.3. Issues

From the discussion above, we can see that there are three major challenges in designing the digital Marauder's map, the malicious wireless tracking system:

(i) *Coverage area*: The coverage area refers to the area the adversary can monitor and pinpoint a mobile within this area. The wireless receiver chain is the component that determines the coverage area. How can the chain be optimized to increase the coverage area?

(ii) *Feasibility*: The success of our localization attack relies on the comprehensive monitoring of probing traffic transmitted between a mobile device and the APs. What if a mobile device is not sending out probe requests? How can the adversary effectively collect probing traffic?

(iii) *Localization accuracy*: The effectiveness of our localization attack is determined by the design of localization algorithms. What factors affect the algorithm design? How to improve the localization accuracy of the adversary?

We will investigate these issues in the following section.

3. Attack Analysis and Algorithms

In this section, we present the detailed design of the digital Marauder's map. We first address the coverage problem of a single high-gain antenna using radio theory. Then, we discuss the collection of probing traffic while reducing the cost and maintaining a reasonable degree of mobility for the monitoring system. Finally, we present and analyze three localization algorithms to locate mobile devices given different (or no) external knowledge about the AP locations.

3.1. Coverage Area

To cover a large area, we need to optimize the wireless receiver chain. Many factors affect the choice of each

component in the chain. An intuitive idea is to use high gain antennas and amplifiers: a high gain antenna can boost the signal receiving power and the amplifier can further increase the power. Unfortunately, an amplifier is often powered and thus may add noise to the signal while amplifying it. We have to carefully analyze the link budget by accounting all the gains and losses from the transmitter, through the medium to the receiver of the wireless interface.

To recognize a wireless signal, a wireless network interface card (WNIC) must have the input signal strength greater than the card's sensitivity, the minimum required signal strength at the input receiver [6]. From radio theory, the wireless receiver chain must meet the constraints in Theorem 1. Its proof is in Appendix A of [7].

Theorem 1: To receive a wireless signal,

$$20 \log_{10} D < G_{rx} - NF_{lna} - SNR_{min} + C \quad (1)$$

where D the distance between receiver and transmitter (i.e. coverage radius in free space), G_{rx} the receiver antenna gain, NF_{lna} the noise factor of the low noise amplifier, SNR_{min} is the minimum signal noise ratio of the receiver to have acceptable demodulation accuracy for digitalized baseband circuitry and C is as follows,

$$C = P_{tx} + G_{tx} - 20 \log \frac{4\pi}{\lambda} - 10 \log B - (-174) \quad (2)$$

where P_{tx} is the transmitter power, G_{tx} the transmit antenna gain, λ free space wavelength, -174 (dBm/hz) is the value of the noise power density of the wireless NIC input impedance (normally 50 Ohm), and B is the receiver's bandwidth in Hz, normally defined by the baseband filter bandwidth.

Theorem 1 confirms the intuition of using an antenna with high gain G_{rx} to increase the coverage radius D . A high-performance WNIC with good sensitivity, smaller SNR_{min} , at the sniffer also helps increase the coverage. Indeed, the adversary can also choose a low noise amplifier to boost the coverage area. However, its capability of increasing the area is limited. We can see that the low noise amplifier gain G_{lna} does not play a role in Equation (1). Noise factor F is the ratio of the noise produced by a real resistor to the thermal noise of an ideal resistor. The noise figure NF is the noise factor converted to the decibel notation. After introducing the LNA, we can change the noise figure of the receiver chain. Based on radio theory, the high gain of the LNA renders the noise figure of the receiver chain as the noise figure of the LNA. Without LNA, the noise figure of the receiver chain is that of the WNIC, NF_{nic} . Therefore, the noise figure of the receiver chain with an LNA decreases by $NF_{nic} - NF_{lna}$. A common WNIC has a noise figure around $4.0 \sim 6.0dB$ [8] and the LNA in our experiment is $1.5dB$. We have a noise figure improvement of $2.5 \sim 4.5dB$. This is also the improvement of the signal to noise ratio of the receiver chain and the increase of the coverage area.

