ENHANCING THE EFFICACY AND SECURITY OF EMERGING WIRELESS SYSTEMS

by

YU ZHANG

A Dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Professor Wade Trappe

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

Enhancing the Efficacy and Security of Emerging Wireless Systems

By Yu Zhang

Dissertation Director:
Professor Wade Trappe

In this thesis, we intend to promote the efficacy and tackle the vulnerabilities in three emerging wireless systems that have recently became popular examples of emerging wireless systems, namely location-based systems, RFID systems, and cognitive radio systems.

In location-based systems, we first address the problem of being able to reliably deliver content to users based on their locations, in spite of the limited resources available in a wireless network. Secondly, allocating content based on a claimed location implies that location information should be verified in order to support these new location-based services.

We examine strategies whereby access points in an infrastructure delegate the responsibility to serve users to other users who have cached requested content. The resulting strategy, which we call Deputy&Forward, uses knowledge of the user mobility pattern to optimize content delivery for location-based services. Additionally, to verify location information, we adapt the classical challenge-response method for authentication to the task of verifying an entity's location. Our scheme utilizes a collection of transmitters, and adapts the power allocations across these transmitters to verify a user's claimed location. This strategy, which we call power-modulated challenge response, is able to be used with existing wireless networks.

As for the RFID systems, we propose a new RFID network prototype that uses transmit-only low-cost tags and lays the burden of the detection and discrimination of collided tag signals on the RFID readers. We present a statistical estimation approach that can perform the detection in the existence of collisions and the near-far problem. Further, we present a low detection error through simulations under realistic system conditions.

Lastly, in cognitive radio systems, we focus on the security problem whereby a cognitive radio node inserts too many packets into the network, thereby disregarding the link quality, the transaction parties' processing speed and other nodes' transmission attempts. An on-board regulative approach is presented to locally enforce the spectrum etiquettes based on the associated link qualities. We evaluate the performance of our scheme with GNU radios in the ORBIT testbed, and show that better transmission efficiency is achieved.

Acknowledgements

I have been greatly appreciated my advisor, Professor Wade Trappe, who has given me invaluable guidance with tremendous patience and kindness. He has taught me that the Ph.D. study is not only just doing the assigned tasks, but involving large amounts of self-motivated study and examining problems beyond their apparent conclusions. I am grateful to Professor Dipankar Raychaudhuri, Professor Yanyong Zhang and Professor Yingying Chen for their feedback and suggestions related to my research. The informal support and encouragement of many fellow WINLAB students have also been indispensable.

Also, I am very thankful to my mom and my sister who have been a constant source of emotional support.

And, of course, I would like to express my gratitude to my husband Zhibin Wu for having always been supportive. Our journey through graduate school together has given me much happiness.

Dedication

Dedicated to my mother.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivations and Problem Overview

Recent advancements across a variety of communication and computing technologies, ranging from wireless communication to techniques for device localization, are driving new forms of wireless systems for various communication purposes. In particular, three emerging wireless systems have recently increased in their popularity: location-based systems, RFID systems, and cognitive radio systems. As these wireless systems allow for more convenience, they also expose new vulnerabilities and inefficacies in performing the very applications they aim to support.

In a location-based system, users will be able to access content at specific places and at specific times. In other words, location information associated with mobile users can support a broad range of new location-oriented services where users' computing experiences will be enhanced according to where they are located. It will become increasingly important that the location information utilized by these services is trustworthy. Notably, before an entity should be allowed access to location-restricted files, as discussed in [1,2], it is essential that position information be verifiable. Currently, the approach taken to obtain location information regarding a specific device is to localize that device by witnessing physical (e.g. signal strength [3] or time of arrival [4]) or network properties (e.g. hop count [5]) associated with that device's transmissions. Although there have been many localization algorithms proposed [3], it has been noted that the perceived position of a device can be easily affected by a malicious entity altering the calibration of the physical measurement process [6]. Although there are efforts to secure the localization process [6–12] by adding conventional authentication fields [13] or applying robust statistical methods, these methods are still not naturally applied to scenarios where proof must be provided to a third party.

In addition, there are numerous other hurdles that are preventing the vision of a pervasive wireless environment, where content is readily available on-demand, from becoming a reality. One notable challenge facing mobile and pervasive computing applications is the ability to provide desired content in a real-time manner to a user as he or she moves about the wireless environment. Even with accurate location information available, getting content from remote servers and pushing this data close to the user in order to facilitate delay-sensitive applications requires an approach that considers both the user's movement patterns and the resources available to the wireless infrastructure (e.g. access points and other mobile users) in order to deliver content with minimal delay. Further complicating matters is the fact that, in a pervasive computing environment, there will be many users moving and requesting services that involve large media files, and ensuring a fair distribution of content to all users will consequently introduce considerable queuing burden on remote network resources if not carefully managed.

RFID systems are used for asset tracking, which requires the system detects the existence of the RFID tags that are attached to those assets. Considering the example of EZPass, the toll station needs to detect the tag signal that is transmitted by an EZPass tag attached to the car, so that a certain amount of money could be charged to the customer's account. Passive tags that depend on harvesting power from a basestation have performance bounded by the regulatory limits of costly high-power basestations (e.g. on the order of 4Watts). Alternatively, at lower frequencies they work well, but only for short range ($\sim$ 1 cm) sensing, which cannot provide continual tracking. Active tags overcome many of these limits and provide improved range and reliability. Unfortunately, the standard assumption that such tags would consume a large amount of power makes it impossible to continuously monitor them over a period of years.

In cognitive radio networks, the adaptability of the lower layer protocol stack becomes an obvious vulnerability. A node might choose a MAC protocol, which does not consider the actual link quality, receiver's processing speed, or is intended to cause interference on other nodes' transmission. In particular, if this node bypasses a higher layer traffic control mechanism and willingly inserts large amount of packets into a frequency channel, the data transactions in this channel are significantly affected.

Both the detection of a malicious or greedy MAC behavior and the defense mechanisms become more complicated in a distributed cognitive radio network. First, the diversity

of lower layer protocols not only greatly enhances the likelihood of interference, but also makes it hard to define a universal legitimate behavior. Further complicating matter is inherently dynamic communication environments where cognitive radios will operate. As cognitive radios are allowed to adapt their MAC protocols and switch channels, there will be significant bursts of traffic, causing communication conditions to vary drastically over the lifetime of a session.

1.2 Our Contribution

Location information should be verifiable in order to support new computing and information services. In this thesis, we adapt the classical challenge-response method for authentication to the task of verifying an entity's location. Our scheme utilizes a collection of transmitters, and adapts the power allocations across these transmitters to verify a user's claimed location. This strategy, which we call power-modulated challenge response, is able to be used with existing wireless sensor networks. First, we propose a direct method, where some transmitters are selected to send "challenges" that the claimant node should be able to witness based on its claimed location, and for which the claimant node must correctly respond in order to prove its location. Second, we reverse the strategy by presenting an indirect method, where some transmitters send challenges that the claimant node should not be able to witness. Then, we present a signal strength based method, where the node responds with its received signal strength and thereby provides improved location verification. To evaluate our schemes, we examine different adversarial models for the claimant, and characterize the performance of our power-modulated challenge response schemes under these adversarial models. Further, we propose a new localization attack, where a set of nodes collaborate to pretend that there is a node at the claimed location. This collusion attack can do tremendous harm to localization and the performance of the above methods under collusion attack are explained. Finally, we propose the use of rotational directional power-modulated challenge response, where directional antennas are used to defend against collusion attacks.

Further, we introduce a delay-sensitive service that involves transmitting large amounts of location-based data to nodes at multiple locations. Given limited number of access points and an abundance of service requests that result from the nodes moving around, a

typical content delivery service would inevitably experience considerable delay. To solve this problem, we analyze the movement pattern of mobile nodes and approximate it as a semi-Markov process. Based on this model, we explore different components of the underlying service delay and propose that access points should use a multicast strategy to minimize the queuing delay component. Furthermore, we demonstrate the feasibility of employing nodes, which already have their own local copies of location-relevant data, to relay such data to other nodes by employing one or multiple communication channels. Lastly, we examine the resulting algorithms and study their performance relative to baseline content delivery schemes through simulations.

Many asset tracking applications demand long-lived, low-cost, and continuous monitoring of a large number of items, which has posed a significant challenge to today's RFID design. In order to satisfy these requirements, we propose to adopt transmit-only tags without a receiver, which can offer both low power and low cost. In spite of their great potential, such a platform faces many challenges since it cannot sense the channel, causing the collisions among tag transmissions to be high. It is thus crucial to employ effective multi-user detection schemes at the tag reader to extract valid information from collided signals. Traditional detection schemes, such as successive cancelation, cannot be directly applied to the targeted system. Firstly, due to the simplicity of receiver-less transmit-only tags, there is no mechanism for feedback to the tags that is traditionally needed for accurate multi-user detection. More importantly, these schemes impose serious processing and memory requirements on the underlying system, which makes real-time tracking impossible. In this thesis, we address these challenges by performing a statistical estimation of the signal amplitude, and by dividing the received signal sequence (from all the tags) and assigning each block to one reader. We also adopt an online learning mechanism so that readers can anticipate the tags that belong to them. We show that the proposed detection algorithm can achieve low detection error under realistic system conditions.

Finally, we are addressing the malicious or greedy behavior of a cognitive radio node, whereby this adversarial node inserts too many packets into the network, ultimately disregarding the link quality, receiver processing speeds or the transmission attempts of the surrounding nodes. Unlike traditional methods, that first try to differentiate these behaviors, such as whether it is jamming [14], and then choose proper solutions to defend against

it, we propose to use an onboard regulation mechanism, that does not categorize the observed behavior, but 'regulates' itself at the exact moment that the corresponding 'Onboard Regulation Module' (ORM), which is in charge of the regulation, detects such a behavior. The regulation is based on an intelligent link quality estimation approach, with which the ORM takes actions according to the specific link condition. Our method not only satisfies the real-time requirements for cognitive radios by offering a responsive and prompt reaction to the traffic condition in the environment, but also takes this action adaptively in a manner according to the level that this node deviates from a normal behavior. In addition, this is also a self-protective mechanism, which helps a resource-limited embedded cognitive radio to smartly use its battery by avoiding or reducing its own transmission when it is interfered by others. We analyze the requirement of the ORM and its relationship between other modules on a cognitive radio. Furthermore, we demonstrate an efficient method for the ORM to collect the environmental information with which the ORM decides whether and how to regulate its own transmission. Lastly, we examine the resulting algorithms and study the performance relative to traditional CSMA and Aloha MAC protocols on the ORBIT testbed using GNU Radios.

1.3 Organization of the Thesis

The rest of thesis is organized as follows. We first present our power-modulated challenge response location-verification system in Chapter 2, and the adaptive content delivery strategies are described in Chapter 3. Next, we examine the problem of the facilitating an active transmit-only RFID system through receiver-based processing in Chapter 4. Our work on reactive on-board regulation of cognitive radios using link quality estimation is proposed in Chapter 5. Finally, we conclude the thesis and discuss opportunities for future works in Chapter 6.

Chapter 2

Evaluation of Localization Attacks on Power-Modulated Challenge-Response Systems

2.1 Introduction

Many new computing services are being proposed that utilize location information, ranging from position-enhanced routing [15] to services that allow access to resources based on a client's claimed position [16]. It will become increasingly important that the location information utilized by these services is trustworthy. Notably, before an entity should be allowed access to location-restricted files, as discussed in [1, 2], it is essential that position information be verifiable.

Currently, the approach taken to obtain location information regarding a specific device is to localize that device by witnessing physical (e.g. signal strength [3] or time of arrival [4]) or network properties (e.g. hop count [5]) associated with that device's transmissions. Although there have been many localization algorithms proposed [3], it has been noted that the perceived position of a device can be easily affected by a malicious entity altering the calibration of the physical measurement process [6]. Although there are efforts to secure the localization process [6–12] by adding conventional authentication fields [13] or applying robust statistical methods, these methods are still not naturally applied to scenarios where proof must be provided to a third party.

Rather, there is a large class of location-oriented services (particularly those that employ location as the basis for access control), where a more natural paradigm is that the client provides a claimed position to a verifying entity. For such computing services, a good model for securing localization is to verify the truthfulness of the claimed location [17, 18]. The verification of a location claim is thus a problem of *authentication*. Consequently, in this thesis, we adapt the classical challenge-response method from authentication to the task of verifying an entity's location. Our approach utilizes a collection of transmitters with fixed locations, and adapts the power allocations across these transmitters to verify a

user's claimed location. This strategy, which we call power-modulated challenge response (PMCR), can be used with existing wireless and sensor networks. Throughout this thesis, we assume a location-based service model where an entity requesting access to a location-based service must provide a claimed location, and that the claimant can only obtain the desired service by successfully completing a location verification. In other words, we consider all other security aspects of the challenge-response and access control process to be addressed through appropriate network security mechanisms.

Power-modulated challenge response can be used in a *direct* method, where the transmission powers of the transmitters are modulated so that a node at the claimed location should be able to witness the beacons from the transmitters. An *indirect* method, however, would involve the transmission powers of some transmitters being set at levels so that their beacons would not be heard by the node at the claimed location. A third method, which we refer to as the *signal strength* method, involves the node replying with the received powers for signals transmitted by a set of transmitters for verification. In this thesis, we study these methods under different adversarial settings, ranging from a single adversary to colluding adversaries, and from a naive adversary to one who attempts to cleverly subvert the verification process. Notably, we extend our basic methods for the single adversary case to colluding adversaries by employing directional antennas.

The thesis is organized as follows. We begin in Section 2.2 with an overview of location verification, and give a high-level description of PMCR. Since the proposed methods rely heavily upon the theory of RF propagation, in Section 2.3, we provide a quick discussion of the salient issues of propagation modeling. Here, we also outline the notation used in the thesis, and discuss the two different adversarial models that will be referred to in the thesis. In Section 2.4, we present a direct method for PMCR, where some transmitters are selected to send "challenges" that the claimant node should be able to witness based on its claimed location, and for which the claimant node must correctly respond in order to prove its location. We then examine an indirect method for location verification in Section 2.5, and finally present our signal strength based method in Section 2.6. Moreover, collusion attacks and their harmful effects on direct, indirect and signal strength methods are in Section 2.7. We also propose to use both angle of arrival and power modulation to detect collusions. Finally, we place our work in context by discussing related work in Section 2.8, and conclude this chapter in Section 2.9.

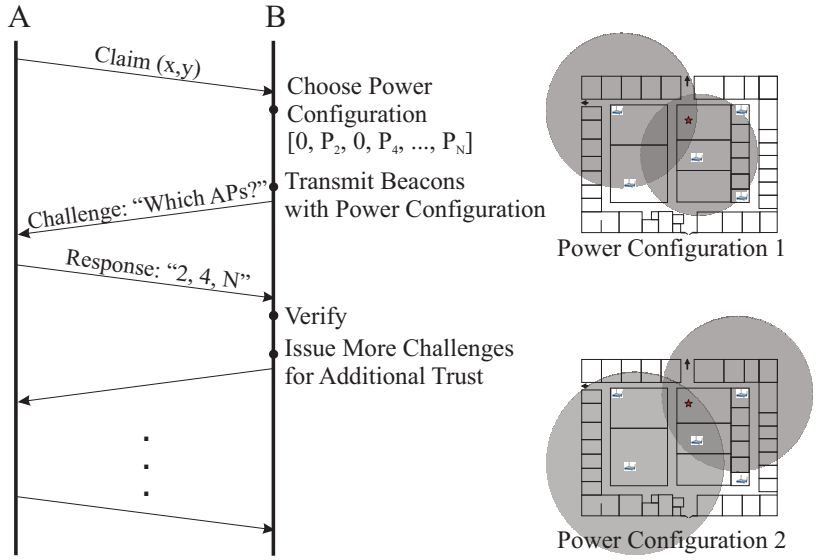


Figure 2.1: Location verification using a generic power modulated challenge-response, where A is the claimant and B is the APs.

2.2 PMCR Overview

Suppose we have an infrastructure of anchor points AP_j of known locations (x_j, y_j), where $j = 1, 2, ..., K$, which are capable of emitting localization beacons, as depicted in Figure 2.1. Suppose that a mobile device contacts the infrastructure, claiming that it is at a location (x, y). To verify the location claim, the infrastructure will issue a *challenge* to the mobile by creating a random test power configuration. This power configuration corresponds to the powers used by the different access points when transmitting their locationing beacons. The power configuration will involve a power of 0 for some access points, meaning that these APs do not transmit, while specific powers are chosen for other APs so as to define a radio region about AP_j such that the node *should* be able to witness the beacon from its claimed position (x, y). The determination of a radio region is done using a propagation model.

The infrastructure now sends the challenge "Which APs do you hear?" to the mobile. The power levels of the APs are temporarily adjusted and location beacons are issued. The mobile then responds with a list of the APs it was able to witness, and the infrastructure checks this response. If a device incorrectly reports that it heard an AP that was not present, then this is clear evidence that the device's truthfulness, and hence its position, is false. However, if a device reports some APs correctly, but fails to report an AP that it

should have heard, then we do not conclude the device's location is false. Rather, it may be that the beacon was simply missed due to poor propagation. We can assert the likelihood that a device misses a beacon using the underlying propagation model, and incorporate this confidence measure into verifying the device's location. In order to enhance the confidence levels of the claimed location, the challenge-response process may be repeated several times with different configurations.

In practice, there are several different variations of PMCR, depending on whether location is verified based on the protocol using APs that the client can or cannot witness according to its claimed location, or whether the client can accurately assess the degree to which it can witness the challenge beacons. In this thesis, we present three different variations of PMCR: Direct, Indirect and Signal Strength PMCR. Later, to defend against collusion attack, we further propose Rotational Directional PMCR.

2.3 System Models

We will first describe the propagation model and the adversary model that we base our work on in this section.

2.3.1 Propagation Model

When a wireless signal propagates in space, it suffers attenuation due to both path loss and shadow fading. A number of statistical propagation models [19–22] have been developed over the years to predict path loss in typical wireless environments. In this work, due to its simplicity and generality, we adopt the combined path loss and shadowing model. For this model, the received power in dB is given by $P_r\left(dBm\right) = P_t\left(dBm\right) + K\left(dB\right) - 10\gamma log_{10}(d/d_0) + \varphi_{dB}$, where P_t is the transmission power, and d is the distance between the transmitter and the receiver. φ_{dB} is a Gaussian distributed random variable with zero mean and variance $\sigma^2_{\varphi_{dB}}$. γ is the path loss exponent, which differs for different environments. K and d_0 are site-specific, constant coefficients. Due to fading, even when the transmission power and the distance are fixed, the actual received power is still a random variable, following a Gaussian distribution $\mathcal{N}\left(f\left(P_t, d\right), \sigma_{\varphi_{dB}}\right)$. The mean received power is $f\left(P_t, d\right) = P_t\left(dBm\right) + K\left(dB\right) - 10\gamma log_{10}(d/d_0)$. For all simulations in this thesis, we use $K = -21.9$, $d_0 = 1$, and $\gamma = 3.71$.

2.3.2 Adversary Model

We consider two adversary [23–25] models: a naive adversary and a smart adversary model. In both models, the adversary claims he is at position (x, y), while his true position is (x', y'). For a naive adversary, we assume he does not know the locations of the access points. Therefore, he cannot estimate the transmission power used by the AP he heard from, and in turn cannot estimate the challenges received at the claimed position (x, y). Hence, he will respond to the challenge like a normal node according to what he hears at (x', y'). For a smart adversary, we assume he knows the locations of the access points, his true location, and the parameters of the propagation model. Thus, he can estimate the transmission power used by the APs he hears. He then estimates the challenges received at position (x, y), and makes a smart response according to his estimates. The difference between the two adversaries will become clear when we apply them to the specific scenarios later.

2.3.3 Assumptions

Our analysis is based on several assumptions. First, we assume all the APs are trustworthy, i.e. the adversary we consider is a node who claims a location different from his true location, and does not compromise the infrastructure. Also, we require that the APs are equipped with radios that can adjust their transmission powers over a continuous range of values.

Second, the WLAN environment is homogeneous. That is, we use the same propagation model for the entire environment. This assumption is not important to our protocol, but it simplifies our analysis and discussion. For the same reason, we also assume that all devices (transmit and receive) are commonly calibrated. For example, this implies that we may assume that all claimants can decode a challenge only if the received signal strength is not less than a fixed, common threshold P_{min}. For all simulations in this thesis, we let $P_{min} = -110$dBm.

Third, we believe a challenge should include a time stamp or nonce, so that if a node does not hear a challenge, it cannot fake a response. Finally, unless otherwise noted, the antennas of the APs are assumed to be omni-directional for computational simplicity. If the antennas [26] are directional, the performance could improve since this would reduce the adversary's chance to hear the challenges when he is away from his claimed position.

2.4 Direct PMCR

In this scheme, we choose k out of K APs to send challenges that can be heard if the node is truly at the claimed location, and keep the other $K-k$ APs silent. We record the indexes of those APs who send challenges in a k-element set H_{c_k}. The transmission power used by each AP depends on the requirement we set on the probability of not being able to verify a normal (trustworthy) claimant node.

For a $j \in H_{c_k}$, the probability that a normal node at its claimed (also true) location (x, y) can hear AP$_j$'s challenge is given by $Pr\left(P_{r_j} \geq P_{min}\right) = Q\left(\frac{P_{min}-f\left(P_{t_j},d_j\right)}{\sigma_{\varphi_{dB}}}\right)$, where P_{t_j} is the transmission power used by AP$_j$, d_j is the node's distance to AP$_j$, P_{r_j} is the received power from AP$_j$ at the node's location, and $Q(\cdot)$ is the standard Gaussian Q-function. The probability that the node can hear all k APs, and thus be verified correctly, is $p_v = \prod_{j\in H_{c_k}}^{k} Q\left(\frac{P_{min}-f\left(P_{t_j},d_j\right)}{\sigma_{\varphi_{dB}}}\right)$. An important design criterion is that the probability of a normal node not being verified be less than threshold a set by the system designer. We call this probability the probability of false negative, and denote it as p_{fn}. Then the criterion is simply $p_{fn} < a$. Since $p_{fn} = 1 - p_v$, this criterion is equivalent to requiring

$$\prod_{j\in H_{c_k}}^{k} Q\left(\frac{P_{min} - f\left(P_{t_j}, d_j\right)}{\sigma_{\varphi_{dB}}}\right) \geq 1 - a. \tag{2.1}$$

For a given set of active APs, there are many valid configurations $\{P_{t_j}\}$ satisfying the above equation. We can choose any of them, or we can simply get one valid configuration by assigning the power P_{t_j} such that

$$Q\left(\frac{P_{min} - f\left(P_{t_j}, d_j\right)}{\sigma_{\varphi_{dB}}}\right) \geq \sqrt[k]{1-a}, \tag{2.2}$$

For every j. Although the above equation only gives the lower bound to each P_{t_j}, we may want to choose the minimum valid power to reduce the chance that the adversary not at the claimed location hears the challenge.

2.4.1 Security analysis

Since all APs should be heard at the claimed location in the direct PMCR scheme, the adversary should respond to all challenges he can hear no matter whether he is a naive or a smart adversary. Therefore, we do not distinguish between them in this section.

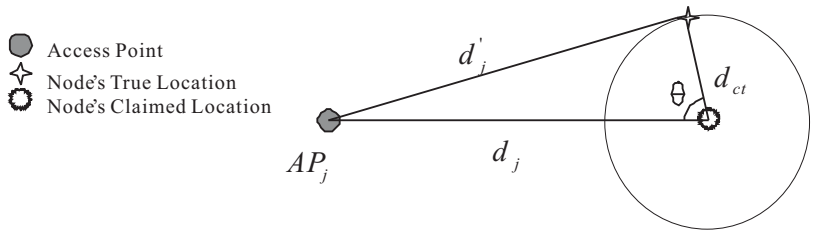


Figure 2.2: The claimed location and the true location.

Suppose the adversary claims his position as (x, y), but is actually at (x', y'), as illustrated in Figure 2.2. Then, the probability that he hears AP_j's challenge is given by $Pr\left(P'_{r_j} \geq P_{min}\right) = Q\left(\frac{P_{min} - f\left(P_{t_j}, d'_j\right)}{\sigma_{\varphi_{dB}}}\right)$, where d'_j is the adversary's actual distance to AP_j. The probability of the adversary hearing all k APs, (and thus is falsely verified), is

$$p_{fp} = \prod_{j \in H_{c_k}}^{k} Q\left(\frac{P_{min} - f\left(P_{t_j}, d'_j\right)}{\sigma_{\varphi_{dB}}}\right). \tag{2.3}$$

It is clear that p_{fp} increases with P_{t_j}, which is why we want to use the minimum valid power for each AP. The probability of false positive is mainly affected by the power configuration and the distance between the claimed location and the true location. We will illustrate their effects with the example network deployment shown in Figure 2.3(a). There are a total of six APs, placed regularly on a grid. The APs are numbered as shown in the figure. The claimed position (x, y) is in the center of the field. Of course, different layouts may affect the appearance of results, but the overall behavior will hold.

Suppose we choose threshold $a = 0.1$, and assign the power of each active AP such that condition (2.2) is satisfied with equality. Then, for every true location (x', y'), there is an associated possibility of false positive, which can be calculated from (2.3). Plotting the equal-p_{fp} contours for different numbers of active APs, we obtain Figure 2.3. The contour labeled 0.9 means that for any adversary located inside this contour claiming a position (x, y), he will be verified with probability greater than 0.9. Because we require a normal node at the claimed position be verified with probability 0.9, the claimed position will lie on the contour. The smaller the area inside the contour, the more reliable the verification is. Clearly, if we have only one active AP, the contours should be circles centered on the AP's location. The closer the adversary is to the AP, the more likely it hears the challenge from this AP. If we increase the number of AP to two, the area with large p_{fp} shrinks

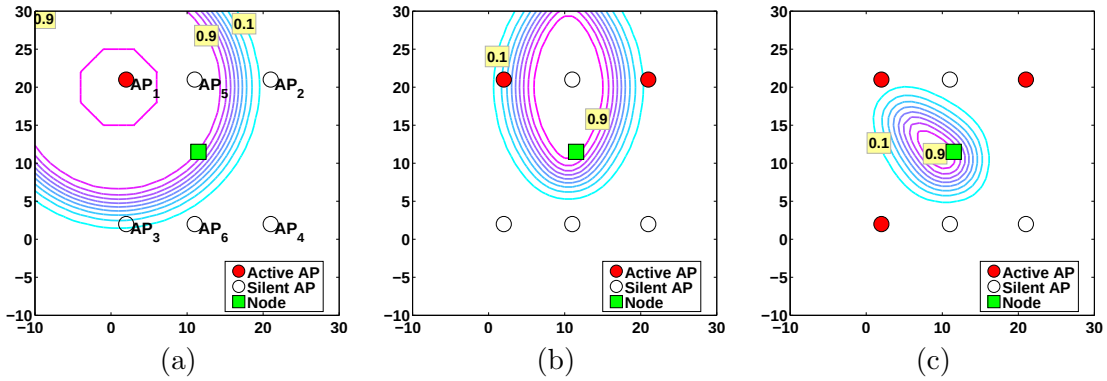


Figure 2.3: Probability of false positive with direct PMCR, (a) $k = 1$ AP, (b) $k = 2$ APs, (c) $k = 3$ APs.

