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

Xinwen Fu, University of Massachusetts Lowell
Nan Zhang & Aniket Pingley, George Washington University
Wei Yu, Cisco Systems, Inc.
Jie Wang, University of Massachusetts Lowell
Wei Zhao, Rensselaer Polytechnic Institute

Abstract—“The Marauder’s Map” is a magical map in J. K. Rowling’s fantasy series, “Harry Potter and the Prisoner of Azkaban”. When being used by wizards who “solemnly swear that I am up to no good”, it shows all moving objects within the boundary of the “Hogwarts School of Witchcraft and Wizardry”. In this paper, we introduce 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.

I. INTRODUCTION

In this paper, we study a class of novel localization attacks to compromise the location privacy of mobile devices in wireless 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 users’ location privacy in wireless networks.

Much work has been done on mobile positioning in WiFi networks [2]–[8]. Nonetheless, the focus was almost exclusively on providing positioning services, instead of localization attacks from a malicious third party [1]. We can classify the existing positioning techniques into four categories: (i) RF signal-strength fingerprinting [9], [10], (ii) Trilateration [2] based on the received signal strength of a device at access points (AP), (iii) Triangulation [8] based on information provided by access points which are equipped with directional antenna and can measure the angle of arrival (AOA) of a wireless signal. (iv) Closest AP [5] directly using the location of APs or sensors with the strongest signal strength.

While effective for mobile devices’ self-positioning, the above-mentioned schemes cannot be directly used by a real-world adversary to locate mobile devices. RF Fingerprinting based positioning often requires formidable training of WiFi networks which must be repeated once some changes (e.g., the addition of buildings) are made to the environment. Trilateration and triangulation techniques are ineffective in urban areas because obstructing buildings often prevent the signal strength and AOA from being accurately measured. The Closest Access Point approach provides poor localization accuracy due to the large coverage area of an AP or requires deployment of a large number of sensors [3], [4].

This paper presents 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 with reasonable implementation cost, 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 [11]. 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 [12] 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:
- 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.
- 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.
- We present M-Loc, AP-Rad, and AP-Loc, three local-
Data localization algorithms for a third-party attacker to accurately locate all monitored mobile devices. The success of our techniques challenges the traditional belief that signal strength information is necessary for the accurate positioning of mobile devices.

- 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. [13] demonstrate 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. In Section II, we introduce the digital Marauder’s map, the WiFi tracking system and discuss its components and related issues. The localization attack theory and algorithms will be given in details in Section III. We evaluate the entire system in Section IV. In Section V, we review the related work. In Section VI, we conclude this paper and discuss future work.

II. SYSTEM

In this section, we develop the main ideas behind the digital marauder’s map, our system for malicious wireless tracking.

A. 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 [12], [14]. This stage is optional because such training data is often available through wireless geographic logging websites. For example, the location of 15 million APs across the world is available at WiGLE [11]. 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 [15] or Kismet [16]. Then, the adversary travels through the target area where the sniffing tools constantly probe APs and record training data including (i) the wireless packets transmitted between APs and the mobile device, and (ii) the spatial coordinates at which those wireless packets are captured. After the collection of training data, the adversary uses the AP-Loc algorithm we will introduce later in the paper to estimate the location of APs.

In the attacking phase, an adversary compromises the location of a wireless device in two steps: First, it identifies a set of APs communicable to 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.

B. 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:

1. The wireless receiver chain. It includes a number of high gain antennas, low noise amplifiers (LNAs), signal spitters, 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.

2. Wireless traffic capture. Each thread of wireless signal is captured by a wireless card, which processes and extracts useful information such as SSIDs and AP MAC addresses from the recorded 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.

3. 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.

4. 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 [17] to overlay the location on top of topology map.

C. 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.

### III. Analysis and Algorithms of Malicious Localization

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.

#### A. 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 (free space, antennas, etc.) 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 [18]. From radio theory, the wireless receiver chain must meet the constraints in Theorem 1. The proof of Theorem 1 can be found in Appendix A in [19].

**Theorem 1.** To receive a wireless signal,

\[
20 \log_{10} D < G_{rx} - N F_{lna} - S N R_{min} + C
\]

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

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

where \(P_{tx}\) is the transmitter power, \(G_{tx}\) the transmit antenna gain, \(\lambda\) 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, which is normally defined by the baseband filter bandwidth.