Although a LNA has limited capability of increasing the coverage are, the high gain LNA amplifies the signal (as

well as noise) so that we can use a signal splitter to split the amplified signal and feed them to multiple wireless cards for later processing. Recall that a WNIC must have the input signal strength greater than the card's sensitivity. Our RF-Lambda LNA has a gain of 45dB. With a 4-way splitter, each thread of signal (and noise) out of the splitter still achieves $45 - 10 \log 4 = 39dB$ of amplification.

3.2. Probing Traffic Collection

3.2.1. Selection of Sniffing Channels. A main challenge for probing traffic collection is the requirement of monitoring a large number of channels. Both 802.11b (DSSS) and 802.11g (OFDM) wireless LANs have 11 channels, each of which has a frequency width of 22 MHz. The only three channels that do not interfere with each concurrently are channels 1, 6 and 11. However, we cannot use a WNIC dwelling on one channel to capture signals on neighbor channels. Although the signal transmitted along a channel may leak energy to neighboring channels, a card listening on neighboring channels may not correctly recognize the signal because the signal picked up at neighboring channels is distorted and the card cannot decode the signal correctly. To address this problem, a simple solution is to use a total of 11 cards, one for monitoring each channel. Unfortunately, this solution not only incurs significant cost to the system design, but also reduces the mobility of the tracking system. Moreover, to support 802.11a, additional 12 cards are required to monitor 12 non-overlapping channels, leading to a total of $11+12 = 23$ cards.

We propose two solutions to the problem: The first is to use statistical information to *listen on most possible channels*. In reality, most APs use the factory settings which are limited to a few channels. The adversary can perform a field training to identify all existing channels from the received beacon frames or probe response traffic, and listen on only those channels.

If the statistical information indicates more existing channels than the number of available network cards, then our second technique, *frequency hopping*, can be used. The adversary hops between a series of channels, dwells on each channel for a period of time to collect wireless traffic. It is worth noting that this approach may miss wireless frames on channels the sniffer is not listening on.

3.2.2. Collection of Probing Traffic. During data communication, a mobile device communicates with only one AP. To obtain the set of APs communicable with a mobile device, we propose a passive and an active attack to collect probing traffic. In a passive attack, the adversary passively sniffs on wireless traffic and does not interfere with the 802.11 protocol. In an active attack, the adversary may manipulate the wireless frames, exploit the protocols and make the victims send extra frames for more sensitive information.

In both attacks, the adversary utilizes the scanning mechanism of the management protocol in 802.11 a/b/g. Before

a mobile can participate in a BSS (basic service set), it must first use scanning to discover the APs that provide the corresponding service. The default scanning technique for wireless cards is to broadcast *probe request* frames along each available wireless channel. The surrounding APs will respond with a *probe response* frame, containing capability information, supported data rates, etc., when they receive a probe request frame [9]. Our experiments in Section 4.2 confirm the popularity of this technique. As such, our *passive attack* technique simply listens on those active scanning frames and record the set of APs which respond to a mobile's probe request frames.

An adversary may also utilize an *active attack*. It can force a mobile into a position where it will automatically start sending probe request frames by abusing other weaknesses in the 802.11 protocol [10]. The adversary can send spoofed "disassociate" requests with the source MAC address and SSID of the current access point used by the mobile. Once the adversary forces a mobile to disassociate from its legitimate network, the mobile enters the scanning process, sends probe request frames and starts searching for an alternate access point for access. The adversary can then collect probing frames from surrounding APs.

3.3. Malicious Localization Theory

Depending on external knowledge of an adversary, we investigate three possible scenarios for malicious localization: (i) the location and maximum transmission distance of each AP is known through external knowledge, (ii) the location is known but the distance is unknown, (iii) neither the location nor the distance is known. In this subsection, we present the main ideas for malicious localization in the three scenarios respectively and then provide the detailed algorithms.