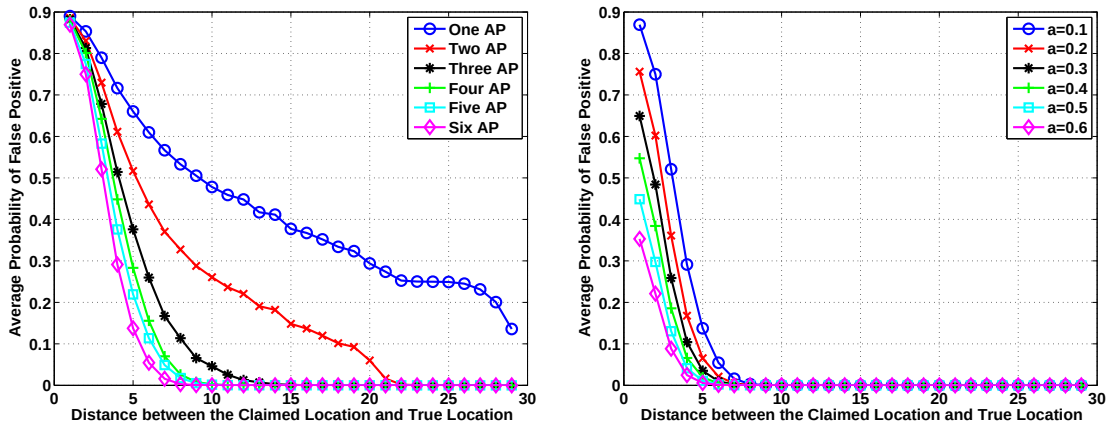
significantly. Only if the adversary is close enough to both APs, can it hear both APs with large probability. The area with large p_{fp} shrinks even further as we increase the number of APs to three. This is as expected since only when the adversary lies in the intersection area of all active APs' communication range, is it able to hear all APs with large probability. The intersection area shrinks quickly as the number of active APs increases.

We also calculated the average probability of false positive $\bar{p}_{fp}(d_{ct})$ when the adversary's actual location is d_{ct} distance away from its claimed location. The curves for different values of k are plotted in Figure 2.4(a). The improvement from $k = 1$ to 2 is very significant, and the improvement slows down as k further increases. Hence, to ensure a low probability of false positive, we need to have a large enough k. On the other hand, we note that it is not true that the larger k, the better. A larger k will result in larger P_{t_j} through condition (2.2), which might help the adversary. Although this side-effect is small compared to the benefit brought by more active APs when k is moderate, it could be detrimental when k is large and the reduction-improvement in intersection area has become negligible.

As in most detection problems, there is a trade-off between the probability of false positive and the probability of false negative. In Figure 2.4(b), we plot the average probability of false positive for different value of a to show this trade-off. As expected, allowing larger p_{fn} reduces the average probability of falsely verifying an adversary.

2.5 Indirect PMCR

In this scheme, we choose k APs to send direct challenges that can be heard and l APs to send indirect challenges that cannot be heard if the claimant is actually at the claimed



(a) (b)

Figure 2.4: (a) Average probability of false positive versus d_{ct} using direct PMCR, for $k = 1, \ldots, 6$ APs, (b) Average probability of false positive decreases with the threshold a for the probability of a false negative, for $k = 6$.

location. The remaining $K - k - l$ APs are kept silent. Here, $K \geq k + l$. We use H_{c_k} to denote the set of indexes of the k APs sending direct challenges, and H_{n_l} to denote the set of indexes of the l APs sending the indirect challenges.

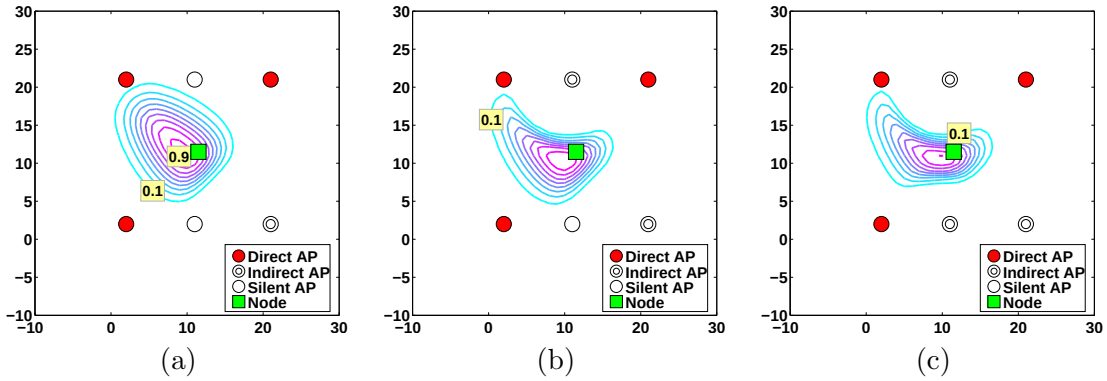
Therefore, the probability that a normal node can hear all k direct APs and cannot hear all of the l indirect APs, and hence can be verified correctly, is $p_v = \prod_{j \in H_{c_k}}^{k} Pr\left(P_{r_j} \geq P_{min}\right) \cdot \prod_{m \in H_{n_l}}^{l} Pr\left(P_{r_m} < P_{min}\right)$. Just as in direct PMCR, we require that the probability of a truthful node not being verified, p_{fn}, to be less than a threshold a. Since $p_{fn} = 1 - p_v$, this criterion is equivalent to requiring $p_v \geq 1 - a$. Again, for a given set of direct and indirect APs, there are many valid power configurations satisfying the above equation. We can choose any of them, or simply obtain a valid configuration by assigning the power such that

$$Q\left(\frac{P_{min} - f\left(P_{t_j}, d_j\right)}{\sigma_{\varphi_{dB}}}\right) \geq \sqrt[k+l]{1-a}, \qquad \forall j \in H_{c_k} \tag{2.4}$$

and

$$Q\left(\frac{P_{min} - f\left(P_{t_m}, d_m\right)}{\sigma_{\varphi_{dB}}}\right) \leq 1 - \sqrt[k+l]{1-a}, \qquad \forall m \in H_{n_l}. \tag{2.5}$$

Although the above equation only gives the lower bound for each P_{t_j} and the upper bound for each P_{t_m}, we may want to choose the powers to reduce the adversary's chance to hear the direct challenge and increase his chance to hear the indirect challenge.



(a) (b) (c)

Figure 2.5: Probability of false positive with indirect PMCR for a naive adversary. (a) $k = 3$, $l = 1$, (b) $k = 3$, $l = 2$, (c) $k = 3$, $l = 3$. Note from now on, we don't label some of the inner contours to give a clearer view.

2.5.1 Security analysis for a naive adversary

We now examine the security issues associated with the indirect PMCR method. The naive adversary will respond to all challenges he can hear, just as a normal node, even though his true location (x', y') is different from his claimed location (x, y). A naive adversary will be falsely verified only if he hears all direct challenges and does not hear all indirect challenges. This probability, the probability of false positive p_{fp}, is given by

$$p_{fp} = \prod_{j \in H_{c_k}}^{k} Pr\left(P'_{r_j} \geq P_{min}\right) \cdot \prod_{m \in H_{n_l}}^{l} Pr\left(P'_{r_m} < P_{min}\right). \tag{2.6}$$

Now we illustrate how introducing indirect APs changes the verification performance. We use the same deployment as earlier with three direct APs. The number of indirect APs varies from one to three. The power used by each active AP is chosen such that (2.4) and (2.5) are satisfied with equality. For every true location (x', y'), there is an associated possibility of false positive, which can be calculated from (2.6). Plotting the equal-p_{fp} contours for different sets of indirect APs l, we obtain Figure 2.5. The change of average probability of false positive with d_{ct} is presented in Figure 2.6(a). The figures show that introducing indirect APs reduces the vulnerable area, and in turn decreases the average probability of false positives.

2.5.2 Security analysis for a smart adversary

For indirect PMCR, the smart adversary responds differently than the naive adversary. When there is an indirect challenge, a smart adversary should not respond to every challenge

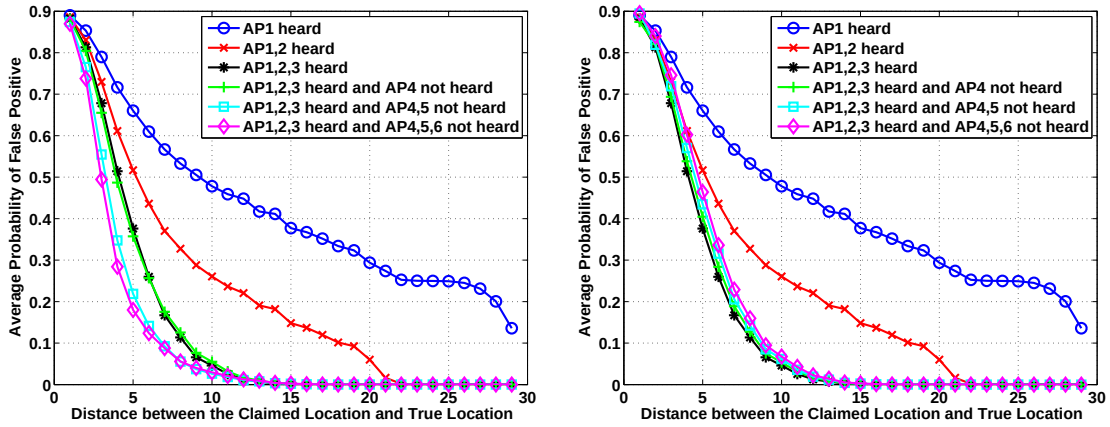


Figure 2.6: Average probability of false positive versus d_{ct} for direct and indirect PMCR. The first three are direct PMCR with $k = 1, 2, 3$ APs, while the last three are Indirect PMCR, with $k = 3$ and $l = 1, 2, 3$. (a) for a naive adversary (b) for a smart adversary.

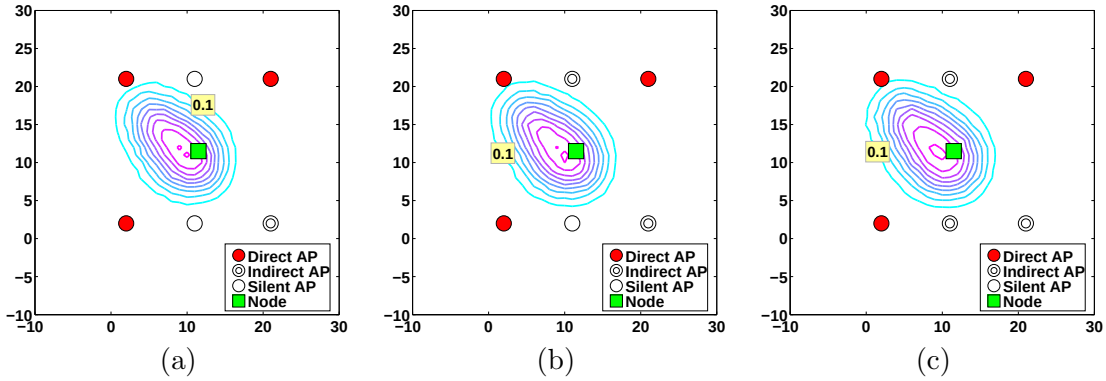


Figure 2.7: Probability of false positive with Indirect PMCR for a smart adversary. (a) $k = 3$, $l = 1$, (b) $k = 3$, $l = 2$, (c) $k = 3$, $l = 3$.

he hears because, if he responds to the false challenge, his location claim will not pass the verification. Instead, since he has knowledge of the APs' locations and the propagation models, he should make a smart judgment on whether he should respond to a particular challenge or not. We now discuss how a smart adversary makes such a judgment and calculate the probability of false positive for a smart adversary.

First, let us assume a smart adversary can hear from AP_j, and the received power is $P'_{r_j} \geq P_{min}$. He needs to make a decision on responding to this challenge or not. To do so, he tries to find the distribution of the received power at the claimed location conditioned on P'_{r_j}. Since he knows the location of AP_j and the underlying propagation model, he can conclude that the transmission power of AP_j follows a Gaussian distribution, that is

$P_{t_j} = P'_{r_j} - K + 10\gamma log_{10}\left(d'_j/d_0\right) + N_1$. Therefore, the received power at the claimed position (x, y) is given by $P_{r_j} = P'_{r_j} + 10\gamma log_{10}\frac{d'_j}{d_j} + N_1 + N_2$. where N_1, N_2 is another Gaussian random variable following $\mathcal{N}(0, \sigma_{\varphi_{dB}})$. If N_1 and N_2 are independent, then $E[N_1 + N_2] = 0$, and $VAR[N_1 + N_2] = VAR[N_1] + VAR[N_2] = 2\sigma^2_{\phi_{dB}}$. Therefore, the distribution of P_{r_j} conditioned on P'_{r_j} is

$$Pr\left(P_{r_j}|P'_{r_j}\right) \sim \mathcal{N}\left(P'_{r_j} + 10\gamma log_{10}\frac{d'_j}{d_j}, \sqrt{2}\sigma_{\varphi_{dB}}\right). \tag{2.7}$$

The smart adversary then estimates the probability that a node at the claimed position can hear the challenge sent by AP_j, and accordingly makes his decision to respond to the challenge or not. In particular, if $Pr\left(P_{r_j} \geq P_{min}|P'_{r_j}\right) \geq \tau$, the adversary decides the challenge is a direct challenge and will respond to it. Otherwise, he will ignore the challenge.

The condition above is equivalent to

$$Q\left(\frac{P_{min} - \left(P'_{r_j} + 10\gamma log_{10}\frac{d'_j}{d_j}\right)}{\sqrt{2}\sigma_{\varphi_{dB}}}\right) \geq \tau.$$

Since $Q(\cdot)$ is a monotonously decreasing function, this is equivalent to

$P_{min} - \left(P'_{r_j} + 10\gamma log_{10}\left(d'_j/d_j\right)\right) \leq \sqrt{2}\sigma_{\varphi_{dB}}Q^{-1}(\tau)$,

which simplifies to

$\delta\left(d'_j, d_j, \tau\right) \triangleq P_{min} - 10\gamma log_{10}\frac{d'_j}{d_j} - \sqrt{2}\sigma_{\varphi_{dB}}Q^{-1}(\tau) \leq P'_{r_j}$.

In summary, a smart adversary will respond to a challenge only if he can hear the challenge ($P'_{r_j} \geq P_{min}$) and $P'_{r_j} \geq \delta\left(d'_j, d_j, \tau\right)$, in other words $P'_{r_j} \geq \max\left(P_{min}, \delta\left(d'_j, d_j, \tau\right)\right)$. If the smart adversary cannot hear a challenge ($P'_{r_j} < P_{min}$), or even if he can hear but $P'_{r_j} < \delta\left(d'_j, d_j, \tau\right)$, he will ignore the challenge. Thus $P'_{r_j} < \max\left(P_{min}, \delta\left(d'_j, d_j, \tau\right)\right)$. The probability for a smart adversary to respond correctly to all direct and indirect challenges, and thus be falsely verified is

$$\begin{aligned} p_{fp} = & \prod_{j \in H_{c_k}}^{k} Pr\left(P'_{r_j} \geq \max\left(P_{min}, \delta\left(d'_j, d_j, \tau\right)\right)\right) \\ & \cdot \prod_{m \in H_{n_l}}^{l} Pr\left(P'_{r_m} < \max\left(P_{min}, \delta\left(d'_m, d_m, \tau\right)\right)\right). \end{aligned} \tag{2.8}$$

If we plot the equal-p_{fp} contours for different set of indirect APs l for $\tau = 0.5$, we obtain Figure 2.7. The change of average probability of false positive versus the distance

between the claimed and true location, d_{ct}, is presented in Figure 2.6(b). The figures show that introducing indirect APs actually increases the probability of false positive when the adversary is smart. The more indirect APs, the larger the detrimental effect. This may seem counter-intuitive at first, but in reality the power used by the direct APs in the indirect PMCR scheme is in fact higher than than is used by the APs in the direct PMCR scheme, when both approaches have the same bound, a, for the probability of false negative. This can be easily seen by comparing (2.4) and (2.2). Thus, when the adversary is smart, the benefit brought by using indirect APs cannot exceed the detrimental effect caused by using a larger transmission power for the direct APs. In fact, for a fixed false negative rate, the indirect method uses more power than the direct method and, as a result, the indirect PMCR system performance actually turns out to be worse than the direct PMCR scheme.

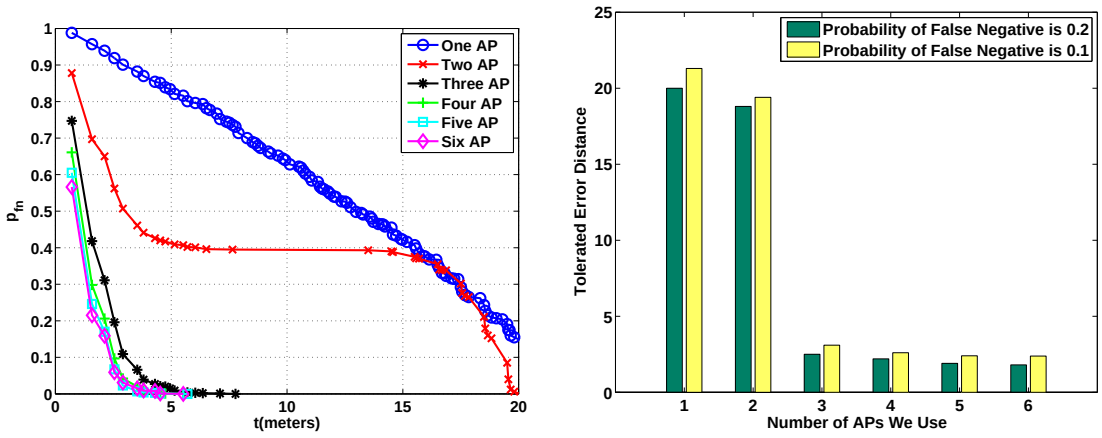
2.6 Signal Strength PMCR

In this scheme, after a node claims its position, k APs are randomly chosen to send challenges with random transmission power $\{P_{t_j}\}$. The power is chosen to be large enough so that a truthful node will hear all the challenges with a high probability. However, unlike the earlier methods, the node is required to report back its received power $\{P_{r_j}\}$ for each AP to the infrastructure. This reported power is then used to verify the node's claimed position.

Due to shadowing, the actual received power P_{r_j} from each AP at location (x, y) follows a Gaussian distribution of $\mathcal{N}\left(f\left(P_{t_j}, d_j\right), \sigma_{\varphi_{dB}}\right)$. Note the location (x, y) plays a role in this probability density function through d_j. With uncorrelated shadowing, the probability density of the set of observed signal powers $\{P_{r_j}\}$ is $Pr\left(\{P_{r_j}\}|\,(x,y)\right) = \prod_{j\in H_{c_k}}^{k} Pr\left(P_{r_j}|\,(x,y)\right)$. To verify a node, the system checks that the response from the claimant includes received powers from all of the active APs. If this is true, the system will make a maximum likelihood estimation of the location of the node based on its reported received power. Denote this location estimate as $(\hat{x}, \hat{y})$, then the maximum likelihood estimate is

$$(\hat{x}, \hat{y}) = \arg\max_{(x,y)} Pr\left(\{P_{r_j}\}|\,(x,y)\right).$$

If the distance between the estimated $(\hat{x}, \hat{y})$ and the claimed (x, y) is smaller than some threshold t, the system will decide that the node is at (x, y). Otherwise, the system rejects the claim. The threshold t determines the probability of not being able to verify a normal



(a) (b)

Figure 2.8: (a) p_{fn} versus threshold t, (b) a clear view of error distance corresponding to p_{fn} of both 0.1 and 0.2 with different number of APs.

node (the probability of false negative p_{fn}), which is given by

$$\begin{aligned} p_{fn} &= Pr\left((\hat{x}-x)^2+(\hat{y}-y)^2 \geq t^2\right) \cdot \prod_{j \in H_{c_k}}^{k} Pr\left(P_{r_j} \geq P_{min}\right) \\ &\quad + \left(1 - \prod_{j \in H_{c_k}}^{k} Pr\left(P_{r_j} \geq P_{min}\right)\right) \\ &\approx Pr\left((\hat{x}-x)^2+(\hat{y}-y)^2 \geq t^2\right). \end{aligned}$$

Here t is chosen to satisfy the system's requirement on p_{fn}. We note the above equation holds if the transmission powers of these k APs are large enough to guarantee that a normal node could hear all the challenges. An analytic relationship between p_{fn} and t is difficult to obtain, and we thus used simulations to explore how p_{fn} changes with t for $k = 1, \cdots, 6$. The results are presented in Figure 2.8(a). Since a large t will result in a large probability of false positive, we would prefer a small t that satisfies the p_{fn} requirement. Figure 2.8(b) shows the value of t for different amounts of active AP's, k, when we require $p_{fn} = 0.1$ and $p_{fn} = 0.2$. Clearly, k should exceed three to ensure that a small t can satisfy the requirement. This is not surprising as three data readings are needed to perform triangulation when estimating a node's location. Beyond $k = 3$, increasing the number of active APs only improves the performance slightly.

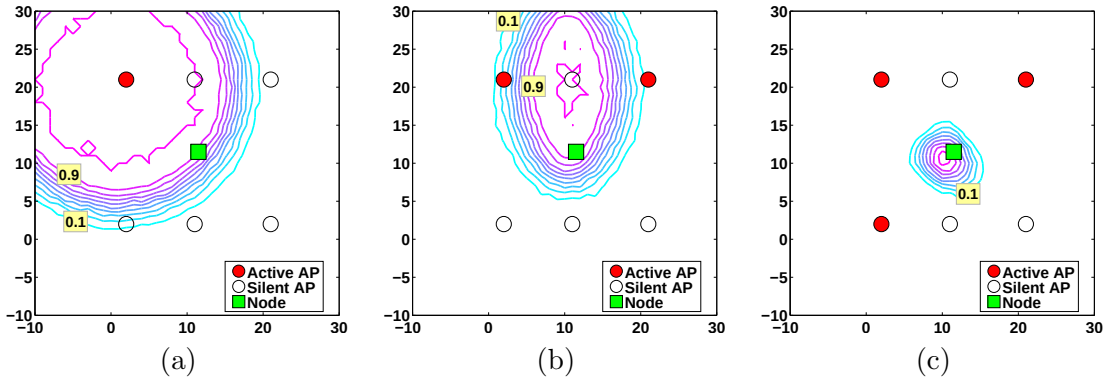


Figure 2.9: Probability of false positive with SS-PMCR for a naive adversary, (a) $k = 1$ AP, (b) $k = 2$ APs, and (c) $k = 3$ APs.

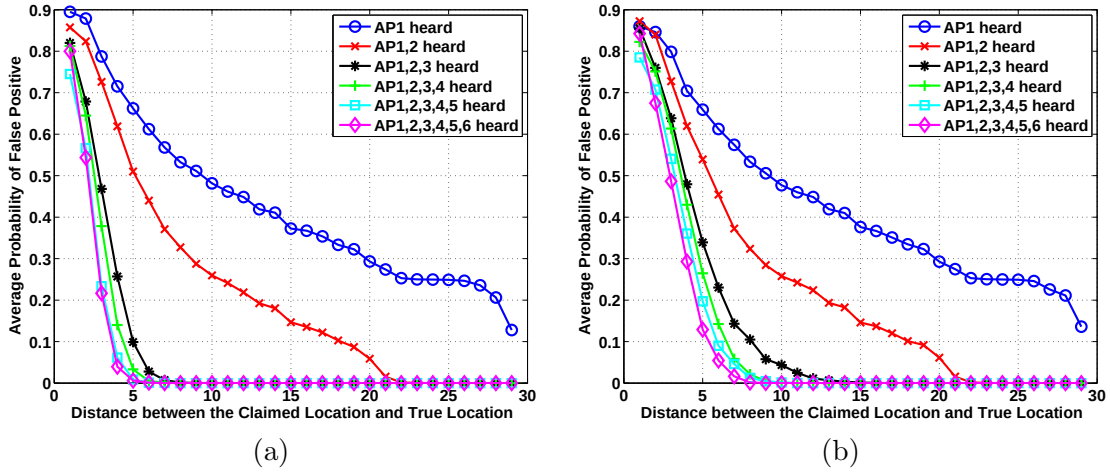


Figure 2.10: Average probability of false positive versus d_{ct} with SS-PMCR, where $k = 1, \cdots, 6$ APs, (a) for a naive adversary, (b) for a smart adversary.

2.6.1 Security analysis for the naive adversary

A naive adversary will simply report its actual received signal strengths $\{P_{r_j}{}'\}$, hoping to pass the verification process. The position estimate obtained at the infrastructure is thus

$$(\hat{x}', \hat{y}') = \arg\min_{(x,y)} \sum_{j \in H_{c_k}}^{k} \left(P'_{r_j} - f\left(P_{t_j}, d_j\right) \right)^2 .$$

The probability of false positive is

$$p_{fp} = Pr\left((\hat{x}' - x)^2 + (\hat{y}' - y)^2 < t^2 \right) \cdot \prod_{j \in H_{c_k}}^{k} Pr\left(P'_{r_j} \geq P_{min} \right). \tag{2.9}$$

We plot the equal-p_{fp} contours for different sets of active APs in Figure 2.9. The change of average probability of false positive versus d_{ct}, the distance between the claimed location

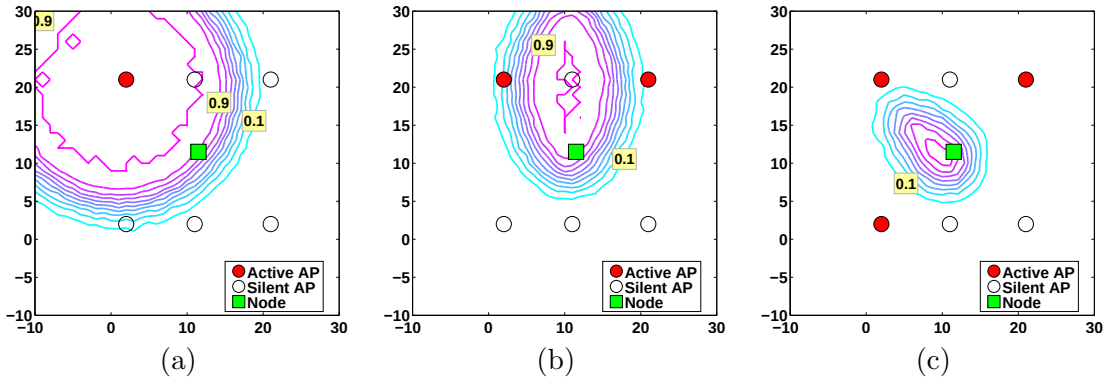


Figure 2.11: Probability of false positive with SS-PMCR for a smart adversary, (a) $k = 1$ AP, (b) $k = 2$ APs, and (c) $k = 3$ APs.

and the true location, is presented in Figure 2.10(a). The figures show that increasing k improves the performance. Notably, when $k \geq 3$ and the adversary is naive, this scheme performs better than the prior schemes.