Theorem 1 is derived via the free space radio transmission model. In urban areas, there are various obstacles such as walls and humans, which cause more electromagnetic wave attenuation. In this paper, the model in Theorem 1 is used as a worst case model locating a mobile since this spherical model overestimates the AP coverage and reduces localization accuracy. Theorem 1 also guides the selection of appropriate electronic components improving the coverage.

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 \(S N R_{min}\), 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 (18). 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 \(N F\) 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, \(N F_{nic}\). Therefore, the noise figure of the receiver chain with an LNA decreases by \(N F_{nic} - N F_{lna}\). A common WNIC has a noise figure around 4.0 ~ 6.0dB [20] and the LNA in our experiment is 1.5dB [21]. We have a noise figure improvement of 2.5 ~ 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 area, 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 [21]. With a 4-way splitter, each thread of signal (and noise) out of the splitter still achieves 45 ~ 10log4 = 39dB of amplification.

#### B. Probing Traffic Collection

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. It was believed [22] that, to capture all 802.11b/g signals at 2.4GHz, we need at least 3 wireless cards, \(NIC_3\), \(NIC_6\), and \(NIC_9\), to monitor Channels 3, 6, and 9, respectively, such that \(NIC_3\) picks up signals for Channels 1 to 5, \(NIC_6\) for Channels 4 to 8, and \(NIC_9\) for Channels 7 to 11. Nonetheless,
our experiments in Section IV (Figure 9) shows this claim does not hold in practice. 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, support for 802.11a requires 12 cards if the location and maximum transmission distance of each AP is known through external knowledge or the distance is unknown, respectively and then provide the detailed algorithms.

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.

One can see from the approach that, as long as the APs’ locations and maximum transmission distances are accurate, the mobile device’s real location is always covered in the intersected area. Thus, 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.

When a mobile device can only communicate with one AP, the intersected area is the maximum coverage area of the AP, and the disc-intersection approach is essentially reduced to the nearest AP approach. 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 can be found in Appendix B in [19].

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

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

Figure 2 depicts the relationship between the intersected area and the number of communicable APs when \( r = 1 \). The results are computed 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
The centroid of communicable APs (i.e., essentially the same. The nearest AP approach unless $k = 1$, the disc-intersection approach always outperforms the density of APs (i.e., $\rho$) when the maximum transmission range of the mobile device increases. This stands in contrast to the $\rho$-steer 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.

We now compare the disc-intersection approach with the previous approach of estimating a mobile device’s location as the centroid of communicable APs (i.e., $x = \sum x_i/n$, $y = \sum y_i/n$, where $(x_i, y_i)$ are the coordinates of communicable APs). When all APs are uniformly distributed (as assumed in the theorem), the two approaches have similar performance as the expected centroid of the intersected area is also the expected centroid of all observable base stations.

However, our disc-intersection approach is significantly more resilient to biased AP distributions than the centroid approach. Consider an example demonstrated in Figure 4, where APs are drawn from a composite of two distributions: 5 APs, $A_1, \ldots, A_5$, distributed uniformly at random in the entire area, and other 10 APs, $A_6, \ldots, A_{10}$, which are located in a small gray area. One can see from the figure that the estimation of centroid approach given $A_1, \ldots, A_{10}$ is much less accurate than given $A_1, \ldots, A_5$ only. On the other hand, note that our approach can only become more accurate when the number of base stations increases because the intersected area can only shrink instead of grow. Thus, our approach becomes significantly more accurate than the centroid approach given the biased distribution of $A_1, \ldots, A_{10}$.

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). The following theorem shows the relationship between the performance of the disc-intersection approach and the maximum transmission distance. The proof of Theorem 3 can be found in Appendix C in [19].