3.3.1. When Both AP Locations and Maximum Transmission Distances are Known. When the location and maximum transmission distance of each AP is known through external knowledge, we can compute a maximum coverage area for each AP as a disc centered as the AP's location with radius of the maximum transmission distance. Such a disc is a superset of all locations that can communicate with the AP. For locating a mobile device in this scenario, we propose a simple *disc-intersection* approach which computes the intersection of the maximum coverage areas of all APs that the mobile device has communicated according to the monitored probing traffic. Then, the intersected area is used as an estimation of the mobile device's location.

The main challenge for this approach is how small the intersected area can be. The smaller the size is, the more accurate the estimation will be. Fortunately, our experiments show that a mobile device can usually communicate with a large number of APs in practice, particularly in urban areas. The following theorem shows that the size of the intersected area decreases rapidly with the number of communicable APs. The proof of Theorem 2 is in Appendix B of [7].

Theorem 2: When APs with maximum transmission distance r are uniformly distributed, the expected size of the intersected area CA generated by the disc-intersection approach for a mobile device communicable with k APs is

$$CA = \frac{2^{k+3}r^2}{\pi^{k-1}} \int_0^1 y(\cos^{-1} y - y\sqrt{1-y^2})^k dy \quad (3)$$

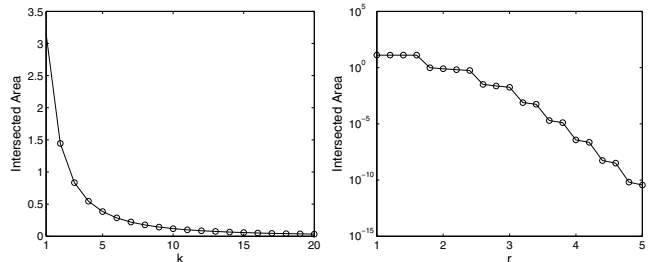


Figure 2: Intersected Area vs Number of Communicable Maximum Transmission Distance

Figure 2 depicts the relationship between the intersected area and number of communicable APs when $r = 1$. They are derived from the theorem using Matlab simulation. As we can see, the intersected area is roughly inversely proportional with the number of communicable APs.

Another interesting observation from the theorem is Figure 3, which depicts the relationship between the intersected area and the maximum transmission distance r when the density of APs (i.e., $\rho = \{\#\text{ of APs}\}/\{\text{Area}\}$) remains constant with $\rho = 0.1$. An interesting observation is that the intersected area size decreases when the maximum transmission range of the mobile device increases. This stands in contrast to the nearest-AP approach, which generates a smaller estimated area when the transmission range is smaller. As we mentioned above, the disc-intersection approach always outperforms the nearest AP approach unless $k = 1$, when both approaches are essentially the same.

3.3.2. When Only AP Locations are Known. In practice, an important challenge for the disc-intersection approach is that the maximum transmission distance varies between different APs, and may not be known through external knowledge. For example, only location but not distance information is available at <http://www.wigle.com>.

A simple approach is to set the maximum transmission distance to a pre-determined value, such as the theoretical upper or lower bound on the transmission distance. Nonetheless, if the value is set too high, the intersected area may become extremely large. If the value is set too low, the mobile device's real location might not be covered by the intersected area (or the area may even become empty).

We propose a linear-programming-based approach to estimate the maximum transmission distance of an AP from the monitored probing traffic. A key observation is that if a mobile device can observe two APs within a short period of

time, then the maximum transmission distances of the two APs, r_1 and r_2 , must satisfy $r_1 + r_2 \geq d_{12}$, where d_{12} is the distance between the two APs which can be computed from their locations pre-known to the adversary. On the other hand, if over a sufficient amount of time, the two APs have never been observed by the same mobile device, then it is highly likely that $r_1 + r_2 < d_{12}$.

As such, we can generate a set of inequalities $r_i + r_j \geq$ (or $<$) d_{ij} from the monitored probing traffic. Considering these inequalities as constraints, we can compute a feasible region for the maximum transmission distances of all APs. Since we prefer overestimates over underestimates, we would like to find a solution in the feasibility region which maximizes $\sum r_i$, the sum of maximum transmission distances for all APs. We solve the optimal maximum through linear programming, and use the solved r_i as the APs' maximum transmission distances. After the maximum transmission distances are estimated, the disc-intersection approach is called to locate the monitored mobile devices.