2.6.2 Security analysis for the smart adversary

A smart adversary uses its knowledge of AP location and propagation model to reports its maximum likelihood estimate of $\hat{P}_{r_j}$, which from (2.7), is $\hat{P}_{r_j} = P'_{r_j} + 10\gamma log_{10}\left(d'_j/d_j\right)$. Then the position estimate obtained at the infrastructure is

$$\left(\hat{x}', \hat{y}'\right) = \arg\min_{(x,y)} \sum_{j \in H_{c_k}}^{k} \left(\hat{P}_{r_j} - f\left(P_{t_j}, d_j\right)\right)^2,$$

and the probability of false positive is still given by (2.9). Plotting the equal-p_{fp} contours for different sets of active APs, we obtain Figure 2.11. The change in the average probability of false positive versus d_{ct} is presented in Figure 2.10(b). Here, we note that a smart adversary has a larger chance of being falsely verified than a naive adversary, and thus the performance ends up being comparable to the direct/indirect schemes.

2.7 Collusion Attack Analysis

The above analyses involved a single adversary, however, a set of adversaries may collude to enhance the effectiveness of an attack. Collusion attacks in localization verification involve multiple adversaries cooperating to cheat the verifiers of the system into believing that there is a node at the claimed location. As long as a node is within an AP's coverage area, it can eavesdrop and share its observation with another colluder. To simplify analysis, we only

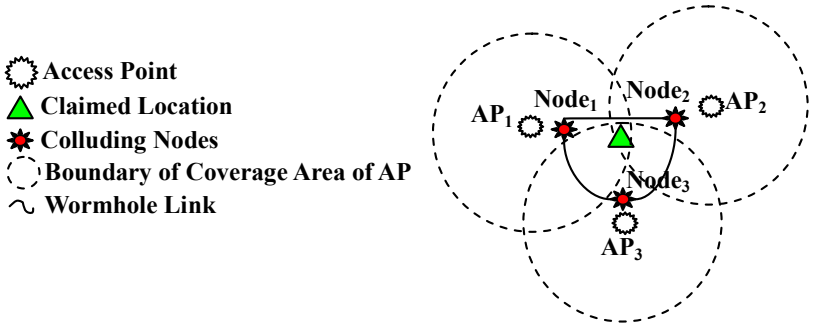
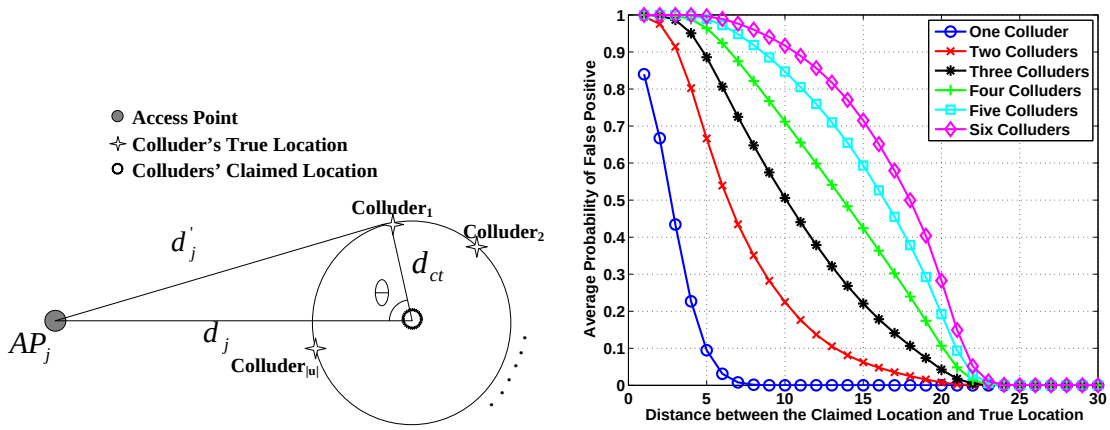


Figure 2.12: Vulnerability of Localization Estimation Parameters to Collusions.

discuss the case where multiple adversaries pretend there is a node at the claimed location and note that more general cases are similar. As shown in Figure 2.12, suppose there are three colluders, Node_1, Node_2, Node_3. None of these nodes can hear all direct challenges from AP_1, AP_2, AP_3. However, because each node can hear a challenge from a distinct AP, in total, the colluding group can hear all the challenges and thus correctly respond to them. In this case, the system is no longer able to make a correct verification.

Suppose there is a set $\mathcal{U}$ of colluding nodes, which cooperate to cheat the system into believing that there is a node at the claimed location, where none of nodes stays. Obviously, if $|\mathcal{U}| = 1$, it reduces to a single adversary case. In this section, we will discuss the collusion behaviors for both naive colluders and smart colluders in the direct, indirect and signal strength PMCR schemes. Here, naive colluders imply each colluder is a naive adversary. None of them knows the locations of access points or estimates the challenges received at the claimed location. If only one colluder receives a certain challenge, he will respond to the challenge like a normal node. If multiple colluders receive a challenge from a certain AP, they still cannot choose whether to reply but have to randomly choose one of them to reply to this challenge. On the other hand, smart colluders imply each colluder is a smart adversary. If only one adversary hears a challenge, he will make an estimate of the transmission power of the AP and make a smart response according to the estimates. If multiple nodes receives a challenge from a certain AP, they smartly choose whether to reply, whom to reply and how to reply.

In this section, the notation follows the same conventions as described in the single adversary case.



(a) (b)

Figure 2.13: (a) The claimed location and the true locations of colluders, (b) Average probability of false positive versus d_{ct} using direct PMCR with $|\mathcal{U}|$ colluders, for $k = 6$ APs and $|\mathcal{U}| = 1, 2, \ldots, 6$.

2.7.1 Direct PMCR

In the direct PMCR method, we will not differentiate between naive colluders and smart colluders since all challenges are direct challenges, and should be answered. Obviously, if the colluders are at different locations, they are more likely to hear all the challenges than a single adversary. As long as one of the colluders hears an AP, that particular colluder is able to respond to this challenge. If all the challenges can be heard by one of the colluders (no matter whether the challenges are heard by the same colluder), these colluders can pass the verification.

Suppose the distance between the colluder u and AP_j is d'_{u_j}, where $u \in \mathcal{U}$, and the received signal strength of colluder u from AP_j is $P'_{r_{u_j}}$. Then the probability of at least one colluders can hear AP_j is

$$1 - \prod_{u \in \mathcal{U}} Pr\left(P'_{r_{u_j}} < P_{min}\right) = 1 - \prod_{u \in \mathcal{U}} Q\left(\frac{f\left(P_{t_j}, d'_{u_j}\right) - P_{min}}{\sigma_{\varphi_{dB}}}\right).$$

The probability of the colluders hearing all k APs, and thus falsely passing the verification, is

$$p_{fp} = \prod_{j \in H_{c_k}}^{k} \left(1 - \prod_{u \in \mathcal{U}} Q\left(\frac{f\left(P_{t_j}, d'_{u_j}\right) - P_{min}}{\sigma_{\varphi_{dB}}}\right)\right). \tag{2.10}$$

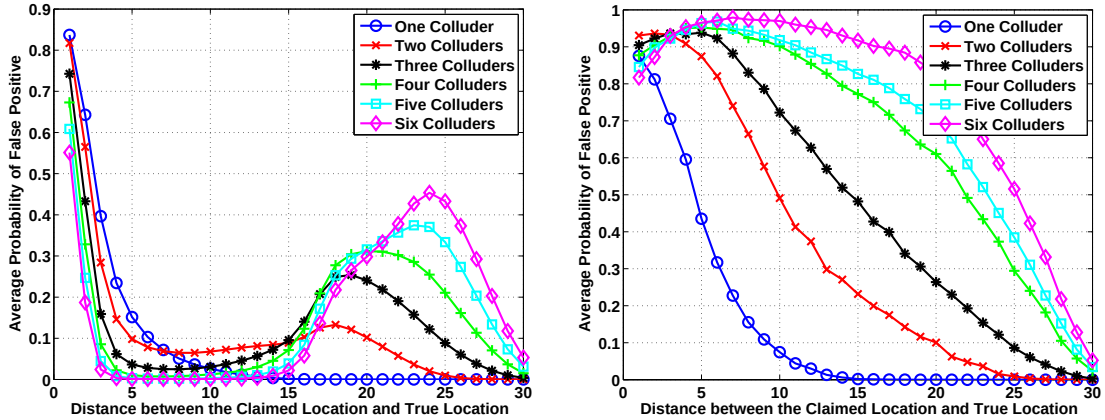


Figure 2.14: Average probability of false positive versus d_{ct} using indirect PMCR with $|\mathcal{U}|$ colluders, for $k = 3$, $l = 3$ and $|\mathcal{U}| = 1, 2, \ldots, 6$. (a) for naive colluders, (b) for smart colluders.

We will still use the example shown in Figure 2.3(a) to show the effect of collusion and set $k = 6$, i.e. all six APs send direct challenges. We will vary the number of colluders $|\mathcal{U}|$ from 1 to 6. In order to give a clear view of the relation between average probability of false positive versus the distance d_{ct} between the claimed location and the colluders, we set each colluder to have the same distance d_{ct} to the claimed location as in Figure 2.13(a), while basically in the different directions. In other words, the $|\mathcal{U}|$ colluders are randomly distributed on the circle that centers on the claimed location with the radius d_{ct}. Certainly, different layouts of colluders may affect the appearance of results, but the overall behavior will hold.

The effects of colluders are illustrated in Figure 2.13(b). If we fix the number of colluders $|\mathcal{U}|$, the average probability of false positive $\bar{p}_{fp}(d_{ct}, |\mathcal{U}|)$ strictly decreases with the increase of d_{ct}. This is intuitively correct, because if we deploy the same number of colluders on a circle, they are more likely to fall out of the coverage area of the APs for a bigger circle. Further, as $|\mathcal{U}| = 1$, the effect is equivalent to the single adversary case shown in Figure 2.4(a). Generally, with the same distance d_{ct} between the colluder and the claimed location, the probability of false positive is higher with more colluders, i.e. they are more likely to hear all challenges and thus falsely pass the verification.

2.7.2 Indirect PMCR

Unlike in the direct method, naive colluders will behave differently from the smart colluders in the indirect PMCR method. When one of the naive colluders hear a challenge, since they are unable to analyze whether it is a direct challenge or not, they must respond to this challenge, hoping this is a direct challenge. While for smart colluders, when one of them receives a challenge, they will analyze whether the node is statistically able to receive this challenge at the claimed location and then decide whether to respond.

Collusion analysis for naive colluders

A set of naive colluders will be falsely verified if, for any direct challenge, at least one of them can hear it (it is unnecessary for one colluder to hear all direct challenges) and none of them can hear any indirect challenges. Thus, the probability of false positive is given by

$$p_{fp} = \prod_{j \in H_{c_k}}^{k} \left(1 - \prod_{u \in \mathcal{U}} Pr\left(P'_{r_{u_j}} < P_{min}\right)\right) \cdot \prod_{m \in H_{n_l}}^{l} \left(\prod_{u \in \mathcal{U}} \left(P'_{r_{um}} < P_{min}\right)\right). \tag{2.11}$$

Now we illustrate the effects of indirect challenges in the face of naive colluders. We still use the same layout of colluders and APs as in the direct PMCR method. However, only AP_1, AP_2 and AP_3 send direct challenges and the other three APs send indirect challenges. The power configurations are the same as in Section 2.5. We plot the curves of the average probability of false positive $\bar{p}_{fp}(d_{ct}, |\mathcal{U}|)$ versus the distance d_{ct} and $|\mathcal{U}|$ in Figure 2.14a. In the near field of the claimed location, i.e. when d_{ct} is small, the average probability of false positive is smaller with more colluders. In other words, the colluders are less able to pass the verification, because the indirect APs are close to the claimed location and the claimed location is near to the indirect APs in our layout, under the same circumstances, more colluders mean that they are more likely to hear some of indirect challenges. Responding to the indirect challenge will reveal that they are not at the claimed location. On the other hand, in the far field of the claimed location, the average probability of false positive is higher with more colluders. In this case, the colluders are far away from the indirect APs, and thus unable to hear the indirect challenges. The performance is therefore similar to the direct PMCR case.

Collusion analysis for smart colluders

For smart colluders, if one or more colluders hear a challenge, they can exchange their signal strength measurements and make a joint decision about whether to respond. Suppose a colluder u hears a challenge from AP j, then the transmission power can be represented as $P_{t_j} = P'_{r_{u_j}} - K + 10\gamma log_{10}\left(d'_{u_j}/d_0\right) + N_1$. We let $P_{t_j} = x_1 + N_1$, where $x_1 = P'_{r_{u_j}} - K + 10\gamma log_{10}\left(d'_{u_j}/d_0\right)$. Suppose there are w colluders can hear this challenge, then we have w equations with $P_{t_j} = x_i + N_i$, where $i = 1, \cdots, w$. Since $P_{t_j} = E\left(x_i\right)$, a good estimation of P_{t_j} is thus $P_{t_j} = \frac{x_1+\ldots+x_w}{w} + \frac{N_1+\ldots+N_w}{w}$. Therefore, the received power at the claimed position (x, y) is given by $P_{r_j} = \frac{x_1+\ldots+x_w}{w} + \frac{N_1+\ldots+N_w}{w} + N_{w+1}$. Since $N_1, \ldots, N_{w+1}$ are independent random variables following $\mathcal{N}\left(0, \sigma_{\varphi_{dB}}\right)$, then $E\left[\frac{N_1+\ldots+N_w}{w} + N_{w+1}\right] = 0$, and $VAR[\frac{N_1+\ldots+N_w}{w} + N_{w+1}] = \frac{w+1}{w}\sigma^2_{\phi_{dB}}$. Then, similar to the single adversary case, we get the condition that the smart colluders respond to a challenge is

$$Q\left(\frac{P_{min} - \frac{\sum_{P'_{r_{u_j}} > P_{min}}\left(P'_{r_{u_j}} + 10\gamma log_{10}\left(d'_{u_j}/d_j\right)\right)}{w}}{\sqrt{\frac{w+1}{w}}\sigma_{\varphi_{dB}}}\right) \geq \tau.$$

The expression of probability of false positive is similar to the single adversary case, thereby we do not reiterate here. The relation between the average probability of false positive versus d_{ct} and the number of colluders $|\mathcal{U}|$ is plotted in Figure 2.14(b) with $\tau = 0.5$. Similar to the naive colluder case, in the very near field of the claimed location, more colluders would be more likely to fail the verification process. This is because in the near field, when the colluders are more likely to hear indirect challenges, they are also more likely to reply to them, although they made adjustments of their strategies already. On the contrary, with bigger d_{ct}, when the colluders are more likely to hear direct challenges rather than indirect ones, the advantages of using this strategy dominate. Therefore, more colluders would be more likely to help them notice direct challenges and also effectively ignore indirect challenges, and thus get a larger average probability of false positive.

2.7.3 Signal Strength PMCR

In the signal strength PMCR method, if only one colluder u_j (no matter whether it is a naive or smart colluder) can hear a challenge from AP_j, colluder u_j has to report a signal strength to AP_j. If more than one naive colluder can hear the challenge, they have to randomly choose one of them, suppose colluder u_j, to report a signal strength, hoping to

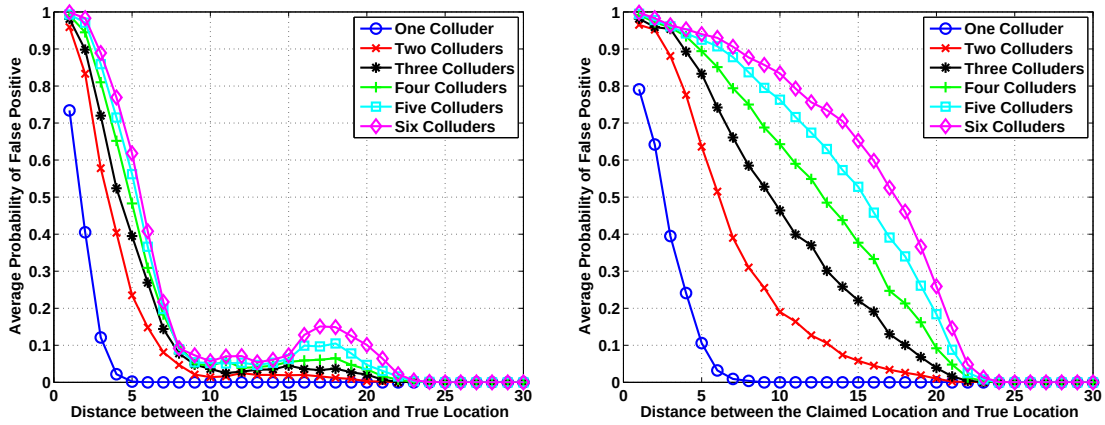


Figure 2.15: Average probability of false positive versus d_{ct} using Signal Strength PMCR with $|\mathcal{U}|$ colluders, for $k = 6$ APs and $|\mathcal{U}| = 1, 2, \ldots, 6$. (a) for naive colluders, (b) for smart colluders.

pass the verification process. In addition, the procedure is also different for naive colluders and smart colluders, in the sense that a smart colluder u_j will respond with altered signal strength values $\hat{P}'_{r_{u_j}} = P'_{r_{u_j}} + 10\gamma log10\left(d'_{u_j}/d_j\right)$, while a naive colluder u_j will report its actual received signal strength $P'_{r_{u_j}}$.

Collusion analysis for naive colluders

If each challenge can be heard by one of colluders, the position estimate obtained at the infrastructure is thus

$$\left(\hat{x}', \hat{y}'\right) = \arg\min_{(x,y)} \sum_{j \in H_{c_k}}^{k} \left(P'_{r_{u_j}} - f\left(P_{t_j}, d_j\right)\right)^2,$$

and the probability of false positive is

$$p_{fp} = Pr\left(\left(\hat{x}' - x\right)^2 + \left(\hat{y}' - y\right)^2 < t^2\right) \cdot \prod_{j \in H_{c_k}}^{k} \left(1 - \prod_{u \in \mathcal{U}} Q\left(\frac{f\left(P_{t_j}, d'_{u_j}\right) - P_{min}}{\sigma_{\varphi_{dB}}}\right)\right). \tag{2.12}$$

We plot the curves of the average probability of false positive $\bar{p}_{fp}\left(d_{ct}, |\mathcal{U}|\right)$ versus the distance d_{ct} between the colluders and the true location and $|\mathcal{U}|$ in Figure 2.15a. With the same distance d_{ct}, the average probability of false positive $\bar{p}_{fp}\left(d_{ct}, |\mathcal{U}|\right)$ is higher with more

colluders. This is obvious because it needs to hear all the challenges and have an estimated location within the distance t to the claimed location to pass the verification, and more colluders are certainly able to hear more challenges and statistically be more likely to pass the verification. Another notable observation is that $\bar{p}_{fp}(d_{ct}, |\mathcal{U}|)$ is non increasing until $d_{ct} \approx 15$ and has a maximum value at $d_{ct} \approx 18$. This is because, at such distances, the colluders who reply to the challenges have similar distances to the claimed location in our layout, and thus can report signal strengths that are easier to pass the verification.

Collusion analysis for smart colluders

Similar to the naive colluders case, if each challenge can be heard by at least one of colluders, the position estimate obtained at the infrastructure is thus

$$(\hat{x}', \hat{y}') = \arg\min_{(x,y)} \sum_{j \in H_{c_k}}^{k} \left(\hat{P}'_{r_{u_j}} - f\left(P_{t_j}, d_j\right) \right)^2.$$

The probability of false positive is thus still given by (2.12).

The curves of the average probability of false positive $\bar{p}_{fp}(d_{ct}, |\mathcal{U}|)$ versus the distance d_{ct} between the colluders and the true location and $|\mathcal{U}|$ in Figure 2.15b. With a distance d_{ct}, the average probability of false positive is higher for more colluders. This is as expected because the challenges are more likely to be heard by more colluders. We also note that the curves have similar shapes as for the direct PMCR method because the smart colluders report altered signal strengths as if from the claimed location.

2.7.4 Rotational Directional PMCR

We now know that the omni directional PMCR is not effective in thwarting colluders, especially smart colluders. This is because for the omni directional direct PMCR method, increasing the number of colluders increases the chances to hear all of the challenges, and additionally, the performance of the omni directional indirect and signal strength PMCR methods are reduced to that of the direct one when the system is attacked by smart colluders.

A natural way to address collusion is to shrink the coverage area of the APs, while ensuring that a node at the claimed location can still hear a direct challenge and not hear an indirect challenge. This would decrease the chance that the colluders could hear all of the direct challenges. In order to achieve this strategy, we may employ directional antennas

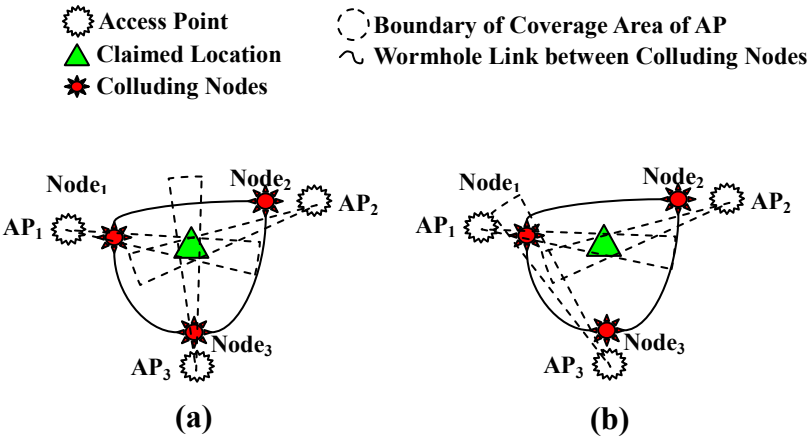


Figure 2.16: Rotational directional PMCR, (a) Node_2 cannot hear the direct challenge from AP_2, (b) Node_1 reveals itself by responding to an indirect challenge from AP_3.

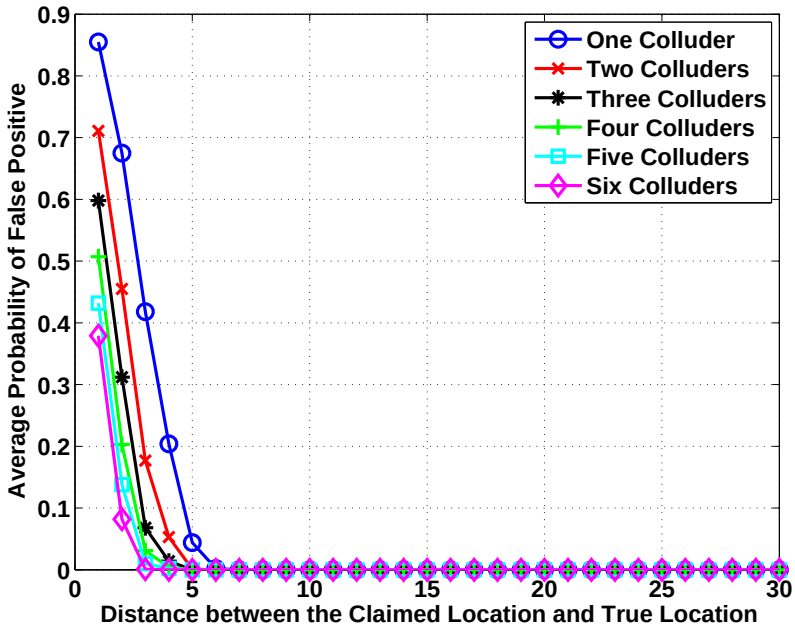


Figure 2.17: Average probability of false positive versus d_{ct} using rotational directional PMCR with $|\mathcal{U}|$ colluders, for $k = 6$ and $|\mathcal{U}| = 1, 2, \ldots, 6$.

to alter the AP coverage region. In particular, an AP with a directional antenna can use power modulation and directivity to send an indirect challenge in the direction of the claimed location, as well as send indirect challenges in other directions. If a node responds to an indirect challenge, we will know that the node is adversarial, regardless of whether it is colluding. The verification process would thereby involve rotating the directions of APs' antennas, and using power modulation to send direct or indirect challenges in many different directions. As before, a node would pass the verification if he can correctly answer

all direct challenges and ignore all indirect challenges.

To explain this scheme, suppose the APs are equipped with directional antennas (either mechanical or electronic). When a node claims to be at a location, the infrastructure selects a valid subset of APs to send direct challenges and another set of decoy APs to send indirect challenges (Note that an AP may send both direct and indirect challenges). The valid APs send direct challenges by setting their transmit powers and directions such that the client is guaranteed to hear these challenges if it is truly in its claimed location. Additionally, the decoy APs send indirect challenges by setting their transmission power or directions so that it is unlikely that the client would witness the challenges if it is at where it claims to be. Let us suppose the layout of APs and colluders is shown as in Figure 2.16(a). If AP_1, AP_2, and AP_3 have omni directional coverages areas, then all the challenges from them could be heard by the colluders, $Node_1$, $Node_2$, and $Node_3$ as in Figure 2.12. Instead, if AP_1, AP_2, and AP_3 send directional challenges to the claimed location, then none of the colluders can hear the direct challenge from AP_2. Another example is shown in Figure 2.16(b). The fact that $Node_1$ responds to the indirect challenge from AP_3 tells the infrastructure it is not at the claimed location.

We plot the curves for average probability of false positive in Figure 2.17. Here we use six APs with antennas having a specular angle of 60°, which send direct challenges in the direction of the claimed location and indirect challenges to all the other direction with equal powers. For all d_{ct} values, the average false positive rate is lower when there are more colluders. This is because, when colluders are at different locations, they are more likely to witness indirect challenges and once a node takes the bait of an indirect challenge, this colluder is detected and fails verification.