**Theorem 3.** When APs with maximum transmission distance $r$ are uniformly distributed and the disc-intersection approach set estimated distance $R$, then if $R \geq r$, the expected size of the intersected area for a mobile device communicable with $k$ APs is

$$CA = \frac{1}{\pi^{k-1}2^k} \int_0^{2R} \left( r^2 \cos^{-1} \left( \frac{x^2 + r^2 - R^2}{2rx} \right) + \frac{R^2 \cos^{-1} \left( \frac{x^2 + R^2 - r^2}{2xR} \right) - \sqrt{(r + R)^2 - x^2}(x^2 - (r - R)^2)}{2} \right)^k dx^2$$

If $R < r$, the probability that the intersected area covers the real location of the mobile device is

$$p = \left( \frac{r}{R} \right)^{2k}$$

Figure 5 and 6 depicts how the intersected area and the coverage probability changes with the estimated maximum transmission distance, respectively, when $k = 10$ and $r = 1$. One can make two observations from the figures: First, an overestimate of $r$ is clearly preferred over an underestimate because when $r' < r$, the probability of the intersected area covering the real location quickly becomes extremely small when $k$ is large. Second, when $r' > r$, the expected size of the intersected area grows rapidly with $r'$. Thus, a theoretical upper bound also does not suffice for the estimation.

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 $\sum r_j$, the sum of maximum transmission distances for all APs. We solve the optimal maximum through linear programming, and use the solved $r_j$ 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) 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

![Fig. 2. Intersected Area vs Number of Communicable APs](image1)

![Fig. 3. Intersected Area vs Maximum Transmission Distance](image2)
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.

D. 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 \( \Delta \), and then computes the estimated location as the centroid of all points in \( \Delta \).

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 tuples, and then 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.

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) \(\Gamma\), 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 \(j\), \(U\) may be empty or contains one or two points.
4: \(\Delta_0 = \Delta_0 \cup U\).
5: end for
6: for each point \((x, y)\) in \(\Delta_0\) 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(\(\Delta\))

IV. Evaluation

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

A. Experiment Setup

We conducted experiments on two campuses: University of Massachusetts Lowell (UML) and George Washington University (GWU). Recall that our system requires an optional training phase and a mandatory attacking phase. We test both scenarios where training is or is not required, respectively. In particular, to test the case where training is not required, we downloaded AP location information from WiGLE. To test the case where training is required, we equipped a Lenovo
University campus demonstrates similar channel distribution. The data is collected via Kismet. The George Washington University campus demonstrates similar channel distribution around the UML campus. Figure 8 shows the channel distribution around the UML campus.

```
AP-Rad: Localization of mobile based on APs’ locations

Require: (i) Location \((x_i, y_i)\) for each AP \(i \in [1, n]\); (ii) \(\Gamma_i\), the set of APs communicating with \(M_i\); (iii) \(\Gamma_k\), 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 \(\Gamma_k\) 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 \(\{r_1, \ldots, r_n\}\) as the result of linear programming with constraints \(C\) and maximized function \(\sum r_j\).
11: Call M-Loc using input parameters \(\{(x_i, y_i)\}, \{r_i\}, \Gamma_k\).
Return the returned value as 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) \(\Gamma_j^t\), the set of APs communicating with \(AP_j\); (iii) \(\Gamma_k\), 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 \(\Gamma_j^t\),
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 \(\Gamma_k\). Return the returned value as the estimated location for mobile device \(M_k\).
```

ThinkPad W500 Notebook with DLink 802.11g Wireless Cardbus Adapter, and used wardriving tool NetStrumbler to collect training data.

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 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. The display of the digital Marauder’s map is illustrated in Figure 7 using Google maps. It shows the location of APs, the real mobile location in red tags and estimated mobile location in blue tags.

To determine how many cards we need to use, we conduct experiments and collect 802.11 channel distribution. Figure 8 shows the channel distribution around the UML campus. The data is collected via Kismet. The George Washington University campus demonstrates similar channel distribution pattern. We can see that most APs (93.7%) use Channels 1, 6 and 11. So we chose to use three cards on three Lenovo W500 Notebooks to monitor these three channels, respectively. The three Lenovo W500 laptops are time synchronized through NTP (network time protocol). Figure 9 shows that when a wireless card is sending packets on Channel 11, other cards listening on neighboring channels can recognize few or none of those packets. This verifies our claim in Section III-B1.

```
Fig. 7. The Digital Marauder’s Map

Fig. 8. Channel Distribution around UML North Campus
```

### B. 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 [25]. 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 10 and 11 show the statistics of probing mobiles. From Figures 10 and 11, we have the following observations.