3.3.3. When No AP Information is Available. When no AP information is available, the adversary must first collect a set of training data tuples before being able to locate the monitored mobile devices. Each training data tuple consists of two parts: an identifier which consists of the longitude and latitude of a training location, and a set of APs a mobile device can communicate with at the training location. Such training data tuples can be collected by using existing wardriving tools such as NetStumbler in a moving vehicle traveling around the monitored area.

After the training data tuples are collected, we propose to compute the location of APs by using, again, the disc-intersection approach. In particular, for each AP, we compute the intersection of discs centered at the training locations which can communicate with the AP. Nonetheless, the exact radius of the discs are unknown and cannot be computed using the linear-programming-based approach due to lack of AP location information. Thus, we propose to use a theoretical upper bound as the radius, and then estimate the AP's location as the centroid of the intersected area.

After the APs' locations are estimated, we estimate the APs' maximum transmission distances using the above-mentioned linear-programming-based approach. Then, we call the disc-intersection approach to locate the monitored mobile devices.

3.4. Malicious Localization Algorithms

Corresponding to the three scenarios discussed above, we develop the following three algorithms for malicious localization. Note that all coordinates used in the three algorithms are for the Earth-Centered, Earth-Fixed (ECEF) Cartesian coordinate system.

Algorithm M-Loc locates a mobile device when APs' locations and radius are provided. The input to M-Loc is

(i) the location and maximum transmission distance (i.e., radius) of each AP, and (ii) a set of APs that are observed to have communicated with the mobile device. The output is an estimated location for the mobile device. In particular, the algorithm first generates all vertices of the intersected area as Δ , and then computes the estimated location as the centroid of all points in Δ .

M-Loc: Localization of mobile based on APs' locations and maximum transmission distances

Require: (i) Location (x_i, y_i) and maximum transmission distance r_i for each AP $_i$; (ii) Γ , the set of APs communicating with the mobile device

- 1: $\Delta_0 = \emptyset, \Delta = \emptyset$.
 - 2: **for** each pair of AP $_i$ and AP $_j$ **do**
 - 3: Compute U as the set of intersected points of the two circles of AP $_i$ and AP $_j$.
 - 4: $\Delta_0 = \Delta_0 \cup U$.
 - 5: **end for**
 - 6: **for** each point (x, y) in Δ_0 **do**
 - 7: **if** $\sqrt{(x - x_j)^2 + (y - y_j)^2} \leq r_j$ for all $j \in \Gamma$ **then**
 - 8: $\Delta = \Delta \cup \{x, y\}$
 - 9: **end if**
 - 10: **end for**
 - 11: Return AVG(Δ)
-

Algorithm AP-Rad estimates the APs' maximum transmission distances based on their locations, and then calls M-Loc to locate a mobile device. The input to AP-Rad is (i) the location of each AP, and (ii) a set of mobile devices and, for each of them, a set of APs that are observed to have communicated with the mobile device. The algorithm generates r_i as the estimated maximum transmission distances for the APs.

Algorithm AP-Loc estimates an AP's location based on training data, and calls AP-Rad and M-Loc for mobile device positioning. The input to AP-Loc is a training dataset which consists of a small number of training locations and, for each, the set of APs that have communicated with the mobile device at that training location. The algorithm generates (x_i, y_i) as the estimated location of the APs.

4. Evaluation

In this section, we evaluate the performance of digital Marauder's map. We first introduce the experiment setup and then discuss the issues of feasibility, coverage area and malicious localization accuracy.