2.8 Related Work

Wireless localization has been an active research area, and many algorithms have been proposed in the last decade. Some of the proposed algorithms measure certain physical metrics to estimate the distances. For example, [27, 28] use RSSI(Received Signal Strength Indication), [29] uses TOA(Time of Arrival), [4] uses TDOA(Time Difference of Arrival), and [5] uses AOA(Angle of Arrival). Other algorithms utilize network properties instead of measuring physical metrics. For example, [30] checks who is within communications range of whom to derive the locations of the nodes in the network; [31] counts the number of hops

between a node and the anchor node, which is then converted into distance.

Given the good performance of many existing localization methods, several location-based services have been proposed. [18] proposed an access control server in the building which requires that the prover give responses at no more than a few meters away from the entrance. [2] presented a spatio-temporal access control scheme, where access to an object or service is based on the user's spatio-temporal context. [32] proposed a location aware approach for key management in sensor networks. In [1], different roles of a user are activated based on its position.

The efficiency of location based services depends on the truthfulness of the localization result. However, as pointed out in [6], localization methods are subject to various adversarial attacks. If the location estimate deviates significantly from a node's true location due to an attack, then location based services will not be able to realize their functionality in a reliable way.

Efforts have been made to deal with the vulnerability of localization algorithms. There are roughly two categories of counter measures. The first category is to design attack-tolerant localization methods to combat the attacks. For example, [9] proposed the SERLOC method which estimates location in an untrusted environment by employing a number of sector antennas for anchors. The anchors transmit beacons in sectors, and a grid table is used to record how many sectors a node can hear. The estimated location is then the centroid of the intersections of all sectors a node can hear. The SERLOC method can handle wormhole attacks, sybil attacks and the compromise of network entities. [6] developed robust statistical methods to make localization attack-tolerant. [33] presents two methods to tolerate malicious attacks against beacon-based location discovery in sensor networks. The second category request a node to claim his own location, and then verify whether the claim is trustable or not. [17] uses time difference to approximate an irregular region with several APs' coverage, in order to verify whether a node is in the region of interest or not. However, special devices allowing both RF and ultrasound are needed for this method. [8] proposed to use time of arrival to resist position and distance spoofing attacks. The method measures distance from verifiers to the prover with RF first, then uses geometric method to validate the location claim. However, since RF is used to measure distance, the devices must be able to resolve time difference in high resolution.

Our work differs from prior work on securing localization by focusing on a challenge-response model. The philosophy of position verification was first proposed in the context of distance bounding protocols by Brands and Chaum in [18], and later in [17]. However, unlike these works, which employ timing information, our verification involves signal strength measurements as the underlying physical property. Further, our work takes advantage of multiple verifiers simultaneously in order to provide enhanced verification through the benefits of triangulation. In comparison with other works on secure localization, we do not have the problem of measurement-based attacks at the collection of receiving base stations. Rather, in our motivating problem, the adversary must respond with what it believes is the appropriate response to a challenge (e.g. which access points it witnesses) and thus there is no advantage for an adversary to conduct an attack of the beacon signals being transmitted by the AP. At best, the adversary can only use the information that it witnesses in order to provide a response to the challenge that would make it appear as if it were in another location. Such a threat, though, has been considered in our adversarial models in this chapter.

Further, power modulation is a different approach to localization that can complement existing methods while also lowering the power requirements of existing methods. As an example of this, consider SERLOC [9], where it is assumed that the location beacons must always be heard. This requirement may imply that the power be large in order to guarantee that an honest node can hear the beacon. On the other hand, our approach allows for different power levels to be assigned across the region of interest. For power modulated location verification, we adjust the transmit power levels based upon a probability of false negative at the claimed location, which can allow us to reduce the overall system power requirements. Similarly, for methods in [6,8,17,33], if power modulation is used, adversaries that are far away from the claimed location will not be able to hear some of verifiers and thus can be more easily be detected.

2.9 Conclusion

In this chapter, we have proposed the technique of modulating the transmission powers in a challenge-response mechanism to verify the truthfulness of an entity's claimed location. Three variations were presented: direct PMCR, indirect PMCR and signal strength PMCR. For these three strategies, we evaluated their effectiveness under different adversarial models.

Specifically, we looked at the probability of falsely declaring a claimant is at a valid position for these three schemes versus the distance between the true and claimed position of the claimant. Additionally, although these three methods are effective in verifying the claimed location against a single adversary, we also showed that these methods are susceptible to collusion, and that the probability of false positive increases notably in the presence of naive and smart colluders. To overcome this issue, we have presented a modification to the power modulated approach that employs directional antennas. The resulting directional power modulated challenge-response protocol can reliably detect collusion and achieves improved performance in spite of additional colluders.

Chapter 3

Adaptive Location-oriented Content Delivery in Delay-Sensitive Pervasive Applications

3.1 Introduction

Recent advancements across a variety of communication and computing technologies, ranging from wireless communication to techniques for device localization, are driving new forms of pervasive applications, where users will be able to access content at anyplace and at anytime. In particular, location information associated with mobile users can support a broad range of new location-oriented services where users' computing experiences will be enhanced according to where they are located. To give an example, consider an art gallery where a user approaches a painting and, as the user approaches, media content describing the painting is cached at a nearby wireless transmitter and delivered to a hand-held device the user is carrying. Such applications represent the vision of pervasive computing services.

In spite of improvements in localization technologies over the past decade [6,8,10,15,34], there are numerous other hurdles that are preventing the vision of a pervasive wireless environment with on-demand content. One notable challenge facing mobile and pervasive computing applications is the ability to provide desired content in a real-time manner to a user as he or she moves about the environment. Even with accurate location information available, getting content from remote servers and pushing this data close to the user in order to facilitate delay-sensitive applications requires an approach that considers both the user's movement patterns and the resources available to the wireless infrastructure (e.g. access points and other mobile users) in order to deliver content with minimal delay. Further complicating matters is the fact that in a pervasive computing environment, there will be many users moving and requesting services that involve large media files. Ensuring a fair distribution of content to all users will consequently introduce considerable queuing burden on remote network resources if not carefully managed. This is because for any available

network, the transmission speed or capacity is limited. If there are many requests and the system chooses First Come First Served, late-coming requests have to wait for the network to finish serving those requests that queued before them to get their turn to be served.

One straight-forward approach to ensuring that content is readily available for a user would be to increase the amount of access points that can deliver content, and cache large repositories of content at every access point so that each user request can be readily handled by the nearest access point. However, such a strategy is costly and faces issues associated with interference between access points. This issue can be alleviated somewhat by carefully assigning channels across the environment, so that different access points operate on different channels in order to serve many nodes simultaneously. An alternative strategy might be to increase the transmission power of each access point and, according to Shannon capacity theory, this would allow for an increase in the transmission rate and thereby decrease the transmission time needed for delivering the same amount of content. However, in reality, an AP's transmission power cannot be arbitrarily increased. A large transmission power may cause interference on other nodes, and thus practical systems must have a limitation on the power that may be used. Consequently, a valid transmission rate should be chosen according to the allowed transmission power and the communication environment in this channel.

In this thesis, we seek to deliver content to as many users as possible while maintaining minimal queuing burdens within the network. We consider several different factors that can be exploited in concert. First, we note that many users will want to access the same content, thereby allowing for broadcast dissemination techniques to be employed (as opposed to point-to-point delivery). Further, by using the statistical properties of mobile users and their behavior, serving nodes can adapt a strategy to serve users while minimizing the queuing delays within the network. Lastly, we can allow mobile nodes to *store* location-based multimedia data. In storing, a user's node collects data for its particular location and, even after that data has been used by that user, when new users enter that location zone, they may receive this content directly from that user without having to burden remote servers or access points. Our strategy of storing-and-forwarding exploits the assignment of channels across user nodes and access points in order to avoid interference while promoting parallelism.

This thesis explores these techniques and integrates them together to provide a collection

of delay-sensitive, location-oriented content-delivery strategies. The thesis is organized as follows. First, we will briefly examine related work in Section 3.2. Then, in Section 3.3, we introduce the problem of location-based services, describe our system model, and examine the associated assumptions we have made. Next, after a brief introduction of our delivery methods in Section 3.4, we study the movement behaviors of a mobile node and provide a semi-Markov process model that is suitable for describing user movement patterns in Section 3.5. In Section 3.6, we analyze the different components that introduce delay in a location-based service, and use this analysis to arrive at a service strategy. We then discuss strategies for performing content storing, and use this to arrive at an improved content delivery mechanism in Section 3.7. We evaluate our schemes in Section 3.8, and conclude the thesis in Section 3.9.

3.2 Related Work and Our Contributions

The development of location-oriented applications has involved a large body of literature that has focused on wireless localization, e.g. [4, 5, 27, 30]. Most of these thesiss involved using properties associated with wireless communications, such as received signal strength, to assist in localization. Complementing this work has been a more recent body of literature devoted to assuring that localization results are trustworthy [6, 8, 10, 15, 34].

Once location information is trustworthy, many types of location-oriented services can be built. A notable example is security services that are built upon location information, such as providing access to data based on a user's location at a particular time [2] [1], or key distribution mechanisms that involve location as a parameter [32] [35]. Further, there are a collection of general location-oriented applications [36–38] that do not involve relationships to security.

One challenge that remains in these location-based data services is addressing important QoS (Quality of Service) issues, such as delay. Supporting delay-sensitive multimedia services that involve location as a driving factor, is thus a desirable objective. This objective, however, is made difficult by the limited bandwidth and resources associated with wireless systems.

An area of related research is multicast and broadcast communications. Although there is an extensive body of literature associated with group dissemination methods [39, 40], our

proposed work seeks to explicitly focus on location-oriented services that can have a QoS benefit by using broadcast/multicasting techniques. The thesis [41] is related to our effort in that they propose a multicast strategy to maximize throughput. However, the authors' main objective is to design a policy that decides when a sender should transmit in order to maximize the system throughput subject to maintaining system stability. In our work, we focus on using location information to optimally deliver location-oriented content in order to minimize delay associated with a user's application.

Additionally, some recent work increases scalability and thus improves QoS by using relay nodes [42] [38] to forward data. In [43], nodes forward data to each other when they are close to each other, and thereby can increase the overall data rate in the network. In another related work, [44] uses information of a user's location to predict the information that a user will need in the future and to transfer this information to his mobile device when it is close to an infostation, thereby increasing QoS performance. We extend this idea in this thesis and propose a method, known as the Deputy&Forward method, whereby a detailed movement pattern for user nodes is analyzed to decide whether or not to use another node as a relay for supporting services. One notable impact of our work, which differs from the results in [45], which states that capacity of the network can be increased at the expense of the delay, is that the capacity (i.e. allowable data rates) in fact can be improved without sacrificing delay if we exploit location information!

We note some differences between the assumptions for DTNs [46–48] and our assumptions in this thesis. Also, we note that DTNs are less concerned about optimizing issues of delay (since they are delay-tolerant), whereas we have explicitly focused our analysis on minimizing delay and maximizing throughput.

3.3 System Overview

We begin by describing our underlying system model and some assumptions that we will use throughout this thesis.

3.3.1 System Model

A basic scenario for location-oriented services is depicted in Fig. 3.1. Here, we have four primary components: mobile users (depicted as PDA's moving around an area), access

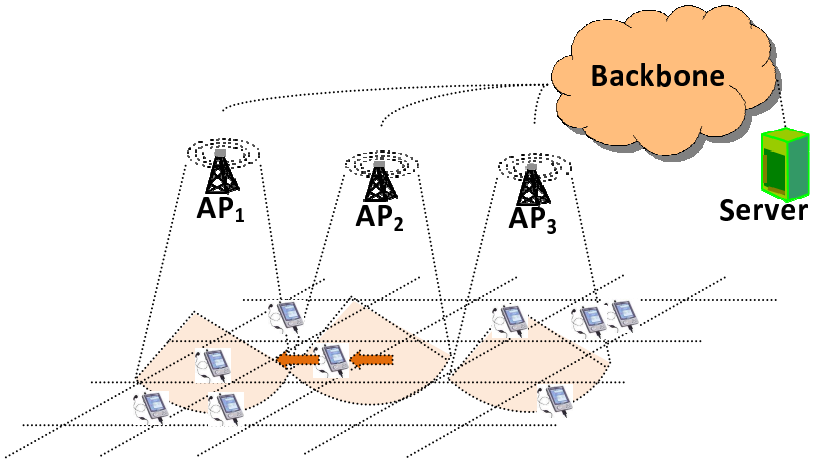


Figure 3.1: System Architecture: a collection of access points (APs) provide service to mobile users according to their locations. Each AP may cover more than one region.

points (depicted as towers transmitting content to users), server (depicted as a data server that contains location-oriented data and decides upon the best strategy for providing services/content to users), and a backbone infrastructure (such as the Internet, which connects the access points and server). A location-oriented application involves nodes (the mobile users) requesting a service or content based on where they are located. Therefore, the administration of a location-oriented service requires the ability to perform localization and being able to track nodes as they move around the environment.

Location-oriented services require defining spatial regions where a user should be located inside to request access to location-oriented content. These regions should be defined to have at least a minimal amount of area in order to cope with inaccuracies in wireless localization. In particular, regardless of whether point-based or area-based localization methods are employed, access to content is given based on whether a user is within a specific *area*. In other words, we visualize an area as a continuous region in two or three dimensional space, instead of a set of discrete points.

The system architecture is compatible with currently available wireless networks, such as 802.11 Wi-Fi or cellular networks. In fact, we envision that our location-based service could ideally work with a multi-AP Wi-Fi system. Here, when a user requests a location-oriented service, the request would be forwarded to the appropriate server by the AP over an IP-backbone. In terms of cohabitation, should a user requests other services (e.g. HTTP), the requests would be appropriately forwarded by following the corresponding network protocol

(e.g. identifiers in the packet header). We note that, when the location-oriented service traffic load is heavy, the proper use of our methods reduces the overall load on APs and consequently would allow more (other) services to coexist.

When the amount of data traffic is light, a traditional transmission scheme that mediates communication (such as CSMA/CA) should be used. On the other hand, when the data traffic becomes heavy, our proposed strategies will make better use of the network capacity and ultimately improve performance.

For quick reference, we list the important notations that we will use in this thesis in Table.3.1.

Expression	*Meaning*
L_i	ith location
p_{ij}	transition probability from L_i to L_j
τ_i	length of time that a node stays at L_i
$\overline{\tau}_i$	mean of τ_i
a_i	multiplicative inverse of $\overline{\tau}_i$
t_{w_i}	length of time that a node waits for service at L_i
t_{p_i}	$t_{p_i} = \tau_i - t_{w_i}$
t_{s_i}	length of service time of a node at L_i from an AP
$\bar{t}_{s_i}$	length of a complete service time from an AP at L_i
M	the total number of locations
$L_{\mathcal{M}}$	the set of indexes of all the locations
K	the number of locations the AP in discussion covers
$L_{\mathcal{K}}$	the set of indexes of locations the AP in discussion covers
n_i	the number of nodes requesting service at L_i
n_{0i}	the number of nodes at L_i
$\bar{t}_{df_i}$	length of a complete service time from a D&F node at L_i
$\mathcal{L}_{W_i}$	the waiting list of L_i
$\mathcal{L}_{D\&F_i}$	the D&F list of L_i
$\mathcal{C}$	the number of channels available
$\texttt{INDEX}(\mathcal{L})$	retrieves the value of `INDEX` in $\mathcal{L}$
q_{op}	the percentage of data been delivered by all nodes
r_{op}	the percentage of data been delivered by one node
//	leads a comment in our algorithms

Table 3.1: Notations

3.3.2 Assumptions

We note that, in any communication system, there are two types of information: content/data and control. In the context of our system, content/data corresponds to multimedia data that is being distributed based on the location of a user. Location-based data are related to a particular location and are the same for all the nodes within the access region

for that content [2]. We note that, for simplicity of discussion in the remainder of the thesis, we will not consider temporal aspects related to content distribution. Rather, we only focus on the spatial aspect of content delivery, and the more general case of content access based on spatial-temporal regions can be handled through straight-forward modifications of the methods presented in this thesis. Here we assume every new node entering a region will request location-based data of this region immediately, and once the request has been fulfilled, it will not request it as long as it possesses the same data. However, due to nodes entering a region at different times, the system would need to deliver certain location-based content multiple times in order to assure that everyone can get this spatially-sensitive information. It is precisely this need for content duplication that we seek to exploit in our work in order to offer better service. On the other hand, we also note that information that is not related to the service, such as control information, typically requires far less bandwidth to deliver than multimedia data and, therefore, we shall focus our discussions on content/data dissemination rather than control message dissemination.

Further, we make an additional assumption about the content being delivered. Although many types of multimedia data can be broken down into segments that may be delivered independently (and possibly out of order with respect to each other), we assume that all content is treated as a single stream file that must be delivered sequentially. In particular, this implies that a receiving node must correctly receive the beginning portion of a content file before being able to receive latter portions of the content. Also, the data stream can be consumed by the receiver before the completion of the entire file transfer. This is useful for users who view a certain portion of the file but decide to stop receiving and move to another location that has more desirable content. The transmission is also performed in a sequential way, in which, an AP needs to finish the data delivery of one node, before it is able to do it for another node. This could avoid unnecessary overhead of switching between different tasks. This may not be necessary when there is a little amount of traffic. However, since we are discussing the heavy traffic load situation, this transmission scheme is valid.

We next make some assumptions regarding the localization algorithms that would be used in support of our location-based service. Wireless authentication and localization have been extensively studied and we thus assume the system can localize and track nodes correctly within the environment. Consequently, we assume that the location information used by our system is accurate and trustworthy. Further, we assume that mobile nodes

have restricted memory resources and consequently, when a node requests content based on its location, it must request this data (as opposed to the rare chance that it might have recorded/stored this content during a previous visit to that location). Finally, in our system, we assume that there are a limited number of access points available for servicing users.

3.4 AP-centric and Deputy&Forward Methods

We will discuss two methods for transmitting location-based data. In the first method, which we call the *AP-centric* method, only APs provide services to the users. In the second method, the *Deputy&Forward* method, mobile nodes *store* location-based multimedia data and then later forward these data to other nodes that enter a specified region. These nodes can thus work as a *deputy* of an AP for content delivery. For the sake of notation, we will call non-AP nodes that can offer services as D&F (Deputy&Forward) nodes for the remainder of this thesis.

The key difference between these two methods is how location-based data are delivered. In AP-centric method, only access points transmit location-relevant data. However, in the Deputy&Forward method, other nodes may assist APs in transmitting data. If a node has already received a location-relevant file and is still in the corresponding location, then that node can become a D&F node and the responsibility for disseminating that data could be assigned to that node. Further, when there are no nodes present that have received a particular data stream, the Deputy&Forward method operates in the same manner as the AP-centric method. For both methods, information that is unrelated to location, such as control information, is managed by the APs. The decision of whether to use a D&F node, or which node is used, is made by the AP in charge of a particular location region. After the AP makes this decision, it will send a control message to the corresponding D&F node to assign the job.

We illustrate these two methods in Fig. 3.2. All the location-oriented data and data delivery assignments originate from the server. At time t, the layout of the network is shown in Fig. 3.2(a), and two regions corresponding to locations where location-relevant content may be accessed are represented by shaded boxes. Node a and node b are in the upper region and node c and node d are in the lower region. Suppose at a later time, $t + \delta t$,

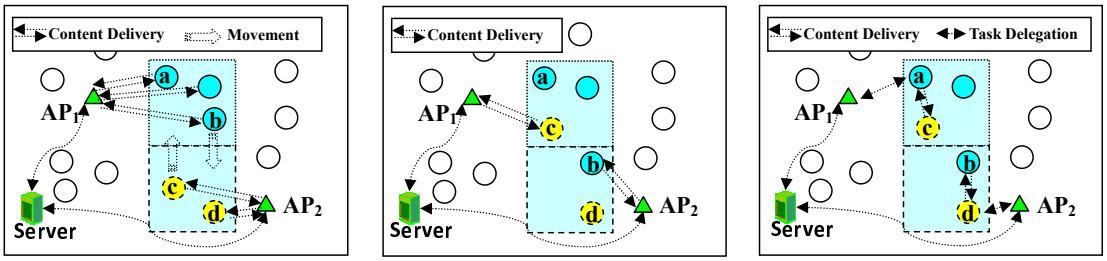


Figure 3.2: AP-Centric and Deputy&Forward methods, (a) Layout at time t, (b) AP-Centric method at time $t + \delta t$, and (c) Deputy&Forward method at time $t + \delta t$.

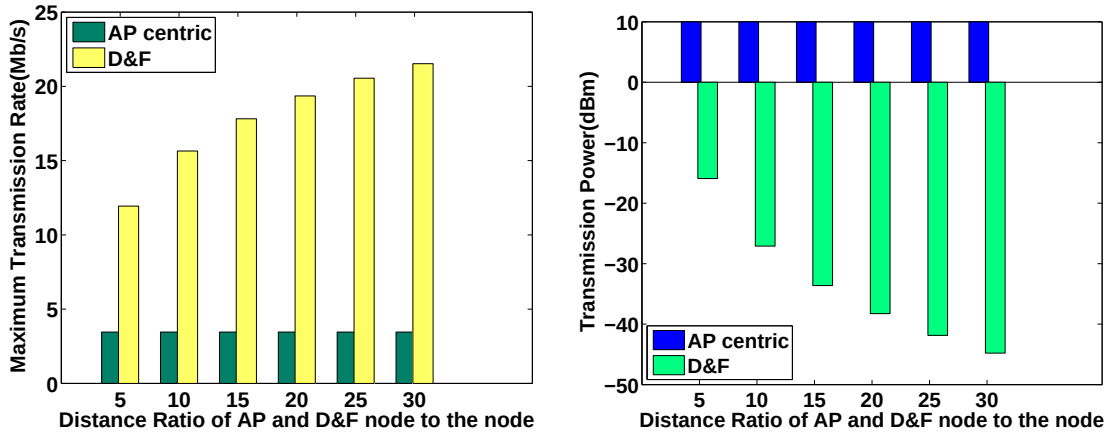
node b moves to lower region and node c moves to the upper region. For the APs shown in Fig. 3.2(b), nodes always choose the AP assigned to the corresponding region, and thus node c will get the location-based data from AP_1 while node b gets the new location-based data from AP_2. In contrast, for the Deputy&Forward method, in Fig. 3.2(c), since node a is still in the upper region, it will forward its stored data to node c, while node d transfers a copy of its location-relevant data to node b. The algorithm will be discussed in detail in Section 3.7. The Deputy&Forward method has some advantages over AP-centric method. Notably, if both the access point and D&F nodes use the same transmission power then, (suppose the noise and interference levels are statistically steady) by Shannon's capacity theorem [49], the D&F node can employ a much larger transmission rate as there will very likely be a D&F node that is closer to the requesting node than the AP. Or, on the other hand, given a certain transmission rate, the required transmission power for a D&F node will be much smaller than that of an access point.

In Fig. 3.2(c), at time $t+\delta t$, AP_1 represents an access point, node c represents a service-requesting node, and node a represents a D&F node. Since node c and node a are within the same region, generally node c will be farther from AP_1 than from another node in the region (since each AP may serve multiple regions). A realistic assumption is implicitly made that it is cost-prohibitive to deploy a large amount of access points and, thus generally, the Euclidean distances satisfy $d_{ac} << d_{1c}$.

Using a fast fading model for a communication channel [50], the channel capacity is given as

$$C = E[log(1 + |h|^2 SNR)], \tag{3.1}$$

where C represents the channel capacity, or the maximum transmission rate, h represents



(a) (b)

Figure 3.3: AP vs. D&F node transmissions(Suppose the noise and interference levels are statistically steady), (a) Maximum transmission rate comparison with the increase of distance ratio $\frac{d_{1c}}{d_{ac}}$, given the transmission power is the same, (b) Required transmission power comparison with the increase of distance ratio $\frac{d_{1c}}{d_{ac}}$, given the capacity is the same and the transmission power of the AP is 10 dBM.

the fading gain, and $|h|^2 SNR$ is the signal-to-noise ratio at the receiver.

In this analysis, we adopt the combined path loss and shadowing model [19]. For this model, the received power in dB is given by

$$P_r\,(dBm) = P_t\,(dBm) + C\,(dB) - 10\gamma log_{10}(d/d_0) + \varphi_{dB},$$

where P_t is the transmission power, and d is the distance between the transmitter and the receiver. φ_{dB} is a Gaussian distributed random variable with zero mean and variance $\sigma^2_{\varphi_{dB}}$. γ is the path loss exponent, which differs for different environments. C and d_0 are site-specific, constant coefficients.

Since we want to characterize an average performance, for simple analysis, we do not need to consider the zero-mean fading effect term φ_{dB}. Further, we suppose the background thermal noise is fixed and the capacity is thus only related to the received power P_r.

In this analysis, we assume that transmitters that are within each other's interference ranges will use different channels to transmit, so that even if they use different transmission powers, they will not interfere with each other. Further, we assume the noise level is statistically steady. With $\gamma = 3.71$, Fig. 3.3(a) shows the maximum transmission rate comparison between these two methods as we increase the ratio of distance between the AP and the node to the distance between a D&F node and that node from 5 to 30, by

keeping the transmission power the same. We can see that the capacity improvement over the AP-centric method increases from roughly a ratio of 3 to over 6. Suppose we need to transmit the same amount of data to each node in both methods, then for a dense environment, the Deputy&Forward method can service more nodes. On the other hand, if D&F nodes have power constraints, by keeping the capacity the same, we can also decrease their transmission power. Fig. 3.3(b) shows the required transmission power comparison under this circumstance, and the required transmission power decreases from 10 dBm to below -40 dBm. With this strategy, the interference between nodes also greatly decreases. In other words, more nodes can transmit data simultaneously and the system throughput increases too. Throughout the rest of the chapter, we shall focus on the first advantage (i.e. the rate improvement of Fig. 3.3(a)), and note that comparable results can be inferred for power reduction.

On the other hand, because Deputy&Forward nodes need to take over the data dissemination tasks, the battery life of these mobile nodes would be affected. We consider Deputy&Forward as a mutual assisting behavior. With which, every node could enjoy more rapid location-based service, while also takes the responsibility of helping other nodes. In order not to exhaust the batteries too fast, we propose three protective measures. The first approach is to balance the battery usage associated with delegation tasks among all the mobile nodes. This avoids the case where we exhaust a particular node's battery significantly faster than any other node. The second approach is to set an upper limit for the overall duration of delegation sessions for each mobile node, and then have the system only delegate a "transmitting task" to a node that has transmitted less than this upper limit. The third approach is to have the node send a low battery warning to the system when that node's battery level is below a preset threshold. In this case, the system would stop delegating tasks to this node. We will discuss our design of the Deputy&Forward approach in Sec. 3.7 in more details.