- 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 IV-B.
D. Localization accuracy

Obstructed by small hills, the area is not flat and the sniffer covers a large area as "LNA". This is due to the geographical feature of the area. The monitoring system is placed on the roof of the Computer Science Department building of UML. A person with a Lenovo X61 Tablet PC walks around the neighborhood to test the coverage radius of the system.

The experiment was conducted in such a way that the mobile has roughly line of sight to the sniffer on the center of the UML north campus and at appropriate height. Our extensive experiments show that positioned around the neighborhood with the Lenovo X61 Tablet PC.

The validates the theoretical analysis of the link budget in Section III-A. The experiment was conducted in such a way that the mobile has roughly line of sight to the sniffer on 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.

For M-Loc and AP-Rad, we obtain the locations of APs from WiGLE [11]. For M-Loc, we further obtain the maximum transmission distances of APs by measuring such distance while traveling around the neighborhood of the monitoring system with a WiFi-enabled Lenovo X61 Tablet PC. The training data required for AP-Loc is also collected by traveling around the neighborhood with the Lenovo X61 Tablet PC.

C. Coverage Area

Figure 12 depicts the coverage radius of different receiver chains. “DLink” refers to a D-Link DWL-G650 PCMCIA card. “SRC” refers to a Ubiquiti Super Range Cardbus SRC 300nW 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 PC walks around the neighborhood to test the coverage radius of the system.

We make the following important observations from Figure 12. (i) “LNA” achieves the best coverage around 1,000 meters. The validates the theoretical analysis of the link budget in Section III-A. 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.

D. Localization accuracy

We now evaluate the accuracy of the three malicious localization algorithms in Section III-D, 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 [26] which computes a mobile device’s location as the centroid of all APs the mobile device is communicable with.

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.

For M-Loc and AP-Rad, we obtain the locations of APs from WiGLE [11]. For M-Loc, we further obtain the maximum transmission distances of APs by measuring such distance while traveling around the neighborhood of the monitoring system with a WiFi-enabled Lenovo X61 Tablet PC. The training data required for AP-Loc is also collected by traveling around the neighborhood with the Lenovo X61 Tablet PC.

Figure 13 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 14 depicts the relationship between the average estimation error and the minimum number of communicable APs observed by a mobile device. 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 III-C 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.

Figure 15 depicts the relationship between the intersected area and the minimum number of communicable APs observed by a mobile device. As we can see, AP-Rad generates a larger intersected area than M-Loc. This is due to the error on the estimation of APs’ radius in AP-Rad. Figure 16 depicts the relationship between the probability of the intersected area covering the mobile device’s location and the minimum number of communicable APs observed by a mobile device. Again, the estimation error on APs’ radius leads to a lower coverage probability for AP-Rad.

We now evaluate the performance of AP-Loc. In particular, we consider the relationship between the average error and the number of training data tuples used in AP-Loc. Figure 17 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.

V. RELATED WORK

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

Positioning Systems: Positioning systems are classified as outdoor and indoor systems due to their vastly different requirements and techniques. The most popular outdoor positioning system is GPS [2]. Many cellular mobile networks allow for the tracking of powered-on handsets through the operator base transceiver stations. Indoor positioning systems include RADAR [27], LANDMARC [6], Lighthouse [7] and VORBA [8]. All these systems position a mobile device based on the measured signal strength. In particular, the former two utilize a dense grid of omnidirectional base stations, while the latter two rely on base stations with revolving unidirectional antennas. Active Badge [4], Active Bat [3] and Cricket [5] can provide better localization accuracy than outdoor systems due to the usage of a large number of positioning-support sensors.

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 [28] 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 [29].

To hide a mobile’s identity, Mix Zones [30] 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 [31] 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 [32]. Singele and Preneel [33] presented cryptographic protocols for randomizing mobile identifiers. However, Pang et al. [13] 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 [26], [34]–[36], 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.

VI. 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 used radio theory to analyze the coverage area of the receiver chain, and deployed a high-gain antenna and a low-noise amplifier to boost the receiving power and the signal-to-noise ratio, which ultimately enlarges the coverage area. For probing traffic collection, we proposed both passive and active attacks which can be used by the adversary to collect probing traffic over mobile devices. For the malicious localization component, we presented three malicious localization algorithms when both AP locations and maximum transmission distances are known (M-Loc), when only AP locations are known (AP-Rad), and when no AP information is available (AP-Loc), respectively. We conducted extensive real-world experiments and validated the feasibility of the proposed attacks. The experimental results showed that the digital Marauder’s map covers a radius of over one kilometer in the northern campus of UMass Lowell. We also tested the accuracy of the malicious localization algorithms, and found that the average error can be limited to around 10 meters which poses a serious threat to location privacy. We also demonstrated the superiority of our algorithms over the existing centroid approach.