4.1. Experiment Setup

For attacks, we set up the tracking system on the roof of Computer Science Department building at UML and Academic building at GWU. The wireless receiver chain includes one HyperLink 2.4 GHz 15 dBi Omnidirectional Antennas, one RF-Lambda Narrow Band LNA with noise figure

AP-Rad: Localization of mobile based on APs' locations

- Require:** (i) Location (x_i, y_i) for each AP_i ($i \in [1, n]$); (ii) For each mobile device M_k , Γ_k , which is the set of APs communicating with M_k .
- 1: $C = \emptyset$.
 - 2: **for** each pair of AP_i and AP_j **do**
 - 3: $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.
 - 4: **if** there exists Γ_k such that $\{AP_i, AP_j\} \subseteq \Gamma_k$ **then**
 - 5: Add a constraint $r_i + r_j \geq d_{ij}$ into C .
 - 6: **else**
 - 7: Add a constraint $r_i + r_j < d_{ij}$ into C .
 - 8: **end if**
 - 9: **end for**
 - 10: Compute $\{r_1, \dots, r_n\}$ as the result of linear programming with constraints C and maximized function $\sum r_i$.
 - 11: Call M-Loc using input parameters $\{(x_i, y_i)\}, \{r_i\}, \Gamma_k$. Return the estimated location for mobile device M_k .
-

AP-Loc: Localization of mobile based on training data points

- Require:** (i) Location (x_i^t, y_i^t) for each training data point t_i ; (ii) Γ_j^t , the set of t_i communicating with AP_j ; (iii) For each mobile device M_k , Γ_k , which is the set of APs communicating with M_k .
- 1: **for** each AP_j **do**
 - 2: Call M-Loc using input parameters $\{(x_i^t, y_i^t)\}$ and Γ_j^t . Assign the returned value as (x_j, y_j) .
 - 3: **end for**
 - 4: Call AP-Rad using input parameters $\{(x_j, y_j)\}$ and $\{\Gamma_k\}$. Assign the returned value as $\{r_j\}$.
 - 5: Call M-Loc using input parameters $\{(x_j, y_j)\}, \{r_j\}$, and Γ_k . Return the estimated location for mobile device M_k .
-

of 1.5dB, one HyperLink 4-way signal splitter, and three Ubiquiti Super Range Cardbus SRC 300mW 802.11a/b/g wireless cards. To test the accuracy of localization attacks, a mobile device is carried around the campus and the wireless tracking system is used to identify the location of the mobile.

4.2. Feasibility

In the experiments, we tested our earlier claim that most mobile devices actively scan for available access points by sending out probing requests. In particular, we equipped a notebook computer with a Ubiquiti Super Range Cardbus SRC 300mW 802.11a/b/g Wireless Card and a tri-band laptop clip mount 4dBi antenna. By using frequency hopping, the network card monitors all 802.11 a/b/g channels sequentially with a dwell time of 4 seconds. We placed the computer in an office of UML and dumped the wireless traffic by *tcpdump* for a duration of 7 days, from October 24 to October 30, 2008.

Figures 4 and 5 show the statistics of probing mobiles.

From Figures 4 and 5, we have the following observations.

- (i) The percentage of probing mobiles is fairly high. In each day, the percentage of probing mobiles within all found mobiles is above 50%. On Oct. 25, 2008, the ratio is 91.61%. This validates the feasibility of passive attacks. Recall that such percentage can be further improved by the active attack described in Section 4.2.
- (ii) There are more mobiles in weekdays than in weekends. This is because the monitoring sniffer is in a school office and students bring their mobile laptops to school in weekdays. The percentage of probing mobiles is lower in weekdays than in weekends.

4.3. Coverage Area

Figure 6 depicts the coverage radius of different receiver chains. “DLink” refers to that a D-Link DWL-G650 PCMCIA card. “SRC” refers to a Ubiquiti Super Range Cardbus SRC 300mW 802.11a/b/g Wireless Card with a tri-band laptop clip mount 4dBi antenna attached. “HG2415U” refers to a HyperLink high gain (15dB) antenna without an LNA. “LNA” refers to that the HyperLink high gain (15dB) antenna with an LNA attached. The monitoring system is placed on the roof of the Computer Science Department building of UML. A person with a Lenovo X61 Tablet walks around to test the coverage radius of the system.