3.5 System Analysis

We now provide a formal analysis of these two strategies. In order to set the stage for our analysis, we shall consider a typical location-oriented service where customers/users will go to specific locations in order to access specific content. For example, one may consider a

museum, where visitors explore exhibitions at different locations and are provided media content that is relevant to that exhibition.

For such a scenario, when the number of visitors is large, their movement patterns can achieve a steady-state condition where there will be a probability that a nearby user will have the relevant content to share. The process of a node moving around can be modeled as a semi-Markov process [51] whose successive location(state) occupancy is governed by the transition probabilities of a Markov process, but the holding time in any location (state) is described by a continuous positive-valued random variable that depends on the state presently occupied and optionally on the state to which the next transition will be made.

Further, at the transition time when a node moves from one location to another, the process is governed by an embedded Markov chain [52]. Let p_{ij} be the probability that the state of an embedded Markov chain, which entered location L_i on the last transition, will enter location L_j on the next transition. We only consider real transitions, in which a transition happens when a node moves from one location to another. In other words, $p_{ii} = 0$, and that there is an inherent relationship $\sum_{\substack{j=1 \\ j \neq i}}^{M} p_{ij} = 1$.

When the number of visitors is large, it is reasonable to further assume the holding time [51] in any location is independent on other locations. As the occurrences of a user's transitions happen between nonoverlapping intervals that are independent, we can describe the holding time τ_i of a mobile user at location L_i as exponentially distributed $h_i(\tau) = a_i e^{-a_i \tau}, \tau \geq 0$, where the mean of the holding time at L_i is $\overline{\tau}_i = \frac{1}{a_i}$ [53].

Since the infrastructure can track nodes moving, the system can measure the time average of holding time $\overline{\tau}_i$ and the transition probabilities p_{ij}. These parameters can be used to optimize the transmission strategy, which we will discuss in Section 3.6 and Section 3.7.

There are five different cases regarding the relationship between the holding time τ_i and the data delivery time, as shown in Fig. 3.4: a node may leave before it gets any location-based data in Fig. 3.4(a); a node leaves when it gets part of the location-based data in Fig. 3.4(b); a node leaves when it gets all location-based data, but does not forward any data in Fig. 3.4(c); a node leaves when it forwards part of data in Fig. 3.4(d); and a node leaves when it forwards all data in Fig. 3.4(e). We note the last two are only relevant to the Deputy&Forward method. The cases where the node serves more than one service-requesting node are similar to Fig. 3.4(d) and (e).

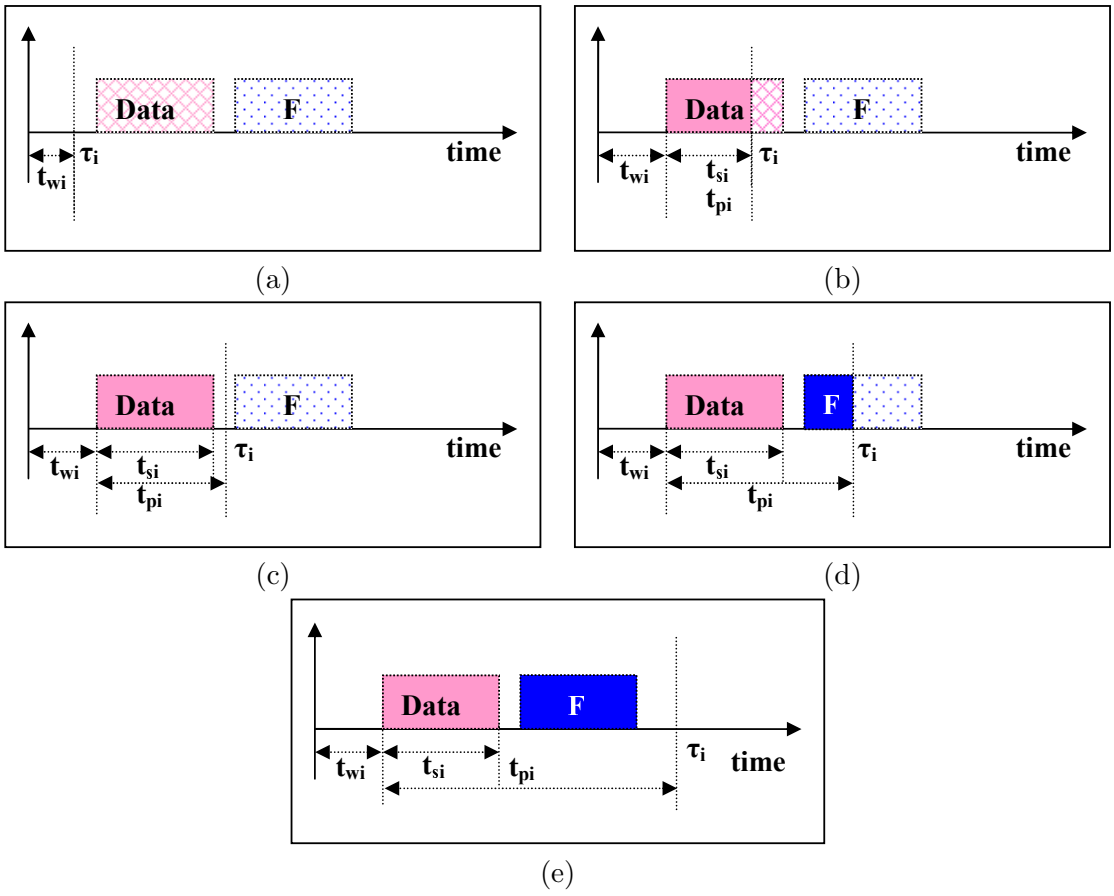


Figure 3.4: Holding time τ_i in Location L_i. (a) Node leaves before getting data, (b) Node leaves when getting part of data, (c) Node leaves when getting all data, (d) Node leaves when forwarding part of data, and (e) Node forwarding all data.

From these five cases, the holding time τ_i is summarized as consisting of two parts, t_{w_i} and t_{p_i}. Here, t_{w_i} is the waiting time, which is the total time from when the node enters location L_i until it begins to get the location-based data delivery or leaves location L_i (whichever is earlier). On the other hand, t_{p_i} is the data possessing time, which is the time since the first bit of data related to this location is possessed by this node till the node leaves this location. The possessing time, as such, at least partly consists of the data transmission time t_{s_i}, as shown in Fig. 3.4. Note that t_{s_i} could be smaller than or equal to the time needed to deliver the entire stream file. The first case is a special case where $t_{p_i} = 0$. The second case is another special case where $t_{s_i} = t_{p_i}$. Taken together, we have

$$\tau_i = t_{w_i} + t_{p_i}. \tag{3.2}$$

When a node(say node c) enters location L_i, if the service-delivering node is busy, node c needs to wait a period of time t_{w_i} for its turn to get served. Since the location-based data is delay sensitive, t_{w_i} should be as small as possible. Immediately after node c begins to receive data, it can process that data. If $t_{p_i} > t_{s_i}$, as in Fig. 3.4(c)(d)(e), within the time $t_{p_i} - t_{s_i}$, if another node moves to location L_i, node c can behave as a D&F node to deliver the stored location-based data to that node. Fig. 3.4(c) is the case where, although $t_{p_i} > t_{s_i}$, this node does not stay long enough to have an opportunity to deliver data to another node. Deputy&Forward method is able to perform well if the holding time for most nodes follows the scenario depicted in Fig. 3.4(e). Beyond reducing the waiting time, it is desirable that one single entire delivery time should be much shorter than the holding time(the system shouldn't plan to deliver too much location-based data). We will show, in Section 3.8, that having the transmission time short compared to the holding time is desirable.

For each location, the amount of location-based data is fixed. Thus, the AP's complete service time is assumed roughly a constant and we use $\widetilde{t}_{s_i}$ to represent the complete service time for the location L_i. Obviously, $t_{s_i} \leq \widetilde{t}_{s_i}$. Suppose the total amount of location-based data at location L_i is S_i, then the amount of data that node c gets at location L_i is $q_i = \frac{t_{s_i}}{\widetilde{t}_{s_i}} S_i$.

From QoS point of view, a good transmission strategy should minimize $\sum t_{w_i}$ and maximize $\sum q_i$. These two criteria are roughly equivalent to each other, since if the waiting time is small, then a node will likely get more data for a fixed holding time than if the waiting time is large. Although these two criteria have different preferences and the resulting performance may differ slightly, their formulations and analysis are very similar. Therefore, we will focus our evaluation only on minimizing $\sum t_{w_i}$.

3.6 AP-centric Method

In the AP-centric method, all of the location-based data is transmitted by the APs, and thus the performance depends on how the APs transmit data. To start, we evaluate some baseline methods and use this to deduce an optimal strategy that minimizes $\sum t_{w_i}$.

3.6.1 Basic Strategies

An AP can serve nodes using either unicast or multicast. If the AP uses unicast, then FCFS (First Come First Served) manner is a good strategy for guaranteeing fairness. In

the case of multicast, if only one node is requesting service, the AP can serve this node immediately, just as in unicast. However, when the AP finishes a communication task and is ready to start the next exchange, there may be multiple nodes requesting service from different locations. The AP needs to choose a subset of nodes to serve, and can either serve nodes/locations in a FCFS manner, or choose to serve the location with maximum number of nodes.

Based on this discussion, we present three basic strategies for AP-centric method to deliver location-based data: FCFS Unicast, FCFS Multicast and Max-Nodes Multicast.

1. **FCFS Unicast**. In this strategy, each AP behaves as a single server with unlimited queue capacity. If only one node requires location-based data, i.e. the queue length is 1, the mobile node can be immediately served. However, if multiple nodes require the service, i.e. the queue length is bigger than 1, nodes must wait for their turn to get service.

2. **FCFS Multicast**. An AP always serves the node with the longest waiting time. When the AP transmits location-based data to the node with the longest waiting time, if there are other nodes in the same location, requesting the service, the AP will multicast the data to these nodes.

3. **Max-Nodes Multicast**. An AP always chooses the maximum number of nodes possible to multicast to. To do this, an AP tracks the number of nodes in each location requesting service, and chooses the location with the most nodes and sends the corresponding location-based data to these nodes. If there are two or more locations with the same number of nodes (which is maximum), the system chooses the location with a node with the longest waiting time to serve.

FCFS multicast is certainly a better strategy than FCFS unicast because the AP serves both the node with longest waiting time and any other node waiting in the same location. Therefore, the overall average waiting time will decrease.

Although FCFS multicast considers fairness, it may not be the best strategy when we consider the whole system's performance. By choosing to serve a location that has the node with the longest waiting time, FCFS multicast does not consider that other locations might have more nodes that collectively have a larger total waiting time. Thus, from the system's

point of view, it is better to choose a location with maximum number of requesting nodes to serve, i.e. the max-nodes multicast method may perform better.

Beyond the three simple strategies, we seek to minimize the total waiting time and thus improve the performance of content delivery. We now look at this more advanced method in detail.

3.6.2 Improved Multicast in the AP-centric Method

In order to improve the multicast performance, we employ a greedy algorithm to allow the AP to always select the location region which, in the short term, will give the best benefit. This selection is done by each AP independent of other APs' choices, so that the decision can be made more efficiently and the system can be more scalable. Here, for discussion, we will focus on a single AP, and will assume the AP covers K locations, collectively represented in the set $L_{\mathcal{K}}$, where $L_{\mathcal{K}} \subseteq L_{\mathcal{M}}$ with $L_{\mathcal{M}}$ is the entire system service area.

The choice of which location to serve is made at each time moment t, when the AP just finishes the previous task and becomes available for a new task (data delivery). Let $X_p(t)$ be the random variable that represents the number of nodes requesting service, and $Y_p(t)$ be the random variable that represents the number of nodes, at moment t and location L_p. Suppose $X_p(t) = n_p$ and $Y_p(t) = n_{0p}$, where n_p and n_{0p} are known by the AP, with $n_p \leq n_{0p}$.

If the optimum choice is location L_i, the overall delay during the transmission time, resulting from choosing location L_i, should be minimum, among all the choices. We note that the nodes requesting the service in location L_i may leave location L_i before they receive the entire data stream, as shown in Fig. 3.4(b). As a result, the AP will abort the delivery procedure in the midway. But here, we assume at least one of these nodes will stay in location L_i till the AP finishes delivering the whole file, and thus the transmission time is $\widetilde{t}_{s_i}$, as in Fig. 3.4(c). The reason is we are considering a system where the average holding time of nodes at a location is much longer than the AP-delivery time, i.e. $\overline{\tau}_i >> \widetilde{t}_{s_i}$. During the transmission time $[t, t + \widetilde{t}_{s_i}]$, the overall delay of nodes in the locations $L_{\mathcal{K}}$ includes following components:

1. **Component I (n_j nodes that are requesting service at location L_j, $j \neq i$ and $j \in L_{\mathcal{K}}$):**

Since the AP chooses to serve nodes at location L_i, then for service-requesting nodes at the other $K-1$ locations in $L_{\mathcal{K}}$, they can behave in two ways.

The first is that the nodes stay in their previous locations even after time $t+\widetilde{t}_{s_i}$ or move to other locations in $L_{\mathcal{K}}$ during $[t, t+\widetilde{t}_{s_i}]$, then the waiting time for these nodes is $\widetilde{t}_{s_i}$. No matter whether there is movement, as long as the nodes are still within the AP's coverage area, they will request location-based data at their destination locations from this AP. Since the channel is engaged during $[t, t+\widetilde{t}_{s_i}]$, those nodes are still waiting for service.

The second way is that the nodes move to locations outside $L_{\mathcal{K}}$ at time $t+\tau \in [t, t+\widetilde{t}_{s_i}]$, where τ is a random variable. After they move out of this AP's coverage, it is no longer this AP's business about whether they get served or not. However, we still need to consider the part of the delay τ before they move out of this AP's coverage.

The probability that the node stays in location L_j during the transmission time is $\int_{\widetilde{t}_{s_i}}^{\infty} a_j e^{-a_j\tau} d\tau$, and the probability that a node moves to other locations in $L_{\mathcal{K}}$ during the transmission time is approximately $\int_0^{\widetilde{t}_{s_i}} a_j e^{-a_j\tau} d\tau \sum_{l\neq j}^{l\in L_{\mathcal{K}}} p_{jl}$. Here, we only consider the case that a node makes only one transition of locations during $[t, t+\widetilde{t}_{s_i}]$, because as we mentioned, we assume the average holding time is much greater than the transmission time, and the probability of multiple transitions is negligible.

Therefore, the waiting time is

$$\sum_{j\neq i}^{j\in L_{\mathcal{K}}} [(\int_{\widetilde{t}_{s_i}}^{\infty} a_j e^{-a_j\tau} d\tau \cdot \widetilde{t}_{s_i} + \int_0^{\widetilde{t}_{s_i}} a_j e^{-a_j\tau} d\tau \cdot \sum_{l\neq j}^{l\in L_{\mathcal{K}}} p_{jl} \cdot \widetilde{t}_{s_i} + \int_0^{\widetilde{t}_{s_i}} \tau a_j e^{-a_j\tau} d\tau \sum_{l\neq j}^{l\notin L_{\mathcal{K}}} p_{jl}) n_j]. \tag{3.3}$$

2. **Component II (Other nodes entering a location in $L_{\mathcal{K}}$ during $[t, t+\widetilde{t}_{s_i}]$):**

   If a node(either previously within this AP's coverage or not) enters a location in $L_{\mathcal{K}}$, it has to wait at least till the end of current delivery to L_i before being considered by this AP.

Since we want to find a strategy that introduces the least delay, we need only to consider the delay that is related to the option of serving location L_i. Thus, we needn't consider

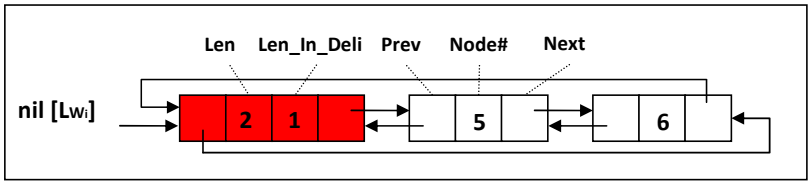


Figure 3.5: An example of waiting list $\mathcal{L}_{W_i}$.

Component II because whatever decisions this AP makes, this part of delay cannot be avoided.

Now we can take a close look at Component I. These n_j nodes at location L_j may move to other locations at time $t+\tau \in [t, t+\widetilde{t}_{s_i}]$ and introduce delay during $[t+\tau, t+\widetilde{t}_{s_i}]$ and this part of the delay is not due to the fact that the AP chooses location L_i. However, these nodes have to wait during $[t, t+\tau]$, due to the system's choice of location L_i. On the other hand, if these n_j nodes remain in location L_i after $t+\widetilde{t}_{s_i}$, they have to wait $\widetilde{t}_{s_i}$. Obviously, this part of the delay is directly related to the system's choice of location L_i. As such, the mathematical expression of the part of delay that is due to the choice of location L_i is

$$D_i = \sum_{j\neq i}^{j\in L_{\mathcal{K}}} [(\int_{\widetilde{t}_{s_i}}^{\infty} a_j e^{-a_j\tau} d\tau \cdot \widetilde{t}_{s_i} + \int_{0}^{\widetilde{t}_{s_i}} \tau a_j e^{-a_j\tau} d\tau) n_j] \tag{3.4}$$

The greedy strategy is to choose a location that will introduce the minimum delay at each instant. Let the index of location of the optimum strategy be *op* and simplify Eqn(3.4), we get

$$\begin{aligned} op &= \arg\min_{i\in L_{\mathcal{K}}} D_i \\ &= \arg\min_{i\in L_{\mathcal{K}}} \sum_{j\neq i}^{j\in L_{\mathcal{K}}} [(\frac{1}{a_j} - \frac{1}{a_j} e^{-a_j \widetilde{t}_{s_i}}) n_j] \end{aligned} \tag{3.5}$$

The AP keeps a list that tracks the requesting nodes for each location it covers, which we call *waiting list* of that location. An example of the waiting list $\mathcal{L}_{W_i}$ is shown in Fig. 3.5. The list can be implemented as a doubly linked list with a sentinel [54] because it is easy to append a node at the end of the list. Unlike traditional dummy sentinels, in $\mathcal{L}_{W_i}$, a sentinel stores the current size of the list $\mathcal{L}_{W_i}$ (denoted as `Len`). It also stores the number of nodes being served by the multicast, (denoted as `Len_In_Deli`, in other words, the number

Algorithm 1 Improved AP-centric Multicast

Require: Topology update, $\mathcal{L}_{W_i}$,a_i,$\widetilde{t}_{s_i}$,$i \in L_{\mathcal{K}}$
Ensure: Improved Serving Strategy
1: **while** Channel is free and Not all $\mathcal{L}_{W_i}$ are empty **do**
2: n_i=Len($\mathcal{L}_{W_i}$), $i \in L_{\mathcal{K}}$
3: *// Choose the optimum location*
4: $op = \arg\min_{i \in L_{\mathcal{K}}} \sum_{j \neq i}^{j \in L_{\mathcal{K}}} [(\frac{1}{a_j} - \frac{1}{a_j} e^{-a_j \widetilde{t}_{s_i}}) n_j]$
5: AP transmits all the data or until all previous requesting nodes leave the location, whichever is earlier
6: Update $\mathcal{L}_{W_{op}}$ as in Alg.2
7: **end while**
8: Update $\mathcal{L}_{W_i}$ as in Alg.2 whenever topology changes

of nodes in delivery) if this location is being served by the AP. If this location currently is not served, the value of `Len_In_Deli` should be 0. Note that the AP will always serve the first one or more nodes in the list because newcomers shall be always appended to the end of the list.

The detailed improved AP-centric multicast is shown in Alg.1. Note in this chapter, to better present the algorithms, we add some comments which are lead by a symbol "//". As long as the channel is free and there is a request, the AP will select a location to serve according to Eqn(3.5). This AP will finish the transmission if at least one of these requesting nodes at the optimum location does not leave this location before the completion of the delivery. Otherwise, the AP will abort this transmission when the last one leaves.

The update algorithm of waiting lists is provided in Alg.2. There are three occasions for updating. The first occasion is when there is a topology change. The AP should append this node into its new location's waiting list. If this node was in the waiting list of its previous location, then it should be removed from this list and the values of its sentinel should be updated as well. The second occasion is when a transmission finishes, then the nodes that get the data should be removed from the waiting list. The third case is when a transmission begins. The value of number of nodes in transmission `Len_In_Deli` should be set as the length of this list `Len`.

3.7 Deputy&Forward Method

We discuss two strategies for the Deputy&Forward method to transmit location-based data: single channel and multiple channel Deputy&Forward.

Algorithm 2 Update of Lists in AP-centric

Require: topology update and waiting lists
Ensure: Update of lists
if Topology update{suppose Node c move from L_i to L_j} **then**
  if node c is in $\mathcal{L}_{W_i}$ **then**
    // Update waiting list of previous location
    Remove node c from $\mathcal{L}_{W_i}$, $\texttt{Len}(\mathcal{L}_{W_i})$=$\texttt{Len}(\mathcal{L}_{W_i})$-1
    if node c is in transmission **then**
      $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_i})$=$\texttt{Len\_In\_Deli}(\mathcal{L}_{W_i})$-1
    end if
  end if
  // Update waiting list of the new location
  Append node c to the end of $\mathcal{L}_{W_j}$, $\texttt{Len}(\mathcal{L}_{W_j})$=$\texttt{Len}(\mathcal{L}_{W_j})$+1
else
  if a transmission is finished **then**
    // Update waiting list of served location
    Remove first $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$ nodes in $\mathcal{L}_{W_{op}}$
    $\texttt{Len}(\mathcal{L}_{W_{op}})$=$\texttt{Len}(\mathcal{L}_{W_{op}})$-$\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$
    $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$=0
  end if
  if a transmission begins **then**
    // Update waiting list of served location
    $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$=$\texttt{Len}(\mathcal{L}_{W_{op}})$
  end if
end if

3.7.1 Single Channel Deputy&Forward method

In this method, under the coverage area of each AP, all the D&F nodes use the same wireless channel that the AP uses, and thus, only one node can transmit at a time. When the last transmission finishes and there are nodes requesting service, the system decides whether an AP or a D&F node will transmit the next data.

Suppose $\widetilde{t}_{df_i}$ is the transmission time needed for a D&F node at location L_i to transmit the entire location-based stream data file. Because of the memoryless property of exponential distribution and $\widetilde{t}_{df_i} < \widetilde{t}_{s_i} << \overline{\tau_i}$, the probability that this D&F node completes one complete transmission is $Pr(t > \widetilde{t}_{df_i}) = e^{-\widetilde{t}_{df_i} a_i} = e^{-\frac{\widetilde{t}_{df_i}}{\overline{\tau_i}}} \rightarrow 1$.

Referring to Eqn(3.5), let $f(t) = \frac{1}{a_j} - \frac{1}{a_j} e^{-a_j \widetilde{t}_{s_i}}$, then $f'(t) = e^{-a_j t} > 0$. Thus $f(t)$ is a monotonically increasing function. Because $0 < \widetilde{t}_{df_i} < \widetilde{t}_{s_i}$, we get $f(\widetilde{t}_{df_i}) < f(\widetilde{t}_{s_i})$. Consequently, $\sum_{j \neq i}^{j \in L_{\mathcal{K}}} [(\frac{1}{a_j} - \frac{1}{a_j} e^{-a_j \widetilde{t}_{df_i}}) n_j] < \sum_{j \neq i}^{j \in L_{\mathcal{K}}} [(\frac{1}{a_j} - \frac{1}{a_j} e^{-a_j \widetilde{t}_{s_i}}) n_j]$. In other words, the D&F node will always introduce less delay than the AP.

Here, we use the same notation and criterion for minimizing the total delay as in last section. Our objective is to minimize the overall delay by choosing to use either a D&F

node or AP. At each moment t, when the channel is free and nodes request location-based data, the system will find whether D&F nodes exist in each location. If, at location L_i, there is a D&F node, then the D&F node should be chosen to transmit the data. Thus, the minimum delay in serving location L_i is $\sum_{j\neq i}^{j\in L_K}[(\frac{1}{a_j}-\frac{1}{a_j}e^{-a_j t_{o_i}})n_j]$, where $t_{o_i}=\widetilde{t}_{s_i}$ if no D&F node exists and $t_{o_i}=\widetilde{t}_{df_i}$ otherwise.

Similar to the AP-centric case, the optimum strategy for selecting a location region to service at a particular instant is

$$op=\arg\min_{i\in L_K}\sum_{j\neq i}^{j\in L_K}[(\frac{1}{a_j}-\frac{1}{a_j}e^{-a_j t_{o_i}})n_j]. \tag{3.6}$$

The AP keeps $2K$ lists of node information, with 2 lists for each of the K locations it covers. For any location L_i, the first list is a sorted one, with a nondescreasing index `usage` (with the unit minute), which describes how much time this node has delivered location-based data to others since it came into the network. Since all the nodes that got the entire data stream file of this location are put into this list, this is a list of nodes that can be used as D&F nodes and we will call it the *D&F list* $\mathcal{L}_{D\&F_i}$. For the other nodes in location L_i that have not received all the location-based data at this moment, they are waiting or in the process of receiving the data. We put these nodes in the second list, which is an unsorted list, called the *Waiting list* $\mathcal{L}_{W_i}$, as was done in the AP-centric method.

We will explain the lists through the example, shown in Fig. 3.6(a). There are four nodes at location L_i. Among these four nodes, node 2 and node 4 have already received location-based data, and thus we keep them in $\mathcal{L}_{D\&F_i}$ as shown in Fig. 3.6(b). Because node 5 and node 6 have not received data, they are in $\mathcal{L}_{W_i}$ as shown in Fig. 3.6(c). In $\mathcal{L}_{D\&F_i}$, the sentinel stores information, which indicates the number of D&F nodes $n_{0i}-n_i$ in location L_i(the index of `Len`) and whether one of them is transmitting location-based data(the index of `Deli_flag`). With this information, an AP can survey how many D&F nodes are present and can update the lists, which we will discuss in Alg.5.

The number of nodes in the waiting list for location L_i is n_i. Given D&F lists and waiting lists in the K locations, as well as the holding time distribution, we can choose the optimum serving location according to Eqn.3.6, as shown in Alg.3.

After the optimum location L_i is chosen, if there are D&F nodes in location L_i, the system can choose any of these D&F nodes to transmit the data. This is because of the

Figure 3.6: D&F list $\mathcal{L}_{D\&F_i}$ and Waiting list $\mathcal{L}_{W_i}$, (a) Node information represented in table, (b) $\mathcal{L}_{D\&F_i}$, (c) $\mathcal{L}_{W_i}$.