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.

REFERENCES

Theorem 1. To receive a wireless signal,
\[ 20 \log_{10} D < G_{rx} - N_{F_{lna}} - SNR_{min} + C \] (6)
where \( D \) the distance between receiver and transmitter (i.e. coverage radius in free space), \( G_{rx} \) the receiver antenna gain, \( N_{F_{lna}} \) 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,
\[ C = P_{tx} + G_{tx} - 20 \log \frac{4\pi}{\lambda} - 10 \log B - (174) \] (7)
where \( P_{tx} \) is the transmitter power, \( G_{tx} \) the transmit antenna gain, \( \lambda \) 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, which is normally defined by the baseband filter bandwidth.

Proof:
To recognize a wireless signal, a wireless network interface card (WNIC) must have the input signal strength, \( P_{rx} \), greater than its sensitivity, \( P_{rx,min} \), which is the minimum required signal strength at the input receiver [18]. The receiver signal strength can be estimated as follows,
\[ P_{rx} = P_{tx} + G_{tx} + G_{rx} - L_{fs} \] (8)
\[ L_{fs} = 20 \log_{10} \left( 4 \times \pi \times \frac{D}{\lambda} \right) \] (9)
where \( P_{tx} \) is the transmitter power, \( G_{tx} \) the transmit antenna gain, \( G_{rx} \) the receiver antenna gain, \( L_{fs} \) propagation loss, \( D \) the distance between receiver and transmitter, \( \lambda \) free space wavelength. Please note: fade margin is ignored in (8) for brevity and fade margin is the loss caused by signal multipath fading. Therefore,
\[ P_{rx} = P_{tx} + G_{tx} + G_{rx} - 20 \log \left( \frac{4\pi D}{\lambda} \right) \] (10)

The sensitivity can be given by following equation:
\[ P_{rx,min} = -174 + NF + SNR_{min} + 10 \log B \] (11)
Where -174 (dBm/Hz) is the value of the noise power density of the WNIC input impedance (normally 50 Ohm), \( NF \) is the noise figure of the receiver chain in dB, \( SNR_{min} \) is the minimum signal-to-noise ratio of the receiver to have acceptable demodulation accuracy for digitalized baseband circuitry, and \( B \) is the receiver’s bandwidth in Hz, which is normally defined by the baseband filter bandwidth.

The noise factor \( F \) is the ratio of the noise produced by a real resistor to the simple thermal noise of an ideal resistor. The noise figure \( NF \) is the noise factor converted to decibel notation. Apparently, our wireless receiver chain consists of a few cascaded blocks, the antenna, a connector connecting the antenna to the powered low noise amplifier (LNA), a LNA, a signal splitter, and the wireless NIC. The \( NF \) of the cascaded receiver blocks can be calculated using noise factor \( F \) of each block,
\[ F = F_1 + \frac{F_2}{G_1} + \cdots + \frac{F_n}{G_1 G_2 G_3 \cdots G_{n-1}} \] (12)
\[ NF = 10 \log F \] (13)
where \( F_i \), \( G_i \) are the noise factor and voltage gain of the \( i^{th} \) block respectively, and \( F \) is the entire noise factor of the cascaded receiver chain. We assume that non-powered blocks don’t introduce noise, and only the LNA and WNIC introduce thermal noise. Their noise factors are denoted as \( F_{lna} \) and \( F_{nic} \) and the gain of the LNA is \( G_{lna} \).
\[ NF = 10 \log \left( F_{lna} + \frac{F_{nic} - 1}{G_{lna}} \right) \] (14)
Since \( G_{lna} \) can be very large, \( (F_{nic} - 1)/G_{lna} \) can be ignored. Therefore,
\[ NF = 10 \log (F_{lna}) = N_{F_{lna}} \] (15)
Substitute (15) into (11), we have
\[ P_{rx,min} = -174 + N_{F_{lna}} + SNR_{min} + 10 \log B \] (16)
Recall that to receive a wireless signal, the wireless receiver chain must meet the following constraint,
\[ P_{rx} > P_{rx,min} \] (17)
therefore,
\[ 20 \log_{10} D < G_{rx} - N_{F_{lna}} - SNR_{min} + C \] (18)