We make the following important observations from Figure 6. (i) “LNA” achieves the best coverage around 1,000 meters. This validates the theoretical analysis of the link budget in Section 3.1. The experiment was conducted in such a way that the mobile has roughly line of sight to the sniffer on the roof. There were not many walls between the mobile and sniffer. Our extensive experiments show that positioned around the center of the UML north campus and at appropriate height, “LNA” can cover the whole campus. (ii) “HG2415U” can cover as a large area as “LNA”. This is due to the geographical feature of the area. The area is not flat and the sniffer is obstructed by small hills.

4.4. Localization accuracy

We now evaluate the accuracy of the three malicious localization algorithms in Section 3.4, M-Loc, AP-Rad, and AP-Loc, for locating monitored mobile devices. The first two algorithms compromise mobile devices' locations by using external knowledge of APs' locations and/or radius. AP-Loc, on the other hand, requires the collection of training data tuples. Thus, we shall first evaluate the accuracy of M-Loc and AP-Rad, and then discuss the performance of AP-Loc. We shall also demonstrate the superiority of the three algorithms over the existing centroid approach [11] which computes a mobile device's location as the centroid of all APs the mobile device is communicable with.

For M-Loc and AP-Rad, we obtain the locations of APs from WiGLE [2]. For M-Loc, we further obtain the maximum transmission distances of APs by measuring such distance while traveling around the neighborhood of the

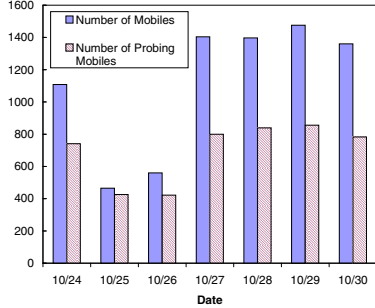


Figure 4: Number of Mobiles and Probing Mobiles

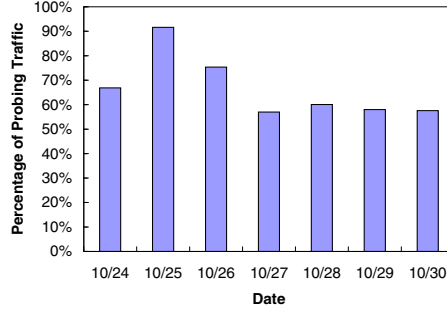


Figure 5: Percentage of Probing Mobiles

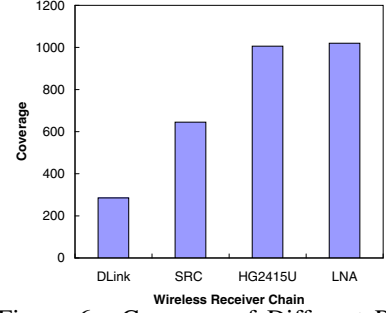


Figure 6: Coverage of Different Receiver Chains

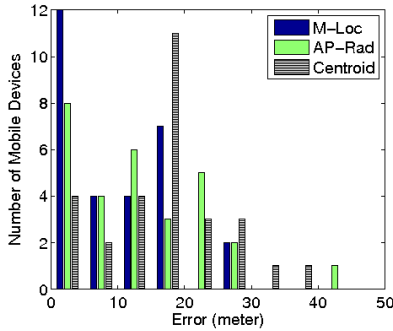


Figure 7: Histogram of Estimation Error for M-Loc, AP-Rad, and Centroid

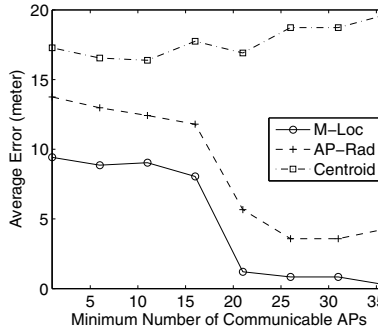


Figure 8: Average Error vs Minimum Number of Communicable APs

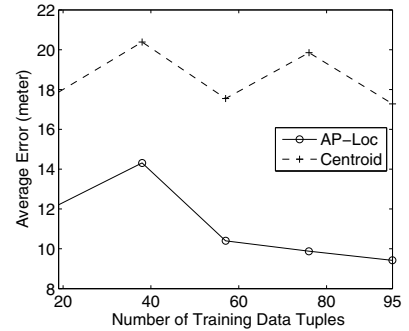


Figure 9: Average Error vs Number of Training Data Tuples

monitoring system with a WiFi-enabled Lenovo X61 Tablet. The training data required for AP-Loc is also collected by traveling around the neighborhood with the Tablet.