Algorithm 3 op=chooseLocation($\mathcal{L}_{D\&F_i}$,$\mathcal{L}_{W_i}$,a_i,$\widetilde{t}_{df_i}$,$\widetilde{t}_{s_i}$,$i \in L_{\mathcal{K}}$)

1: **for** $i \in L_{\mathcal{K}}$ **do**
2: n_j =Len($\mathcal{L}_{W_i}$)
3: **if** Len_In_Deli($\mathcal{L}_{D\&F_i}$)> 0 **then**
4: $t_{o_i} = \widetilde{t}_{df_i}$
5: **else**
6: $t_{o_i} = \widetilde{t}_{s_i}$
7: **end if**
8: $D_i = \arg\min_{i \in L_{\mathcal{K}}} \sum_{j \neq i}^{j \in L_{\mathcal{K}}} [(\frac{1}{a_j} - \frac{1}{a_j} e^{-a_j t_{o_i}}) n_j]$
9: **end for**
10: *// Choose the optimum location*
11: $op = \arg\min_{i \in L_{\mathcal{K}}} D_i$

memoryless property of the node's holding time, (i.e. the time that a D&F node will continue holding in location L_i is independent how much time it has stayed in this location). However, to balance the battery usage, we introduce the index of `usage` for each node in the D&F list.

Note that the `usage` field (see Fig. 3.6) is proposed to make proper use of limited power resources of mobile nodes. Because $\mathcal{L}_{D\&F_i}(,i \in L_{\mathcal{K}})$ is a list sorted according to the ascending value of `usage` and the delegation task is always assigned to the first node in the list, it is guaranteed that the node that has been involved in the least amount of

service is selected and assigned a task. The transmission load is therefore balanced among the mobile nodes. Second, we introduce MAXIMUM ALLOWED USAGE as the upper limit for the amount of tasks that are allowed to be granted to a mobile node. The value of MAXIMUM ALLOWED USAGE is chosen according to the specific settings of a device and an application. For example, if the battery life for transmitting is 300 minutes for a device in an application, MAXIMUM ALLOWED USAGE may be set 120 minutes. If `usage` is above 120 minutes of transmission usage, the node will be no longer considered for having a task delegated to it. Third, before a data dissemination task is assigned to a node, the system can inquire about the battery level for a node. If a "battery low" alarm is reported by that node, the system could assign the current task to another D&F node or just handle the task. In our implementation, when an alarm is reported, the system updates the `usage` field for that node to a value greater than MAXIMUM ALLOWED USAGE.

The details for the single channel Deputy&Forward algorithm is shown in Alg.4. If a D&F node leaves location L_i before it finishes the entire transmission, the system should check whether there is another D&F node in this location. If the result is affirmative, that D&F node should continue the current task from the point where the previous transmission pauses, otherwise, the AP should continue. This is like a relay process and the multiple relays are still counted as parts of one entire transmission.

To keep the real-time service information of each location, an AP should update its D&F lists and waiting lists in three cases as in Alg.5.

1. **Topology update, e.g. node c moves from location L_i to location L_j.** Node c may be in either $\mathcal{L}_{D\&F_i}$ or $\mathcal{L}_{W_i}$. If it is in $\mathcal{L}_{W_i}$, then it can be removed from $\mathcal{L}_{W_i}$ and append it to $\mathcal{L}_{W_j}$. If it is in $\mathcal{L}_{D\&F_i}$, then we need to decide whether it is transmitting data. If the index of `Deli_flag` of $\mathcal{L}_{D\&F_i}$ is 1 and node c is the first node in $\mathcal{L}_{D\&F_i}$, we know it is transmitting data. This is because $\mathcal{L}_{D\&F_i}$ is a sorted list with a nondescreasing value of `usage` and the D&F node with least usage should transmit data. Otherwise, we can just remove it from $\mathcal{L}_{D\&F_i}$ and append it to $\mathcal{L}_{W_j}$. Meanwhile, we need to update the statistical information in the sentinels of both lists.

2. **A delivery of the entire stream data file is finished**. In this case, we will check whether a D&F node or AP offered the delivery. If it is a D&F node, we should update the value if its `usage` and sort $\mathcal{L}_{D\&F_i}$. Additionally, since `Len_In_Deli` nodes

Algorithm 4 Single Channel Deputy&Forward

Require: Topology update, $\mathcal{L}_{D\&F_i}$,$\mathcal{L}_{W_i}$,a_i,$\widetilde{t}_{df_i}$,$\widetilde{t}_{s_i}$,$i \in L_{\mathcal{K}}$
Ensure: Deputy&Forward Serving Strategy

while Channel is free and Not all $\mathcal{L}_{W_i}$ are empty **do**
  // Choose the optimum location
  op=chooseLocation($\mathcal{L}_{D\&F_i}$,$\mathcal{L}_{W_i}$,a_i,$\widetilde{t}_{df_i}$,$\widetilde{t}_{s_i}$) as in Alg.3
  // Set $q_{op} \in [0,1]$ as percentage of data been delivered
  $q_{op} = 0$
  // Assign delivery task to an available D&F node or AP
  while `Len`($\mathcal{L}_{D\&F_{op}}$) > 0 and $q_{op} < 1$ and `Usage`(node df_1) $<$ Maximum Allowed Usage **do**
    AP inquiries the battery level of the first D&F node df_1
    if Battery Low Alarm is received **then**
      // Do not assign task to this low-battery node
      Update $\mathcal{L}_{D\&F_{op}}$ as in Alg.5
    else
      // Assign task to the first D&F node df_1
      The first D&F node df_1 in this list transmits the part of data $[q_{op}, q_{op} + r_{op}]$,{where $r_{op} \in [0, 1 - q_{op}]$ and the resulting value of `usage` is less than Maximum Allowed Usage}
      $q_{op} = q_{op} + r_{op}$
      if $q_{op} < 1$ **then**
        Update $\mathcal{L}_{D\&F_{op}}$ as in Alg.5
      end if
    end if
  end while
  if $q_{op} < 1$ **then**
    // Assign task to AP
    AP transmits the part of data $[q_{op}, 1]$
  end if
  Update $\mathcal{L}_{D\&F_{op}}$ and $\mathcal{L}_{W_{op}}$ as in Alg.5
end while
Update $\mathcal{L}_{D\&F_i}$ or $\mathcal{L}_{W_i}$ as in Alg.5 whenever topology changes

in $\mathcal{L}_{W_i}$ got location-based data already, we remove them from $\mathcal{L}_{W_i}$ and insert them into $\mathcal{L}_{D\&F_i}$. During the insertion, we may change the order of nodes to keep $\mathcal{L}_{D\&F_i}$ sorted.

3. **A transmission begins**. As a transmission begins, if $\mathcal{L}_{D\&F_i}$ is not empty, then the first node in $\mathcal{L}_{D\&F_i}$ is in service, and we should set the index of `Deli_flag` in the sentinel as 1. Since all n_i nodes in $\mathcal{L}_{W_i}$ will be multicasted location-based data, we need to set the index of `Len_In_Deli` in $\mathcal{L}_{W_i}$ as n_i.

4. **Battery Low Alarm is received by AP**. In our design, we intend to keep $\mathcal{L}_{D\&F_i}$ sorted, so that the task could be always assigned to the first D&F node in $\mathcal{L}_{D\&F_i}$. This is accomplished by adding the value of `usage` with the Maximum Allowed Usage

plus one. As a node can only be assigned a task with the value of its `usage` less than Maximum Allowed Usage, this node is exempt from the delegation.

3.7.2 Multiple Channel Deputy&Forward method

In the multiple channel case, we assume that at most $\mathcal{C}$ nodes can transmit simultaneously in $\mathcal{C}$ orthogonal channels at K locations within an AP's coverage area. When one of the $\mathcal{C}$ channels is free and there exist nodes requesting the service, the system will first find the set of locations at which nodes are receiving location-based data. We use $L_{\mathcal{C}}$ to record the set of locations. Considering fairness, the system will choose a location among the location set $L_{\mathcal{K}} - L_{\mathcal{C}}$ if $L_{\mathcal{K}} \supset L_{\mathcal{C}}$. In practice, since the number of available orthogonal channels should be fewer than the number of location zones an AP covers, we will not consider the case of $L_{\mathcal{K}} \subseteq L_{\mathcal{C}}$. The criterion thus is

$$op = \arg \min_{i \notin L_{\mathcal{C}}}^{i \in L_{\mathcal{K}}} \sum_{j \neq i}^{j \in L_{\mathcal{K}}} [(\frac{1}{a_j} - \frac{1}{a_j} e^{-a_j t_{o_i}}) n_j]. \tag{3.7}$$

The description of the multiple channel Deputy&Forward algorithm and list update algorithm are similar to the single channel case. Therefore, we will not discuss them further.

3.8 Evaluation

We now evaluate our proposed algorithms and compare the performance between the AP-centric and the Deputy&Forward method. More specifically, we intend to explore the delay and throughput of our proposed methods, with various settings, such as different number of nodes, ratio of holding time versus transmission time, and ratio of D&F node transmission rate over AP.

Performance is evaluated in terms of three parameters: the normalized waiting time, the normalized number of deliveries and the percentage of data successfully received by the nodes. We now explain these metrics. Suppose the length of simulation time is T and that a node spends a total of τ waiting time to receive the location-based data. This waiting time τ is an accumulated waiting time across all the locations it stays during this simulation period T. If the mean waiting time of all nodes is $\overline{\tau}$, the *normalized waiting time* is $\overline{\tau}/T$.

Suppose that the total number of location changes across all nodes is n, and that there is a total of N data delivery operations(either from an AP or a D&F node), then the *normalized number of transmissions* is $\frac{n}{N}$. We note if a location-based data delivery procedure is started by one node and finished by another, then this is counted as one transmission.

Suppose, for each location, the amount of location-based data is regarded as 1 data "unit". In order to understand service performance, we will introduce a "service ratio" to indicate the service satisfaction at a location. Suppose in a location, there are possibly p nodes entering this location during the simulation. If all of them received the file successfully, the whole throughput of the delivery service is p. However, as some nodes may not be served or only served partially, the overall data throughout will, in reality, be a smaller value, which we denote as q. The service ratio(*average percentage of data* the nodes get) is thus defined as q/p.

A good transmission method would have smaller delay (*normalized waiting time*) and higher throughput (*average percentage of data*). We intend to exploring which method has better performance and the underlying reasons. The metric (*normalized number of transmissions*) would give us some details to help disclosing the reasons.

The simulation layout is shown in Fig. 3.7. There is one AP and the system is divided into 9 locations. In other words, $L_{\mathcal{K}} = L_{\mathcal{M}}$ and $M = 9$. $|\mathcal{S}_c|$ nodes are moving around in this closed environment. The rounded rectangle represents a location zone and is labeled as L_i, $i = 1, \ldots, 9$. The arrow between location L_i and location L_j illustrates that users may move between adjacent zones, and hence reflects a non-zero transition probability p_{ij}. As we mentioned, there are only a few APs in the system, and the APs work independently. Therefore, one AP is chosen in our simulation. We suppose this AP covers nine locations. While in practice, the number of locations an AP cover depends on the application and how the location regions are divided.

We would like to choose reasonable values for the simulation parameters. When a user visits a location (say in a museum), we envision that it is normal for that user to hang around that location for a few minutes. Thus, we have chosen the average holding time to be 200 seconds.

In order to justify our selection of 20 seconds for the transmission time, we conducted two "empirical" studies. The first experiment involved testing the actual transmission speed in a Wi-Fi environment. We set up two laptops with an 802.11g router (an AP) in our lab,

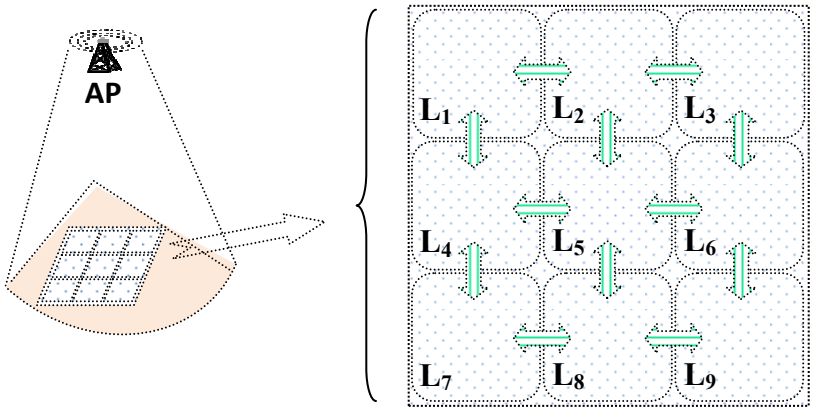


Figure 3.7: Simulation Layout(a single AP services all 9 location regions).

The claimed maximum (physical layer) speed for 802.11g is 54Mb/s. We did 10 experiments where we copied a file from one laptop to another through the AP, and the actual application layer speed varied from 300KB/s to 900KB/s. A reasonable location-based video file may occupy a few megabytes, and thus tens of seconds for transmission time is a good estimate for the time needed to to finish the delivery. In our second experiment, we downloaded an MTV file (a video on demand service) that lasts roughly 3 minutes, and we delivered over Verizon's EVDO Rev A network (which claims a forward link speed up to about 3.1Mb/s) to a local Motorola Krave video phone, the observed delivering time took from 15 to 25 seconds. Therefore, we believe 20 seconds to be a good value for the transmission time for a location-based media file.

As we have mentioned, a D&F node spends less time to deliver the location-based data. On average, it takes less than half of the time the AP does, as shown in Fig.3(a). Therefore, we chose 10 seconds as the delivery time from a D&F node to another node.

Our simulations are initiated by randomly distributing nodes in the 9 locations in Fig. 3.7. In each location, after an exponentially-distributed holding time, the node transitions to another location. At the transition moment, the node chooses its next location according to the following transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}. \tag{3.8}$$

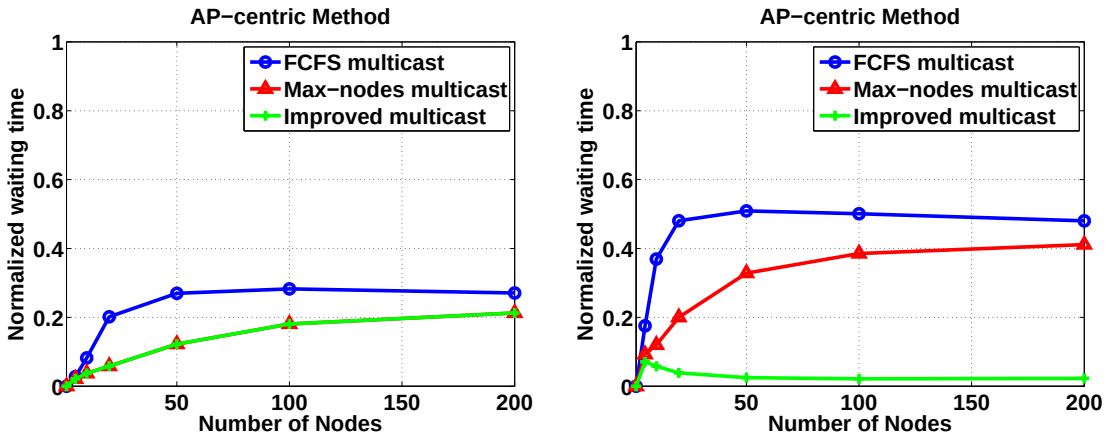
As we have mentioned, each item in the transition matrix p_{ij} represents the probability that the state of an embedded Markov chain, which entered location L_i on the last transition, will enter location L_j on the next transition. Without prior knowledge of actual application, in our simulations, it is reasonable to allow the users to go to any direction with equal probability. For example, if a user is in Location L_1, he can only go to Location L_2 and Location L_4 in the next transition. The transition probabilities p_{12} and p_{14} are thus $\frac{1}{2}$, while $p_{1j} = 0$, for $j = 1, 3, 5, 6, 7, 8, 9$.

After a node moves to another location, it must wait for its turn to be served. The way it is served corresponds to the used strategy. Throughout the simulation, we keep track of the user location, its subjected delay and the corresponding location-oriented data throughput. By comparing the delay and throughput between methods, we would clearly get which method has better peformance over others.

We now discuss several experimental scenarios we investigated. Note that we have done some additional simulations by varying the transmission time from 2 seconds to 20 seconds. (Due to space limitations, we could not include these studies.) In these studies, the performance changes very slightly, but the general results (such as which method has better performance) would hold. We thus believe our results to hold more generally.

3.8.1 Multicast Strategies in AP-centric Method

In this simulation, we will test the performance of three different multicast strategies: FCFS, max-nodes, improved multicast strategies. The performance is evaluated in two scenarios. In the first scenario, we set the transmission time for the location-based data



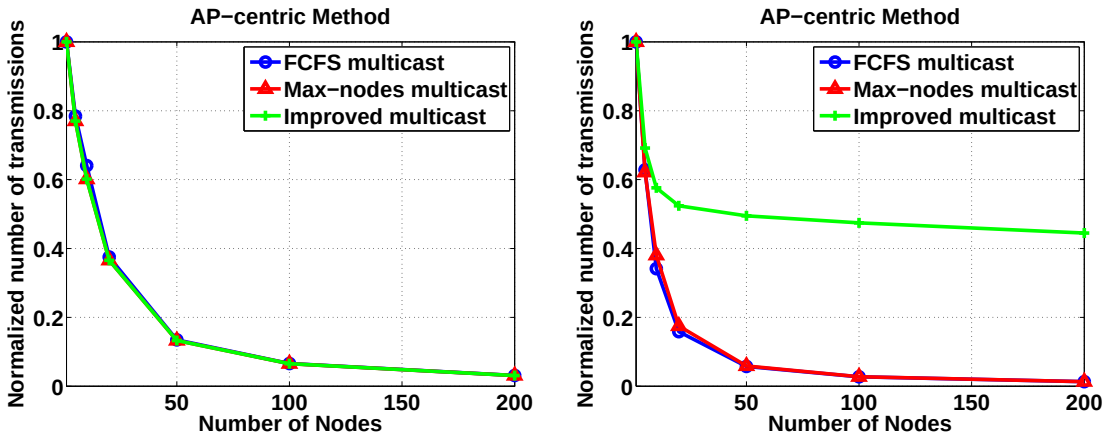
(a) (b)

Figure 3.8: Normalized waiting time of FCFS, Max-nodes and Improved Multicast Strategies in AP-centric Method, (a) with transmission time of all locations 20, (b) with transmission time of locations $[10, 20, 30, 40, 50, 60, 70, 80, 90]$, respectively.

for all locations to 20 seconds. In the second scenario, the transmission time is different as $[10, 20, 30, 40, 50, 60, 70, 80, 90]$ seconds for the nine locations, respectively. Further, the number of nodes $|\mathcal{S}_c|$ is varied from 1 to 200.

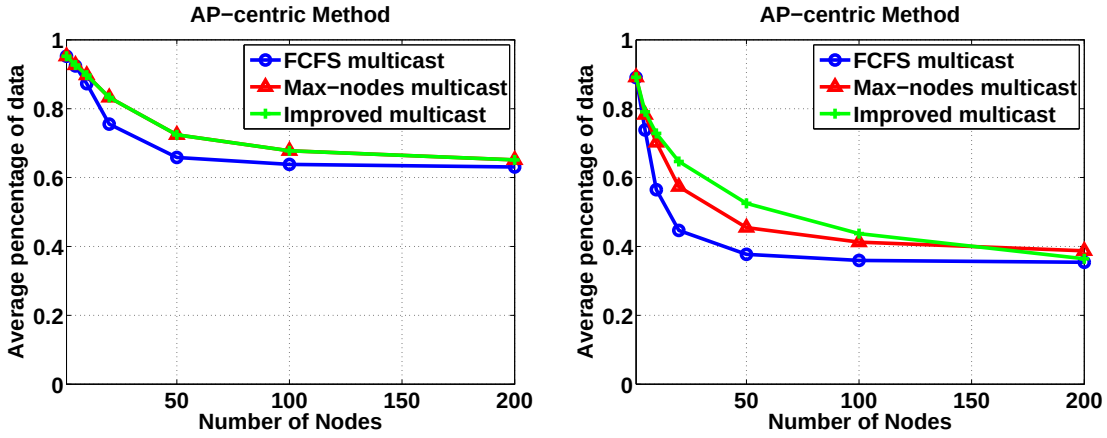
In the first scenario, the delay increases with the number of nodes for all three strategies in Fig. 3.8(a). When the number of nodes is 100, the delay achieves its steady state. Max-nodes and improved multicast algorithms have the same performance, and are better than FCFS multicast. However, when the transmission time is not the same(in the second scenario), the improved multicast strategy has least delay and shows much better performance than the other two methods, as shown in Fig. 3.8(b). It shows that, in the case where each location's data needs the same transmission time and the users have the same holding time distribution, the max-nodes multicast method is an improved multicast strategy. However, it is no longer the case when the above parameters are different.

Fig. 3.9 shows the normalized number of transmissions for these three multicast strategies. The number of transmissions is almost the same with the same transmission time for all locations. When the transmission time is not the same, the improved strategy is able to transmit more. This is because the improved multicast method always chooses the location that that introduces least delay to serve. Although the system considers both the transmission time and number of nodes in that location, compared to FCFS and max-nodes multicast, it tends to choose a location with less transmission time and thus can support



(a) (b)

Figure 3.9: Normalized Number of Transmissions of FCFS, Max-nodes and Improved Multicast Strategies in AP-centric Method, (a) with transmission time of all locations 20, (b) with transmission time of locations $[10, 20, 30, 40, 50, 60, 70, 80, 90]$, respectively.



(a) (b)

Figure 3.10: Average Percentage of Location-based Data Each Node Gets in Each Transmission of FCFS, Max-nodes and Improved Multicast Strategies in AP-centric Method, (a) with transmission time of all locations 20, (b) with transmission time of locations $[10, 20, 30, 40, 50, 60, 70, 80, 90]$, respectively.

more transmissions.

As we mentioned, a good strategy should guarantee that users get as much location-based data as possible. If the transmission time is the same, the max-nodes multicast and improved multicast methods have the same performance(again better than FCFS multicast) as shown in Fig. 3.10(a). When the data delivery time varies at different locations, the improved multicast method has the best performance. This shows that when the waiting

delay is smaller, nodes statistically get more data, given that the holding time is fixed. However, when the number of nodes is large, the max-nodes multicast method has best performance. In this case, the advantage of serving more nodes in max-nodes multicast is dominant over the advantages of minimizing the delay in the improved multicast method.

From these simulations, FCFS multicast has the worst performance. In the case that all the parameters are the same, max-nodes multicast is actually the improved multicast strategy. However, in cases where the parameters are not the same, the improved multicast method can serve with minimized delay and thus normally has the best performance.

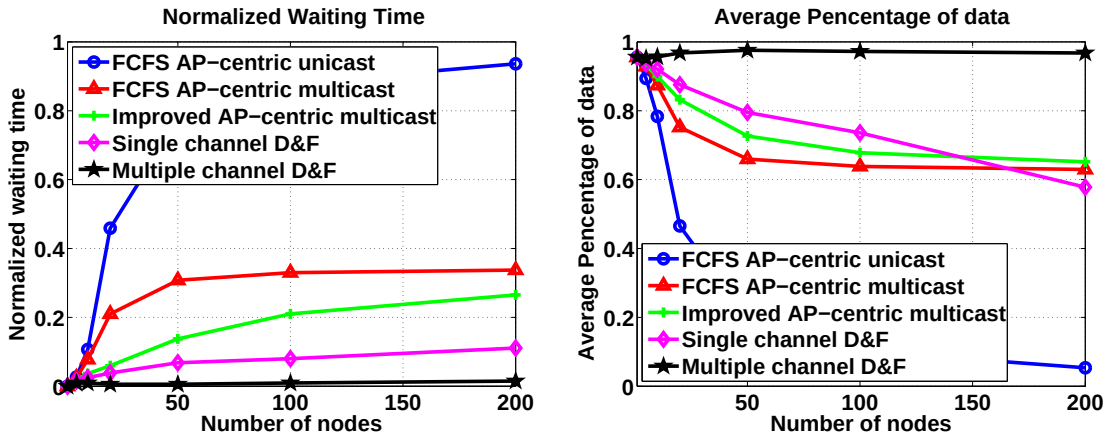
3.8.2 AP-centric and Deputy&Forward Methods

Here we will test the performance of five strategies, FCFS unicast, FCFS multicast, improved multicast, single channel Deputy&Forward and multiple channel Deputy&Forward. In the multiple channel case, we suppose there are nine channels available, so that if there exists D&F nodes in each location, they can transmit simultaneously.

Effects of the Number of Nodes

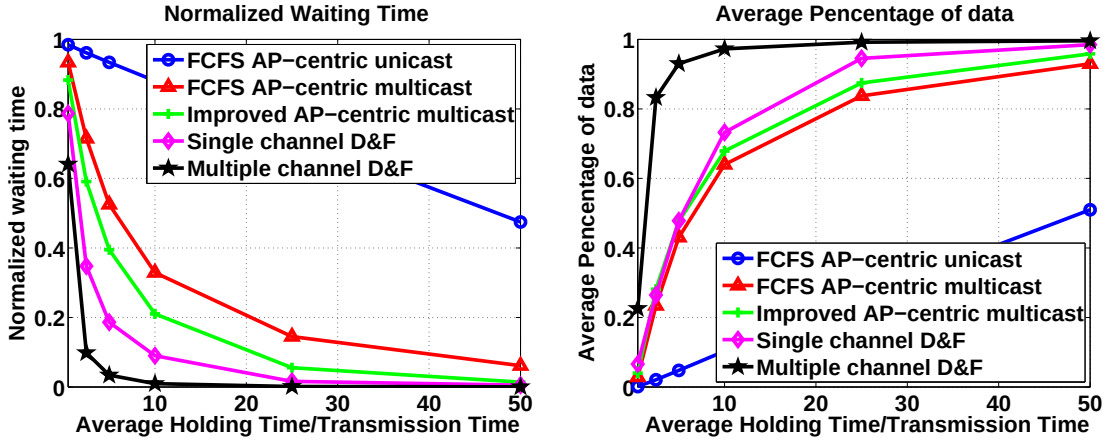
As we increase the number of nodes in the system from 1 to 200, the Deputy&Forward method has less delay than the AP-centric method, as shown in Fig. 3.11(a). This result is because a D&F node needs less transmission time than an AP and thus it has more transmissions within the same period of time. Additionally, the multiple channel method is better than the single channel case as it behaves as multiple servers that work simultaneously.