APPENDIX A

Theorem 2. When APs with maximum transmission distance \( r \) are uniformly distributed, the expected size of the intersected area (or cross area CA) generated by the disc-intersection approach for a mobile device communicable with \( k \) APs is
\[ CA = \frac{2^{k+3} \pi^{2}}{\pi^{k-1}} \int_{0}^{1} y (\cos^{-1} y - y \sqrt{1 - y^2})^{k} dy \] (20)

Proof:
Assume in an area, APs are uniformly distributed. Those APs are of the same brand with the same factory parameter setting. Their transmission contour follows a sphere model and is a circle in the two-dimensional plane. Therefore, a mobile’s receiver range is a circle with radius \( r \). There are \( k \) APs within the mobile receiver range.

If there is a point \( A \), which is in the intersected area generated by the disc-intersection approach, then \( A \) must be able to see all the \( k \) APs that the mobile can see. This implies all the \( k \) APs are within the shaded area in Figure 18. Assume that \( A \) is \( x \) meters away from the mobile. The shaded area is the intersected area of the mobile’s transmission range disc and the disc of radius \( r \) with \( A \) as the center. The intersected area can be calculated as follows,
\[ CA_x = 2r^2 \cos^{-1} \frac{x}{2r} - \frac{x}{2} \sqrt{4r^2 - x^2} \] (21)
The probability, \( p_x \), that an AP and the mobile are in the same intersected area, given that A is at \( x \), is

\[
p_x = \frac{CA_x}{\pi r^2}
\]

\[
= \frac{2}{\pi r^2} \cos^{-1} \left( \frac{x}{2r} \right) \sqrt{1 - (x/2r)^2}
\]

\[
= \frac{2}{\pi r^2} \cos^{-1} \left( \frac{x}{2r} \right) \frac{\pi}{\sqrt{1 - (x/2r)^2}}
\]

The probability that \( k \) APs and the mobile are in the same area is \( p_k \), given that point A is \( x \) meters away from the mobile. Since A is at most \( 2r \) away from the mobile and A is within an area of \( d(\pi x^2) \) (given that A is \( x \) meters away from the mobile), the mean CA can be calculated as follows

\[
CA = \int_{x=0}^{2r} p_k d(\pi x^2) = 2\pi \int_{x=0}^{2r} xp_k dx
\]

Let \( y = x/2r \),

\[
CA = 8\pi r^2 \int_{0}^{1} y p_k^k dy
\]

\[
= 8\pi r^2 \int_{0}^{1} y (2\cos^{-1} y - y\sqrt{1 - y^2})^k dy
\]

\[
= \frac{2^{k+1}}{\pi^{k-1}} \int_{0}^{1} y (\cos^{-1} y - y\sqrt{1 - y^2})^k dy
\]

**Corollary 1.** CA in Theorem 2 monotonically decreases with receiver range radius \( r \) and AP density \( \rho \).

**Proof:** We can see from Theorem 2 that CA is a function of \( k \), and

\[
\frac{\partial CA}{\partial k} = \frac{r^{2k+3}}{\pi^{k-1}} \int_{0}^{1} \text{term}_1\text{term}_2 dy
\]

\[
\text{term}_1 = y(\arccos y - y\sqrt{1 - y^2})^k
\]

\[
\text{term}_2 = -\ln \frac{\pi}{2} + \ln(\cos^{-1} y - y\sqrt{1 - y^2})^k
\]

Since \( 0 \leq y \leq 1 \), \( \text{term}_1 \geq 0 \) while \( \text{term}_2 \leq 0 \) so that \( \text{term}_1\text{term}_2 \leq 0 \). Therefore,

\[
\frac{\partial CA}{\partial k} < 0
\]

CA is a decreasing function of \( k \), the number of APs.

Assume in an area, APs are uniformly distributed. The density is \( \rho \). If there are \( k \) APs within the mobile receiver range, then

\[
k = \pi r^2 \rho
\]

Since \( k \) is an increasing function of \( \rho \), and \( CA \) decreases with \( k \) (the number of APs), \( CA \) monotonically decreases with AP density \( \rho \) too.