Figure 7 depicts the histogram of estimation errors for M-Loc, AP-Rad, and Centroid. One can see from the figure that our two approaches, M-Loc and AP-Rad, achieve significantly better accuracy than the Centroid approach. In addition, M-Loc generates more accurate results than AP-Rad. This is consistent with our intuition that M-Loc benefits from the knowledge of AP radius. In particular, the average estimation error of M-Loc and AP-Rad is only 9.41 and 13.75 meters, respectively, in comparison with an average error of 17.28 meters for the Centroid approach.

Figure 8 depicts the relationship between the average estimation error and the minimum number of APs communicating with a mobile device, observed by the tracking system. Again, one can see the superiority of our M-Loc and AP-Rad approaches over the simple Centroid approach. Another interesting observation is that our approaches (particularly M-Loc) has average error monotonically decreasing with the number of communicable APs, while the average error of Centroid is increasing. This is consistent with our discussion in Section 3.3 that, due to the vulnerability of the Centroid approach on skewed AP distributions, more communicable APs may increase the estimation error for Centroid, but will always reduce the error for our M-Loc approach.

We now evaluate the performance of AP-Loc. In particu-

lar, we consider the relationship between the average error and the number of training data tuples used in AP-Loc. Figure 9 depicts such relationship. As we can see, AP-Loc achieves much better accuracy than the Centroid approach even when the number of training tuples is fairly small. For example, given 19 training tuples, AP-Loc can achieve an average error of only 12.21 meters for locating monitored mobile devices.

5. Related Work

There are a large number of brilliant papers on positioning and localization in wireless networks. Because of the page limit, here we will only briefly review most related work.

Location Privacy Protection: Much work has been done on the protection of location privacy in wireless networks. The existing work can be generally classified into two categories: location data protection and network identity hiding. To prevent location data from leaking sensitive location information, Temporal and Spatial Cloaking [12] uses a special middleware agent providing positioning data that would preserve only the location resolution essential for location-based applications. The agent sends the data to a location-based application via an anonymous communication network. There are a large number of data based location privacy works such as [13].

To hide a mobile's identity, Mix Zones [14] uses the idea of silent zones, where users keep silent by not sending any

requests in order to mix the identities of people within this zone. This approach may incur extensive inconvenience. Hu and Wang [15] present a framework of location privacy using random identity addresses such as IP and MAC addresses and random silent period in which mobile nodes don't transmit or receive frames. They implemented a similar framework in [16]. Singele and Preneel [17] presented cryptographic protocols for randomizing mobile identifiers. However, Pang *et al.* [4] demonstrate that many implicit identifiers such as network names in probing traffic may break those pseudonyms.

Range-Free Positioning in Sensor Networks: Our work is also related to various range-free localization approaches in wireless sensor networks [18], [19], [11], [20], by which a sensor locates itself. The basic idea is that the sensor senses its surrounding anchor nodes and calculate its location based on the locations of anchor nodes. Those schemes generally require corresponding sensing protocols for self-positioning. The problem investigated in this paper is different from that in sensor networks because our attack is initialized by a malicious third party without support of those positioning protocols in sensor networks. The complicated WiFi environment also poses a great challenge.