For the same reason, nodes can get more data via a multiple channel Deputy&Forward method. But as the number of nodes becomes large(more than 200) in Fig. 3.11(b), nodes in the single channel Deputy&Forward get less data than the improved multicast AP-centric method. This is because, in single channel Deputy&Forward, a D&F node generally serves less nodes in each transmission. We can understand this better using a simple example. Suppose that for one location during a complete duration of a D&F node's data delivery, 10 nodes come from other locations, and during a complete AP's data delivery, 20 nodes come from other locations(because D&F nodes require less transmission time). Then, although the average delay decreases, fewer nodes get the broadcast messages in the next D&F node's data delivery. As the number of nodes increases, the advantage of serving more nodes in



(a) (b)

Figure 3.11: AP-Centric vs. Deputy&Forward methods with Varying Number of Nodes, (a) Normalized waiting time of each node, (b) Average percentage of location-based data each node gets in each transmission.



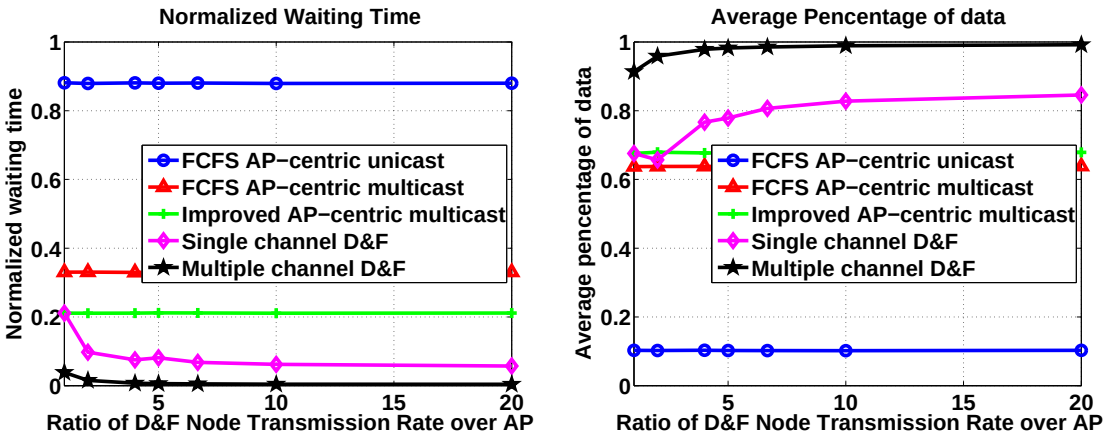
(a) (b)

Figure 3.12: AP-Centric vs. Deputy&Forward methods with Varying Ratio of Holding Time over AP's Transmission Time, (a) Normalized waiting time of each node, (b) Average percentage of location-based data each node gets in each transmission.

improved multicast AP-centric method is dominant and provides nodes with more data.

Effects of Ratio of Holding Time over AP's Transmission Time

In this simulation, we fix the transmission time of each location as 20 seconds and increase the mean holding time from 20 to 1000 seconds. Thus, the ratio of holding time over AP's transmission time ranges from 1 to 50. In Fig. 3.12, the Deputy&Forward method



(a) (b)

Figure 3.13: AP-Centric vs. Deputy&Forward methods with Varying Ratio of D&F Node's Transmission Rate over AP's, (a) Normalized waiting time of each node, (b) Average percentage of location-based data each node gets in each transmission.

shows better performance than the AP-centric method. In addition, it shows the better performance when the ratio is larger. As holding time is large, the number of transitions is less, and the workload is less. Fig. 3.12(b) also shows when the ratio is small, nodes in the single channel Deputy&Forward method get a similar amount of data as in the improved AP-centric multicast method. This is because, in this case, nodes do not have sufficient holding time to receive data themselves, and thus cannot become D&F nodes. However, the single channel Deputy&Forward becomes much better as the ratio increases, since more nodes can behave as D&F nodes as holding time has increased.

Effects of Ratio of Transmission Rate of D&F node over AP's

In this study, we fix the AP's transmission time for location-based data to be 20 seconds, and we varied the D&F nodes' transmission time from 1 to 20 seconds. Thus, the ratio of D&F Node's Transmission Rate over AP's varies from 1 to 20.

An increase in this ratio does not affect the performance of the AP-centric method. But for Deputy&Forward methods, the delay decreases with the increase of this ratio. As a result of an increase in this ratio, the D&F method will have more opportunities for the D&F nodes to transmit location-based data. An interesting phenomenon is that for the single channel Deputy&Forward method, as the ratio approaches one, we get the same performance as the

improved AP-centric multicast method. This justifies the fact that Deputy&Forward strategy has better performance because it exploits smaller transmission times. Additionally, the percentages of data that each node increases with the ratio of the D&F node's transmission rate over an AP's keeps increasing for multiple channel Deputy&Forward method. However, when the ratio becomes 2, Fig. 3.13(b) shows there is a decrease of percentage of data nodes get in single channel Deputy&Forward. This is as we mentioned, D&F nodes may serve fewer nodes in each transmission. Thereby, although the delay decreases, nodes get less data. But as the ratio continues to increase, the advantage of fast rates dominates, and nodes receive more data.

3.9 Conclusion

In this chapter, we examine a service where mobile nodes access data according to their locations. As this data will likely be multimedia data, the transmission time is not negligible and will introduce tremendous delays when the system supports many users simultaneously. In order to solve this problem, we describe the moving pattern of a mobile node as a semi-Markov process and formulate a criterion for optimizing the service strategy. We first evaluate this criterion for an AP-centric method, where all the data is transmitted by APs. Since an AP should choose a service cluster of nodes that would introduce the least delay, we have proposed an improved multicast strategy. To further improve performance, we presented the Deputy&Forward method, in which nodes who have previously received location-based data can assist the system by serving nodes that newly arrive at the location. The Deputy&Forward method is a better strategy as these nodes can transmit with faster rates. We discuss two Deputy&Forward methods, single channel and multiple channel Deputy&Forward and analyze their serving strategies. Based on simulation studies, we have shown that the improved AP-centric method has better performance than three baseline strategies and that the Deputy&Forward method can achieve better latency and throughput than the AP-centric method.

Algorithm 5 Update of Lists in Single Channel Deputy&Forward

Require: op, df_1, r_{op}, q_{op}, topology update, D&F lists and waiting lists
Ensure: Update of lists

if Topology update{suppose Node c move from L_i to L_j} **then**
  if Node c was in $\mathcal{L}_{D\&F_i}$ **then**
    // Update D&F list of the previous location
    if $i == op$ and $\texttt{Deli\_flag}(\mathcal{L}_{D\&F_{op}})==1$ and $r_{op} \neq 0$ and $c == df_1$ **then**
      $\texttt{Usage}$(node c)=$\texttt{Usage}$(node c)+$r_{op}\widetilde{t}_{df_i}$
      if $q_{op} == 1$ || $\texttt{Len}(\mathcal{L}_{D\&F_{op}})== 1$ **then**
        $\texttt{Deli\_flag}(\mathcal{L}_{D\&F_{op}})$=0
      end if
    end if
    Remove node c from $\mathcal{L}_{D\&F_i}$, $\texttt{Len}(\mathcal{L}_{D\&F_i})$=$\texttt{Len}(\mathcal{L}_{D\&F_i})$-1
  else
    // Update waiting list of the previous location
    Remove node c from $\mathcal{L}_{W_i}$, $\texttt{Len}(\mathcal{L}_{W_i})$=$\texttt{Len}(\mathcal{L}_{W_i})$-1
    if node c is receiving data **then**
      $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_i})$=$\texttt{Len\_In\_Deli}(\mathcal{L}_{W_i})$-1
    end if
  end if
  // Update waiting list of the new location
  Append node c to the end of $\mathcal{L}_{W_j}$, $\texttt{Len}(\mathcal{L}_{W_j})$=$\texttt{Len}(\mathcal{L}_{W_j})$+1
else
  if a transmission is finished **then**
    // Update D&F list and waiting list of served location
    $\texttt{Usage}$(node df_1)=$\texttt{Usage}$(node $df1$)+$r_{op}\widetilde{t}_{df_i}$
    $\texttt{Deli\_flag}(\mathcal{L}_{D\&F_{op}})$=0, reorder $\mathcal{L}_{D\&F_{op}}$
    Remove first $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$ nodes in $\mathcal{L}_{W_{op}}$
    Insert them to $\mathcal{L}_{D\&F_{op}}$ and keep it orderly, $\texttt{Len}(\mathcal{L}_{D\&F_{op}})$=$\texttt{Len}(\mathcal{L}_{D\&F_{op}})$+$\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$
    $\texttt{Len}(\mathcal{L}_{W_{op}})$=$\texttt{Len}(\mathcal{L}_{W_{op}})$-$\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$, $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$=0
  end if
else
  if a transmission begins **then**
    // Update D&F list and waiting list of served location
    $\texttt{Deli\_flag}(\mathcal{L}_{D\&F_{op}})$=1
    $\texttt{Len\_In\_Deli}(\mathcal{L}_{W_{op}})$=$\texttt{Len}(\mathcal{L}_{D\&F_{op}})$
  end if
else
  if Battery Low Alarm is received **then**
    // Update D&F list of served location
    $\texttt{Usage}$(node df_1)=$\texttt{Usage}$(node $df1$)+Maximum Allowed Usage+1
    Sort $\mathcal{L}_{D\&F_{op}}$
  end if
end if

Chapter 4

Facilitating an Active Transmit-only RFID System Through Receiver-based Processing

RFID technologies promise the ability to monitor a variety of assets [55–57]. Although RFID technologies have had many success stories, such as the EZ-Pass system for electronic toll collection, the ability for RFID systems to simultaneously monitor a large amount of items, such as would be needed for more tightly managing inventory, while also having low-cost, has continued to be a significant challenge. Ideally, in order for an enterprise to track its assets, it is desirable to identify precisely where individual items are at any moment over an extended period of time. Passive tags that depend on harvesting power from a basestation have performance bounded by the regulatory limits of costly high-power basestations (e.g. on the order of 4Watts). Alternatively, at lower frequencies they work well, but only for short range ($\sim$ 1 cm) sensing, which cannot provide continual tracking. Active tags overcome many of these limits and provide improved range and reliability. Unfortunately, the standard assumption that such tags would consume a large amount of power has made it impossible to continuously monitor over a period of years.

In order to avoid the shortcomings associated with both types of tags, in [58], we proposed to adopt transmit-only active tags, which have the long range of traditional active tags, but without their high power consumption. In a system built on such transmit-only tags, tags periodically announce their presence by sending out their tag IDs, and the processing burden is placed on the tag reader. Since transmit-only tags cannot sense the channel, their transmissions are likely to collide with each other, especially for a system with a large number of tags. Thus, the tag reader must employ effective multi-user detection schemes to extract tag IDs from collided signals. In our earlier work [58], we tested the feasibility of putting together such a system by using several simple detection schemes. While our results in [58] provided some initial support towards building a realistic tracking system using transmit-only tags, the detection accuracy left room for significant improvement, especially

for a dense RFID system.

To address this need, in this study, we focus on the development of a specialized multi-user detection scheme suitable for transmit-only RFID systems. Our starting point is the popular successive cancelation algorithm. We formulate the successive cancelation algorithm in the context of our random on-off keyed tag signals, and discuss the algorithm for both coherent and non-coherent detection scenarios. As successive cancelation suffers from high computational and memory complexity, which is disadvantageous for real-time asset tracking, we then present an improved detection scheme that significantly reduces resource complexity while maintaining desirable detection performance. We then examine the performance of our tag detection scheme by evaluating it in the context of a broader system. For this system, we make each tag reader responsible for multiple tags, and discuss several approaches to achieve efficient tag handoff between readers.

The thesis is organized as follows. First, we will examine related work in Section 4.1. Then, in Section 4.2, we introduce our basic RFID system and communication model. Next, in Section 4.3, we turn to the problem of identifying tags in spite of collisions by using a successive cancelation algorithm that has been customized for our RFID problem. In Section 4.4, we support the feasibility of our approach by providing a scalability analysis, and also propose an algorithm for updating tag lists as tagged items move through the environment. An evaluation is provided in Section 4.5, and we conclude the thesis in Section 4.6.

4.1 Related Work

Recently, radio frequency identification has attracted significant research interest [56, 57]. However, in order for RFID to succeed, it is necessary to have low-cost RFID systems that have good performance [57], and thus there has been extensive work to design a low cost tag [59]. One particularly difficult challenge that is faced by RFID tags is the issue of multiple tags transmitting at the same time. For contact-read tags, this issue is not serious, but for non-contact tags, the issue of identifying tags in spite of the potential for collisions is very significant. One approach to handling collisions is to employ a basic CSMA-style medium access control mechanism [60], or other MAC protocol, such as an ALOHA [61–65] strategy. More recently, for tags with receive capabilities, a tree based approach has been

explored [66–70].

All of the above methods require the tags to have receive functionality, which inevitably increases the cost and power consumption of a tag. In this thesis, we propose to completely remove the receiver from the tag. Rather, each tag transmits periodically, and we place the task of detection and collision resolution on the readers. This is similar to DS/CDMA [71,72], which allows multiple users transmit simultaneously, and the receiver decodes transmissions via de-spreading. For such systems, there are many detection methods, such as the optimum receiver, MMSE receiver, and successive cancelation methods [72–75].

However, unlike conventional multi-user communications, we do not have a feedback mechanism for power control, as is used in [72–75]. Rather, in this work we estimate which tags are present using a modified successive cancelation method that estimates the amount of tags present as well as the received signal level at the base station. Recently, [58] introduced a method using derivatives of correlation functions to estimate the transmission times of tags in an RFID system. However, because of near-far problems, the derivative method does not perform well in some cases. To handle these problems, this thesis improves upon traditional successive cancelation [71, 72, 74, 75] and the derivative method.

4.2 System Model

4.2.1 RFID System Model

A typical RFID system is composed of four components: the tag, the reader, application software that makes use of the data at the reader, and a computer system that is connected by the reader.

Similarly, in our model, there are a number of tags. Each tag has a unique identifier of length L (in this thesis, we shall use $L = 100\delta$ for our discussions, where δ is the time taken to transmit one bit and thus serves as the unit). In general, a proper L is chosen to ensure that the necessary number of tags is far less than 2^L. We shall assume that the tags transmit their tag identifiers as beacons in order to support the detection of individual tags, and that the tags transmit periodically with a period of T bits of transmission duration. In order to reduce the collisions, $T >> L$. On the other hand, T should not be too large, so that real-time detection is guaranteed. A data rate of $1Mb/s$ was chosen as this corresponds to the rate supported by the current generation of low-cost radio chips, giving $L = 10^{-4}$sec

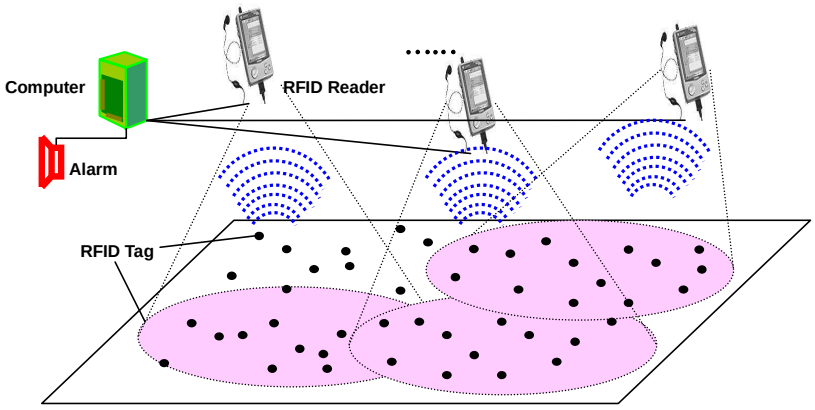


Figure 4.1: RFID System Model.

and we used $T = 1$ sec for the period.

In our paradigm, the communication is one-way and asynchronous, i.e. tags only transmit signals. This greatly simplifies the logic on board a tag, and thus a tag's cost can be significantly reduced. There are two meanings of asynchronization. First, there is no synchronization between the transmission of tags. Since we need the logic of tags to be simple, we can neither require all tags transmit simultaneously nor in a TDMA manner. Further, the communication between tags and the reader is without synchronization, which implies that, in order to detect each tag, we need to also estimate the transmission time of that tag.

The tag reader, which we shall often refer to as a basestation, consists of an RF frontend that downshifts the received waveform to baseband, performs A/D, and supplies an appropriately sampled waveform to a processor for the detection and identification of the tags. The processing of the sampled waveform could be performed on-board (if the reader has a sufficiently powerful DSP/FPGA), or can be performed off-board by a PC (e.g. samples may be transferred via PCI Express), as depicted in Figure 4.1. Our objective is to detect the tags in collisions without MAC or synchronization through received-based processing.

4.2.2 RFID Communication Model

For the sake of simplicity and cost-effectiveness, our tags use on-off keying (OOK) as the basic modulation scheme. This choice is motivated by our system implementation effort, where we have chosen a radio chip that uses OOK. For tag i, suppose that the transmitting

baseband signal is $C_i(t)$, which is a randomly generated sequence composed of 1 or 0. Here, $i \in \{1, \ldots, N\}$, represents the index of the tag, and N is the total number of tags in the system. Since we use OOK modulation, given the carrier frequency ω_c and the phase ϕ_i, the modulated signal will be $C_i(t)\, cos(\omega_c t + \phi_i)$.

Suppose the distance of tag i to the reader is d_i and the starting time of the transmission is τ_i, with τ_i being an integer multiple of the unit δ. In a wireless fading environment, because the tag periodically transmits with period T and $T >> L$, the received signal for tag i is

$$r_i(t) = \sum_{n=0}^{\infty} f_i(d_i)\, C_i(t - nT - \tau_i)\, cos(\omega_c t + \phi_i) \tag{4.1}$$

where $f_i(d_i)$ is the received amplitude resulting from path loss and the fading of tag i's signal.

Thus, the complete received signal is

$$\begin{aligned} r(t) =& \sum_{i=1}^{N} r_i(t) + n_w(t) \\ =& \sum_{i=1}^{N} \sum_{n=0}^{\infty} f_i(d_i)\, C_i(t - nT - \tau_i)\, cos(\omega_c t + \phi_i) \\ &+ n_w(t). \end{aligned} \tag{4.2}$$

At the demodulator, we first pass the received signal through a local oscillator, downshift and low-pass filter, to obtain the demodulated signal

$$\begin{aligned} R_I(t) =& LP\left\{ r(t)\, cos\left(\omega_c t + \widehat{\phi}\right) \right\} \\ =& \sum_{i=1}^{N} \sum_{n=0}^{\infty} f_i(d_i) C_i(t - nT - \tau_i)\, cos(\phi_i - \widehat{\phi})/2 \\ &+ n_I(t) \\ =& \sum_{i=1}^{N} \sum_{n=0}^{\infty} A_{I_i} C_i(t - nT - \tau_i) + n_I(t) \end{aligned} \tag{4.3}$$

where $n_I(t)$ is the I-phase filtered Gaussian noise and $A_{I_i} = f_i(d_i) cos(\phi_i - \widehat{\phi})/2$. Similarly,

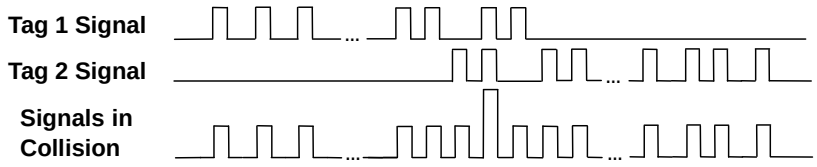


Figure 4.2: Tag Signals in Collision.

we can get the Q-phase received signal:

$$\begin{aligned} R_Q(t) =& - LP\left\{r(t)\, sin\left(\omega_c t + \widehat{\phi}\right)\right\} \\ =& \sum_{i=1}^{N}\sum_{n=0}^{\infty} f_i(d_i) C_i(t - nT - \tau_i)\, sin(\phi_i - \widehat{\phi})/2 \\ &+ n_Q(t) \\ =& \sum_{i=1}^{N}\sum_{n=0}^{\infty} A_{Q_i} C_i(t - nT - \tau_i) + n_Q(t) \end{aligned} \tag{4.4}$$

where $n_Q(t)$ is the Q-phase filtered Gaussian noise and $A_{Q_i} = f_i(d_i) sin(\phi_i - \widehat{\phi})/2$.

4.3 Detection Algorithm

Due to collisions, the received signal at the reader during any time interval is the composition of one or more tag signals, as shown in (4.3). These signals may mask or compromise each other and thus corrupt the detection process. In this section, we explain our strategy and our algorithms to solve the detection problem.

4.3.1 Detection Strategy

Since the transmission times of these tags are not known, and the received signal at any time may be just a corrupted signal, we will focus on a period T of the received signal as the input for processing. Because tags' transmissions are periodic, any received signal of duration T contains a complete description of the information needed for detection.

In order to detect the N RFID tags, we estimate the signal amplitude for each of the N tags, and decide whether the estimation is valid. Here, we use a predefined minimum received signal strength of $Threshold$ and declare the tag is present if the estimated signal strength is no smaller than $Threshold$. If the estimation is valid, we declare that the corresponding tag is present.

Suppose tags transmit independently, for tag j, the estimation of its received signal amplitude just depends on the received signal at the duration $[\tau_j, \tau_j + L)$. Thus, before the estimation of the signal amplitude, the transmission time of each tag must be estimated. The baseband tag signal is composed of L pulses, each corresponding to either a 0 or 1. For analysis, we assume the pulse $p(t)$ has duration δ, with $\int_0^\delta p(\tau)\, d\tau = \int_0^\delta p^2(\tau)\, d\tau = 1$, and represents a 1. Similarly, a pulse $g(t)$, with the same duration and $\int_0^\delta g(\tau)\, d\tau = \int_0^\delta g^2(\tau)\, d\tau = 1$ represents a 0. It is straightforward to show that the average autocorrelation of each tag signal is $L/2$, and that the crosscorrelation of tag signals is $L/4$. Therefore, we can perform a correlation between the received signal of duration T with each tag signal C_j of duration L. The position of the peak in the correlated signal is the estimated transmission time $\hat{\tau_j}$. For $0 \leq t < T$, the correlated signal is

$$\begin{aligned} \rho_j(t) = &\int_t^{t+L} \sum_{i=1}^{N} A_i C_i(\tau - \tau_i)\, C_j(\tau - t)\, d\tau \\ &+ \int_t^{t+L} C_j(\tau - t)\, n(\tau)\, d\tau. \end{aligned} \tag{4.5}$$

However, due to collisions, the estimation $\hat{\tau_j}$ may not be accurate, and the estimate only deteriorates further because of near-far issues. A natural solution is to estimate the tag with the maximum signal strength received at the reader, because it is most robust to collisions, and subtract its contribution from the received signal. By repeating this process, as long as the estimation of the received signal strength of each tag is sufficiently accurate, we remove large high-confidence components and amplify the presence of less powerful tags for further processing– a process known as successive cancelation [72].

4.3.2 Coherent Detection

For analytical simplicity, we first examine a coherent scheme. However, because coherent detection would require an increase in the cost of tags, we will not use it in our system. In coherent detection, we know the carrier phase of each tag. For simplicity, we do not align the phase $\hat{\phi}$ of the generated signal from the oscillator to the carrier phase ϕ_i. The phase information is used only at the later validation. Suppose the recovered baseband received signal from the I-phase and Q-phase component is $R(t) = \sum_{i=1}^{N} \sum_{n=0}^{\infty} A_i C_i(t - \tau_i) + n(t)$, which is composed of N tag signals, with a demodulated signal strength of A_i. Successive cancelation finds the tag signal with the maximum signal strength in each step, and subtracts

Set $R_0(t) = R(t)$, the received signal
$S_0 = \{indices\ of\ all\ tags\}$
for k=1 to N
  Get $\hat{r}_{k_{max}}$
  $R_k(t) = R_{k-1}(t) - \hat{r}_{k_{max}}$
  $S_k = S_{k-1} - k_{max}$
end for
for k=1 to N
  if $\frac{Amplitude(\hat{r}_{k_{max}})}{cos(\phi_i - \hat{\phi})} \geq Threshold$
  Then $\text{Tag}_{k_{max}}$ exists
end for

Figure 4.3: Successive Cancelation Method for Coherent Detection

for $j \in S_k$
  $\rho_{k_j} = correlate(R_{k-1}, C_j)$
  $c_j(\alpha_j) = max\left(\left|\rho_{k_j}\right|\right), \alpha_j \in [0, T)$
end for
$c_{max}(\hat{\alpha}_{k_{max}}) = argmax\ c_j(\alpha_j)$
$\hat{A}_{k_{max}} = 2\frac{\rho_{k_{k_{max}}}(\hat{\alpha}_{k_{max}}) - \sum_{l \neq k_{max}} \rho_{k_l}(\hat{\alpha}_{k_{max}})/[(N-1)]}{\int_0^L C_{k_{max}}^2(\tau)d\tau}$
$\hat{r}_{k_{max}} = \hat{A}_{k_{max}} C_{k_{max}}(t - \hat{\alpha}_{k_{max}})$

Figure 4.4: Estimate of $r_{k_{max}}$ from the residual signal $R_{k-1}(t)$.

this signal from the received signal, until all the tag signals are found, as shown in Fig. 4.3. S is the set which contains the index of undetected tags. At the first step, S_0 will consist of all tags indices. After each iteration, the index of the detected tag will be removed from the set.

Fig. 4.4 shows the estimation of both the transmission time and tag signal for the kth round. First, we get the maximum value $c_j(\alpha_j)$ of the absolute value of the correlated signal ρ_{k_j}, where $j \in S_{k-1}$. We use absolute value because A_i may be negative due to the phase offset. Next, the maximum value $c_{max}(\hat{\alpha}_{k_{max}})$ among all $c_j(\alpha_j)$ is obtained, which is the maximum value of all the correlations at the kth step. Then, $\hat{\alpha}_{k_{max}}$ is the corresponding estimation of the transmission time of tag k_{max}, and is the position of the maximum value at the correlated signal $\rho_{k_{k_{max}}}$. For notational convenience, we assume $j = k_{max}$.

In a correct detection, $\hat{\alpha}_j = \tau_j$. Without considering the noise,

$$\rho_{k_j}(\tau_j) = \int_{\tau_j}^{\tau_j + L} A_j {C_j}^2(\tau - \tau_j)\, d\tau + \sum_{\substack{|\tau_i - \tau_j| < L \\ i \neq j}} \int_{\tau_j}^{\tau_j + L} A_i C_i(\tau - \tau_i) C_j(\tau - \tau_j)\, d\tau \tag{4.6}$$

For other $l \neq j$,

$$\sum_{l \neq j} \rho_{k_j}(\tau_j) = \sum_{|\tau_i - \tau_j| < L} \int_{\tau_j}^{\tau_j + L} A_i C_i(\tau - \tau_i) \cdot \sum_{l \neq k_{max}} C_l(\tau - \tau_j)\, d\tau \tag{4.7}$$

Since each C_l is a random sequence of 0 or 1, and as N is big, according to law of large numbers, each bit of the sequence $\sum_{l \neq j} C_l(\tau - \tau_j)$ has roughly value $(N-1)/2$. Thus, $\sum_{l \neq j} \rho_{k_l}(\tau_j)/[(N-1)] = \sum_{|\tau_i - \tau_j| < L} \int_{\tau_j}^{\tau_j + L} A_i C_i(\tau - \tau_i)/2 d\tau$. In addition, we know $E\left(\int_{\tau_j}^{\tau_j + L} A_i C_i(\tau - \tau_i) C_l(\tau - \tau_j)\, d\tau\right)$ is equal to $E\left(\int_{\tau_j}^{\tau_j + L} A_i C_i(\tau - \tau_i)\, d\tau\right)/2$.