**APPENDIX C**

**Theorem 3.** When APs with maximum transmission distance \( r \) are uniformly distributed and the disc-intersection approach set estimated distance \( R \), then if \( R \geq r \), the expected size of the intersected area for a mobile device communicable with \( k \) APs is

\[
CA = \frac{1}{\pi^{k-1}r^{2k}} \int_{0}^{2R} \left( r^2 \cos^{-1} \left( \frac{x^2 + r^2 - R^2}{2xr} \right) + \right.
\]

\[
R^2 \cos^{-1} \left( \frac{x^2 + R^2 - r^2}{2xR} \right) - \sqrt{(r + R)^2 - x^2}(x^2 - (r - R)^2) \right)^k \frac{d x^2}{2}
\]

If \( R < r \), the probability that the intersected area covers the real location of the mobile device is

\[
p = \left( \frac{R}{r} \right)^{2k}
\]

**Proof:** Let \( C_1 \) be a circle centered at the location of the mobile device and has radius of \( r \). Let \( \Omega \) the set of all APs communicable with the mobile device. One can see that every AP in \( \Omega \) is located within \( C_1 \).

We focus on the case where \( R \geq r \) first. Let \( \Theta \) be the intersected area. Consider a location \( \alpha \) with distance \( x \) from the mobile device. We only consider the cases where \( x \leq 2R \) because otherwise it is impossible for \( x \) to fall into \( \Theta \). Let \( C_2 \) be a circle centered at \( \alpha \) and has radius of \( R \). Clearly, \( \alpha \) falls into \( \Theta \) iff all APs in \( \Omega \) are located within \( C_2 \). Since all APs in \( \Omega \) are also located within \( C_1 \), \( \alpha \) falls into \( \Theta \) iff all APs in \( \Omega \) are located within the overlapping area of \( C_1 \) and \( C_2 \), denoted by \( C_{12} \). Formally,

\[
\Pr \{ \alpha \in \Theta \} = \Pr \{ \Omega \subseteq C_{12} \}.
\]

The area of \( C_{12} \) is determined by \( R, r, \) and \( x \), and can be easily derived with plane geometry:

\[
A(C_{12}) = r^2 \cos^{-1} \left( \frac{x^2 + r^2 - R^2}{2xr} \right) + \frac{R^2 \cos^{-1} \left( \frac{x^2 + R^2 - r^2}{2xR} \right) - \sqrt{(r + R)^2 - x^2}(x^2 - (r - R)^2)}{2}
\]

Since the APs are i.i.d. with uniform distribution on the area covered by \( C_1 \), the probability that an AP falls into \( C_{12} \)
is $A(C_{12})/(\pi r^2)$, where $\pi r^2$ is the area of $C_1$. Thus, the probability that all APs in $\Omega$ fall into $C_{12}$ is

$$\Pr\{\Omega \subseteq C_{12}\} = \left(\frac{A(C_{12})}{\pi r^2}\right)^k, \quad (38)$$

Since $\Pr\{\alpha \in \Theta\} = \Pr\{\Omega \subseteq C_{12}\}$, the expected size of the overlapping area is

$$CA = \pi \int_0^{2R} \Pr\{\alpha \in \Theta\} dx^2 \quad (39)$$

$$= \pi \int_0^{2R} \left(\frac{A(C_{12})}{\pi r^2}\right)^k dx^2 \quad (40)$$

$$= \frac{1}{\pi^{k-1}2^k} \int_0^{2R} (A(C_{12}))^k dx^2 \quad (41)$$

We now consider the case where $R < r$. Let $C_3$ be a circle centered at the mobile device and has radius of $R$. If the real location of the mobile device falls into $\Theta$, then all APs in $\Omega$ must be located within $C_3$ (i.e., $\Omega \subseteq C_3$). We have

$$\Pr\{\Omega \subseteq C_3\} = \left(\frac{\pi R^2}{\pi r^2}\right)^k = \left(\frac{R}{r}\right)^{2k}. \quad (42)$$

Thus, the probability for the intersected area to cover the real location of the mobile device is $p = (R/r)^{2k}$. ■