6. Conclusion

In this paper, we presented the digital Marauder's map, a malicious wireless tracking system to locate mobile devices in WiFi networks. The system consists of a wireless receiver chain, a probing traffic capturing component, a malicious localization component, and the display for digital Marauder's map. We present a comprehensive set of theoretical analysis and experimental results which demonstrate the coverage and localization accuracy of the digital Marauder's map. To the best of our knowledge, our work is the first to study attacks against location privacy through a full-fledged malicious tracking system utilizing high-gain antennas in WiFi networks. We expect the results of this paper to stimulate the implementation of a set of mobile identity camouflaging protocols to preserve user location privacy in pervasive WiFi networks.

Acknowledgement

This work was supported in part by the National Science Foundation under grants 0907964, 0722856, 0324988, 0329181, 0721571, 0808419, 0845644, 0852673, and 0852674. Any opinions, findings, conclusions, and/or recommendations expressed in this material, either expressed or implied, are those of the authors and do not necessarily reflect the views of the sponsor listed above.

References

[1] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis, "A study on the value of location privacy," in *Proceedings of the 5th ACM workshop on Privacy in electronic society*, 2006.

[2] Arkasha and Bobzilla, "WiGLE - wireless geographic logging engine - plotting wifi on maps," <http://www.wigle.net/>, 2008.

[3] M. Gast, *802.11 Wireless Networks: The Definitive Guide, Second Edition (Definitive Guide)*. O'Reilly Media, Inc., April 25 2005.

[4] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, "802.11 user fingerprinting," in *Proceedings of the 13th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Sep. 2007.

[5] Google, "What is the google maps api?" <http://code.google.com/apis/maps/>, October 2008.

[6] J. Zyren and A. Petrick, "Tutorial on basic link budget analysis," <http://sss-mag.com/pdf/an9804.pdf>, June 1998.

[7] X. Fu, N. Zhang, A. Pingley, W. Yu, J. Wang, and W. Zhao, "The digital marauder's map," Department of Computer Science, UMass Lowell, <http://www.cs.uml.edu/~xinwenfu/Marauder.pdf>, Tech. Rep., November 2008.

[8] E. Perahia and S. Li, "doc.: Ieee 802.11-06/0330r4," <http://www.ieee802.org/19/pub/2006/11-06-0338-03-000n-p802-11n-ca-document.doc>, March 2006.

[9] V. Gupta, "A characterization of wireless network interface card active scanning algorithms," Master Thesis, Georgia State University, 2006.

[10] M. Kershaw and J. Wright, "802.11b firmware-level attacks," <http://code.google.com/apis/maps/>, September 2006.

[11] N. Bulusu, J. Heidemann, and D. Estrin, "Gps-less low cost outdoor localization for very small devices," *IEEE Personal Communications Magazine*, vol. 7, no. 5, pp. 28-34, October 2000. [Online]. Available: <http://lecs.cs.ucla.edu/~bulusu/papers/Bulusu00a.html>

[12] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of MobiSys*, 2003.

[13] B. Gedik and L. Liu, "Location-privacy in mobile systems: A personalized anonymization model," in *Proceedings of International Conference on Distributed Computing Systems (ICDCS)*, 2005.

[14] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," in *Proceedings of Information Hiding Workshop*, 2003.

[15] Y.-C. Hu and H. J. Wang, "Location privacy in wireless networks," in *Proceedings of the ACM SIGCOMM Asia Workshop*, April 2005.

[16] T. Jiang, H. J. Wang, and Y.-C. Hu, "Location privacy in wireless networks," in *Proceedings The 5th International Conference on Mobile Systems, Applications, and Service (MobiSys)*, June 2007.

[17] D. Singele and B. Preneel, "Location privacy in wireless personal area networks," in *Proceedings of the 5th ACM workshop on Wireless security (WiSe)*, 2006.

[18] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of MobiCom*, 2003.

[19] Y.-Q. Chen, Y.-K. Kim, and S.-J. Yoo, "Accurate sensor position estimation for wireless sensor networks using neighborhood relationship," *IEICE Transactions on Communications*, vol. E91-B, no. 9, September 2008.

[20] J. Hwang, Y. Gu, T. He, and Y. Kim, "Realistic sensing area modeling," in *Proceedings of the 26th Conference on Computer Communications (INFOCOM) Minisymposia*, May 2007.