Then, we can get

$$E\left(\rho_{k_j}(\tau_j)\right) = \int_{\tau_j}^{\tau_j + L} A_i {C_j}^2(\tau - \tau_j)/2 d\tau + \sum_{l \neq j} \rho_{l_i}(\tau_j)/[(N-1)] \tag{4.8}$$

Because every basis has similar weight, and each position is uniformly distributed with 1 or 0, we have

$$\hat{A}_j \approx 2 \frac{\rho_{k_j}(\tau_j) - \sum_{l \neq j} \rho_{l_i}(\tau_j)/[(N-1)]}{\int_0^L C_j^2(\tau)\, d\tau}. \tag{4.9}$$

Therefore, $\hat{r}_j = \hat{A}_j C_j(t - \tau_j)$. As shown in Fig. 4.4, there is an estimation of $\hat{A}_{k_{max}}$ for every cancelation. For coherent detection, we can set a threshold, and a tag is declared to be found if $\hat{A}_{k_{max}}/cos(\phi_i - \widehat{\phi}) \geq Threshold$.

4.3.3 Noncoherent Detection

Now we extend the coherent scheme to handle the more general, noncoherent case where we do not know the carrier phase of each tag. Since the carrier phase of tags are randomly distributed, the contribution of some tags may be very small for either I-phase or Q-phase received signal, which complicates the near-far effect. Further, for non-coherent detection, because we don't know each tag's carrier phase, it is hard to set a proper $Threshold$ to validate a correct estimation.

However, we note that typically, if the I-phase tag signal is small its Q-phase signal is large (and vice versa), due to the complementary properties of the trigonometric functions. This allows us to perform successive cancelation separately for both I-phase and Q-phase signals.

Set $R_{I_0}(t) = R_I(t)$, $R_{Q_0}(t) = R_Q(t)$
$S_{I_0} = S_{Q_0} = \{indices\ of\ all\ tags\}$
$\underline{\hat{A_I}} = \underline{\hat{A_Q}} = zeros(1, N)$
for k=1 to N
  Perform estimation on R_{I_0} and R_{Q_0} as in Fig. 4.3
  Store estimation in $\underline{\hat{A_I}}$ and $\underline{\hat{A_Q}}$, respectively
end for
for k=1 to N
  if $\sqrt{\underline{\hat{A_I}}^2(k) + \underline{\hat{A_Q}}^2(k)} \geq Threshold$
  Then Tag_k exists
end for

Figure 4.5: Successive Cancelation Method for Noncoherent Detection

The algorithm is summarized in Fig. 4.5. For each channel of detection, the estimation is similar to the coherent case. However, after we get the estimation of $\hat{A_{I_i}}$ and $\hat{A_{Q_i}}$, for $i = 1, \ldots, N$, we declare we find tag_i as long as $\sqrt{\hat{A_{I_i}}^2(i) + \hat{A_{Q_i}}^2(i)} \geq Threshold$. By making use of both results, the final detection and estimation will be more accurate and complete. Further, the phase information is no longer important to the validation.

Naturally, we may get two different estimations of transmission times $\hat{\tau_{I_i}}$ and $\hat{\tau_{Q_i}}$ for tag_i. We believe a large $\hat{A}$ will give a more accurate $\hat{\tau}$. Therefore, we decide $\hat{\tau_i} = \hat{\tau_{I_i}}$ if $\hat{A_{I_i}} \geq \hat{A_{Q_i}}$, otherwise, $\hat{\tau_i} = \hat{\tau_{Q_i}}$. Further, if either $\hat{A_{I_i}}$ or $\hat{A_{Q_i}}$ is small and inaccurate, the implications are minimal. In addition, we can set $Threshold$ slightly smaller than the minimum received signal strength to allow some fault tolerance. For the remainder of the chapter, we restrict our discussion to non-coherent detection.

4.3.4 Overlap Reduced Successive Cancelation Method

A disadvantage of successive cancelation is its intensive computational and memory requirements. Further, the computational load doubles for non-coherent detection. We now propose an algorithm which can reduce the cost of computation by dividing the long received signal into contiguous overlapping blocks of manageable length, say T_s samples, then performing successive cancelations for each block. On average, since the transmission time of all tags are randomly distributed with $[0, T)$, we can detect NT_s/T tags for each block. Thus, at the second block, we need only perform correlation for $N - NT_s/T$ tags. As for the last block, only NT_s/T correlations need to be performed. For example, if $T = 10^6$, $T_s = 10^4$, the computation can be reduced by about 2. Since there might exist tags at the boundaries of each block, the neighboring blocks needs to overlap each other by a length of L, in order to detect every tag.

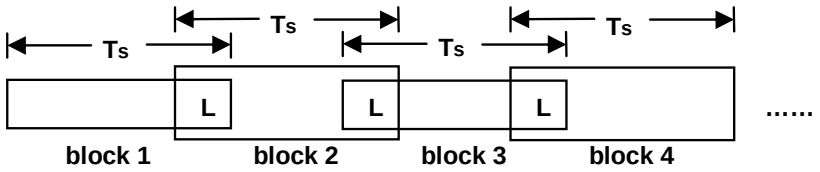


Figure 4.6: Overlap Reduce Successive Cancelation Method.

In order to keep the correlated values of tags for direct successive cancelation method, we need $4NT$ bytes of memory if every correlated value needs four bytes. However, for our overlap reduced successive cancelation method, since we perform detection block-by-block, the necessary memory to store the correlated values is only $4NT_s$ bytes. Thus, the relative necessary memory of overlap method over the conventional method is T_s/T. For example, if $T = 10^6$, $T_s = 10^4$, the memory needed is reduced by a factor of 100.

4.4 Scalability Analysis

We now consider a general asset tracking application with a couple of thousand tags in a fairly spread out environment. Detection of all tags at a single receiver in such a scenario based on multi-user detection (MUD) will be extremely computationally intensive. A single receiver deployed in such an environment will also experience the near-far effect due to the large area covered.

In considering scaling, we will look at a deployment, as in Fig. 4.1, where the networked basestations are mounted in the ceiling and spread out over the volume to be monitored. The density of basestations is chosen to guarantee each tag can be loudly heard by at least one reader, and the number of tags per basestation is within the tolerance bounds with our algorithm. We believe such a basestation can be manufactured at a cost low enough to make this architecture attractive.

Each basestation ensures that it is dealing with a limited subset of tags from the universe of tags in the environment by maintaining a tag list $\underline{L_i}$, where $i = 1, \ldots, M$ and M is the number of readers. Reader$_i$ is only responsible for the tags present in its tag list $\underline{L_i}$. As long as each tag is properly taken care of by one or more readers, the total detection is complete and accurate. During operation, we define $\underline{\widehat{L_i}} = \{i_1, \cdots, i_j, \cdots\}$ as the list of tags that reader$_i$ detected in one or more rounds of detection. In addition, the estimation

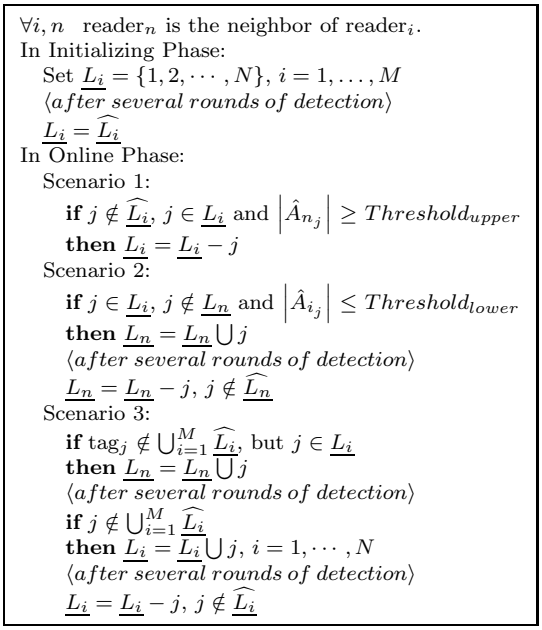


Figure 4.7: Update of Tag List

of signal strength of each tag from reader$_i$ is kept in $\underline{\hat{A}_i} = \left\{\left|\hat{A}_{i_1}\right|, \cdots, \left|\hat{A}_{i_j}\right|, \cdots\right\}$. These lists are initialized during system setup by having each reader check for all possible tags in the environment using an initialization algorithm and then switches to a higher speed operational algorithm afterwards.

4.4.1 Initializing Phase and Online Phase

The process is composed of two phases. In the initialization phase, readers don't know the distribution of the tags, and thus they have to scan all the tags. If a tag is detected by a reader, we say it is under this reader's coverage. A tag which is covered by multiple readers is called a boundary tag. In our simple propagation model, these tags will be at the boundary between the coverage areas shown in Fig. 4.1. For more complex environments, the geographical relation between the tags will not be as simple. Each reader keeps the tags that it covers in its tag list. In the initializing phase, we may run the detection for several rounds, to guarantee the tag list keeps complete information. The subsequent detections are all called second phase, or the online phase. In the second phase, each reader only needs to track the tags in its tag list.

4.4.2 Soft Handoff and Update of Tag List

The capability of the system to seamlessly monitor the movement of these tags in an integrated environment depends on the soft-handoff ability of our system. Periodic updates of the tag lists at each of the readers ensures a proper configuration of tag lists in realtime and makes systematic tracking of the tags feasible with low computation. Fig. 4.7 gives an overview of our distributed tag list update algorithm.

For initialization, we let each reader's tag list $\underline{L_i}$ contain all the tags. After several rounds of detections, each reader records the tags that it can detect in $\underline{\widehat{L_i}}$, and then sets $\underline{L_i} = \underline{\widehat{L_i}}$. In the online phase, there are three cases. The first case is that a tag loses the coverage of some readers. Meanwhile, the estimated signal strengths from that tag at the neighbors of these readers are above $Threshold_{upper}$. Hence, we know it is a boundary tag that has moved and remove its index from those readers' tag lists.

The second case happens if the estimated signal strength from a tag previously covered by only one reader drops below $Threshold_{lower}$. Then, we know this is because this tag has moved away from this reader. The surrounding readers will add this tag into its tag list. Over the next few rounds of detection, these readers will find out whether this tag is under their coverage. Those readers which don't cover it will remove it from its tag list.

The third case is that the system loses one tag. This may be because it has moved too fast or may be due to environmental interference. Since we know which readers last detected this tag, we add this tag into the neighbor readers' tag lists. After several rounds of detection, if we are still unable to find the tag, we add this tag into the tag lists of all the readers and remove the tag from the tag lists of the readers which don't cover the tag. We call this procedure detection compensation.

4.5 Simulation

Our simulations aim to study the effect of varying tag densities, near-far situations and signal-to-noise ratio on the overall detection performance. The detection performance of the system is measured in terms of probability of wrong detection of a tag and the mean squared error (ϵ) of the system in the amplitude estimation. Tag error rate is defined as the ratio of tags whose estimated transmission times do not conform to the actual ones. Suppose each tag has a correct amplitude A_i of demodulated baseband signal at the receiver,

while the amplitude estimate is $\hat{A}_i$, then the mean square error is $\epsilon = \sum_{i=1}^{N} \left(A_i - \hat{A}_i\right)^2 / N$. The ϵ parameter indicates the ability of the estimation of the amplitudes of tag signals and the tag probability of wrong detection describes the probability of a detection failure with an arbitrary packet. Though these two metrics are correlated, the information conveyed by these parameters is individually significant. The simulations compare the relative performance of four detection algorithms: Correlation with first and second order derivative based post processing [58], traditional successive interference cancelation scheme [74, 76] and our improved successive cancelation algorithm.

4.5.1 Simulation Setup

Our simulation environment consists of 100 tags with each tag broadcasting a 100 bit sequence in a $10^4 \mu s$ burst. The average channel utilization corresponds to 10^4 tags broadcasting with a period $1s$. The start time for the $T = 10^4 \mu s$ burst is randomly chosen by each tag. We assume the receiver correctly demodulates and uses a sufficient sampling rate, and thus our analysis is for baseband signals only.

The physical layout of the experiment matches Fig. 4.1 where the tags are randomly distributed in a square plane, and the networked basestations are positioned above this plane. The minimum distance between tags and basestation is d_{min}, and the maximum distance is d_{max}. Periodically, each tag sends its tag ID signal with a transmission time uniformly distributed in the range $[0, T)$. We employ a free space path loss model, where the relationship between the received signal strength P_r and the transmitting power P_t is $P_r = \left(\sqrt{G_l}\lambda/(4\pi d)\right)^2 P_t$ [19], for G_l, π, and λ are constants, and d is the distance between the transmitter and the receiver.

4.5.2 Effect Of Tag Collisions

To study the effect of collisions on the detection scheme, we vary the number of collisions in the system by varying the number of tag transmissions from 10 to 150 in a single burst of $10^4 \mu s$. Other parameters are set as $d_{min} = d_{max}$, and ambient noise in the system is set to 0. From Fig. 4.8 and Fig. 4.9, we see that generally the detection accuracy deteriorates as more collisions occur, for both the estimation of the transmission times and the amplitude. Due to the fact that tag signals will have varying levels of crosscorrelation, the overlapped

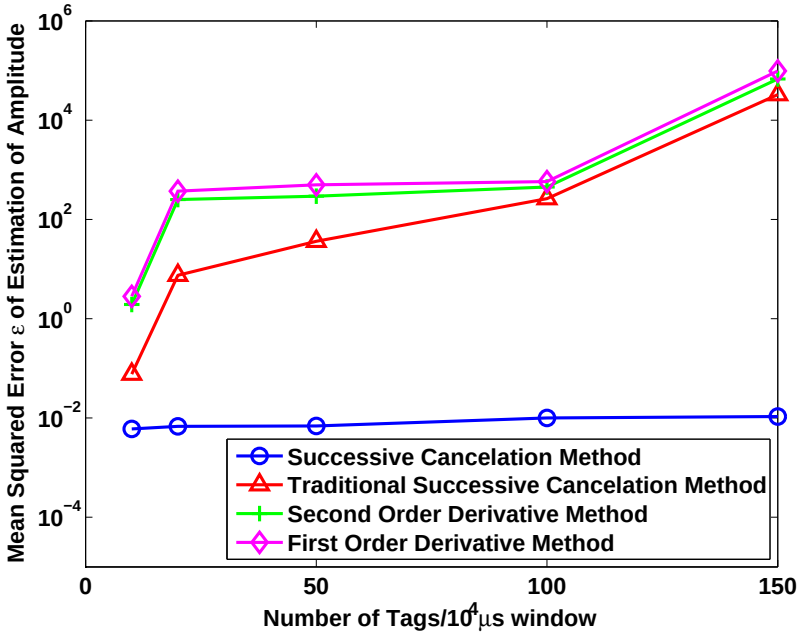


Figure 4.8: The Mean Square Error of Estimation of the Amplitude versus the Number of Tags per Reader, $d_{max} = d_{min}$ and noise is 0.

transmission of multiple tags can affect detection performance. In the plots, this is evident for the derivative methods of [58], but we note that successive interference cancelation has better results than the other approaches. The reason is that successive cancelation method deducts the loudest signal in each cancelation, thereby alleviating the collisions among tags. Moreover, since we use a statistical method to cancel the effect of cross correlations, our successive cancelation is superior to the traditional approach.

4.5.3 Near Far Effects

Large variations of the path loss experienced by the transmission of different tags will affect the detection accuracy of our algorithm. To study this effect we consider a scenario with 100 tags in the environment, the noise levels fixed at 0, and vary the d_{max}/d_{min} from 1 to 10 to test the consequences of the near far effect.

Fig. 4.11 shows that successive cancelation has a better performance than the derivative based post processing method. Elimination of the loudest tags in consecutive estimations by the successive cancelation approach enables it to assuage the effect of the loud tags on the detection thereby allowing the transmission of the soft tags to stand out in the residue

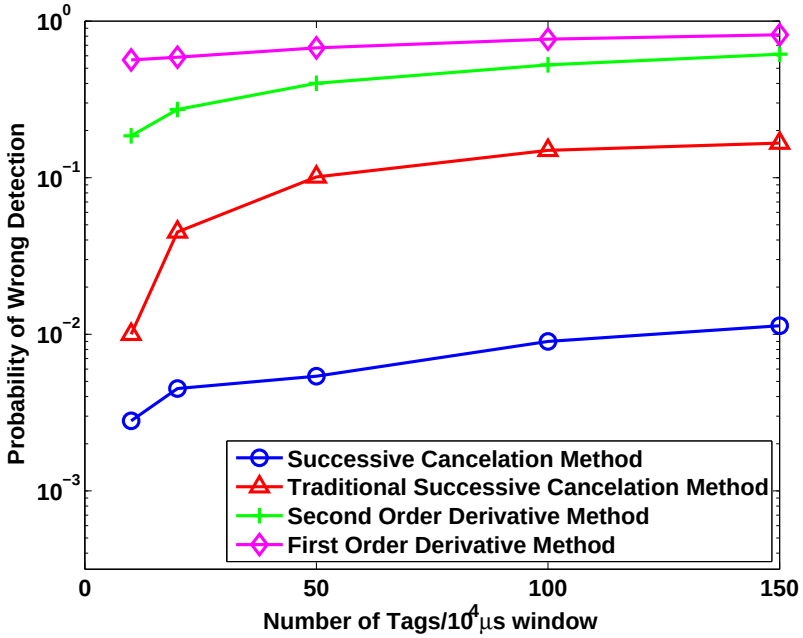


Figure 4.9: The Tag Error Rate with Increasing Number of Tags per Reader, $d_{max} = d_{min}$ and noise is 0.

of the received signal. Our successive cancelation approach also manages to outperform the conventional successive cancelation scheme since the estimation approach used with our algorithm is better suited to solve our problem than the generic one. The signal strength of tags signals are generally smaller as d_{max}/d_{min} increases, and the mean square error automatically decreases as shown in Fig. 4.10. We can see that mean square errors of successive cancelation methods decrease rapidly, which again supports its ability to handle near-far effects.

4.5.4 Noise Effects

The presence of noise on the channel will bias the received signal and thus produce estimation errors with the time and amplitude estimations. To explicitly study the effect of noise on the detection accuracy, simulations were done with 100 tags in a $10^4 \mu s$ transmission burst. The SNR was varied from 40 to 5 dB. To emphasize the effect of noise the d_{max}/d_{min} ratio was set at 5.

Fig. 4.12 plots the mean square error of the estimated amplitude as a function of the signal to noise ratio for the tag transmissions. Results show that the estimation error with

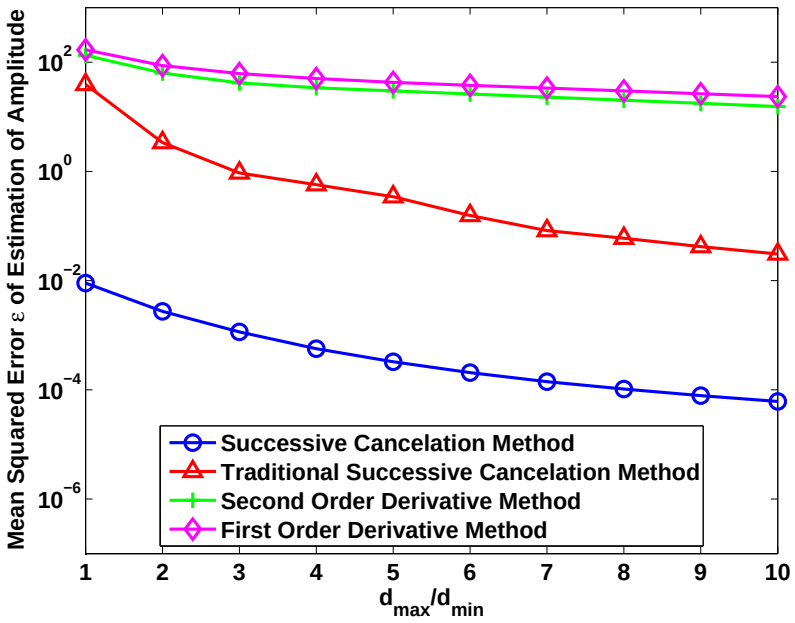


Figure 4.10: The Mean Square Error of Estimation of the Amplitude versus d_{max}/d_{min}, the Ratio of Maximum Distance to Minimum Distance between AP and Tags, with $N = 100$ in $10^4 \mu s$ Window and noise is 0.

our modified successive cancelation approach is far less than the other three. Since the other three methods do not attempt to make an accurate estimate of the amplitude of the received signal, the estimation error of these approaches with no power control is significant. Fig. 4.13 shows the probability of tag error for the same experiment. The plot shows that as the signal strength improves, the performance of our algorithm improves significantly as compared to the others due to better tag amplitude and transmission time estimation. From Fig. 4.12, we can see that our method overall has best performance among the four algorithms. The performance deteriorates as SNR decreases since the estimation of a tag's signal power degrades as tag power levels approach the noise level.

4.5.5 Scalability Test

The use of multiple readers in an integrated environment reduces the near-far effect thereby allowing for an improved detection accuracy with no power control mechanism. The goal of this experiment is to test the effect of varying reader count and using multiple detection rounds on a set of 100 tags transmitting over a $10^4 \mu s$ burst. Since this experiment aims at testing the performance in a real environment we are considering a worst case distance

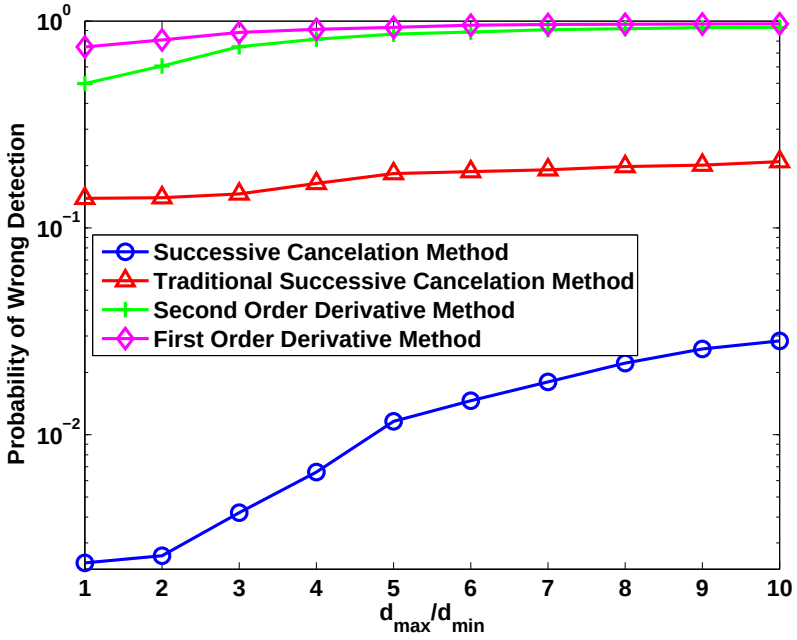


Figure 4.11: The Tag Error Rate versus d_{max}/d_{min}, the Ratio of Maximum Distance to Minimum Distance between AP and Tags, with $N = 100$ in $10^4 \mu s$ Window and noise is 0.

ratio with the $d_{max}/d_{min} = 10$. We consider the transmissions from the tags to be at a minimum 15dB SNR.

Fig. 4.14 plots the improvement in the detection performance with various amounts of readers, and detection performed over multiple rounds. A tag is *detected* if it is successfully identified by at least one of the readers in at least one of the multiple rounds. The initial reading with 1 reader and a single round of detection shows a particularly poor performance because of the exceptionally harsh conditions chosen our simulations. Even under these conditions, it can be observed that as the number of readers and the rounds increase there is a considerable improvement in detection accuracy.

Fig. 4.15 tests the non-coherent successive interference cancelation scheme versus the number of readers at different phases of operation. The online phase shows the performance of the detection scheme when each basestations is responsible for all the tags in the environment. To scale computation, the online phase optimizes the tag list at each basestation. However, it may be observed that excessive truncation of the tag list results in missing some important collision information, which degrades performance. To correct this, the online phase with compensation ensures that the tag lists are properly updated to deal with the

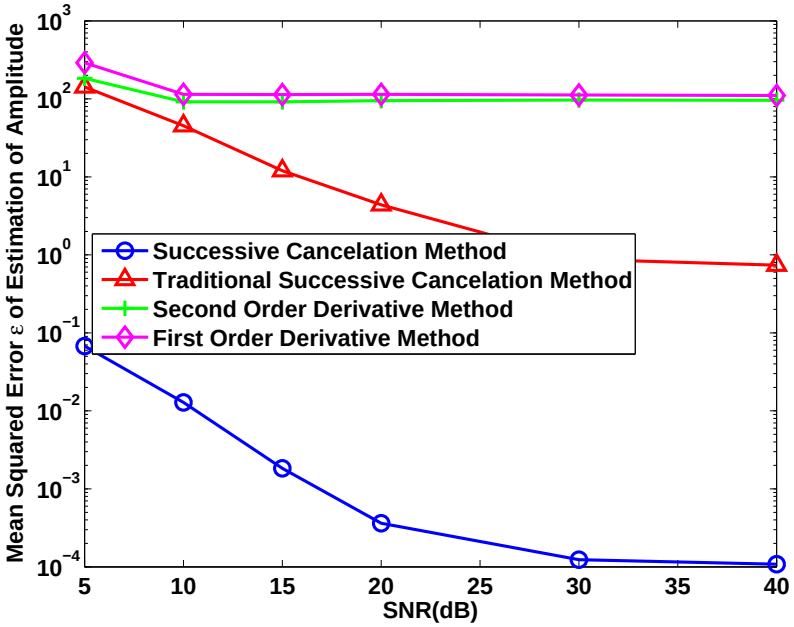


Figure 4.12: The Mean Square Error of Estimation of the Amplitude versus SNR(dB), with $d_{max}/d_{min} = 5$ and $N = 100$ in $10^4 \mu s$ Window.

correct set of tags. The corresponding improvement results in a near accurate detection even under extremely harsh testing conditions. For these experiments, we observed that each reader was responsible for roughly 20% of the tags during the online phase, thus reflecting a reduction in computational load at each reader.

4.5.6 Overlap Reduce Successive Cancelation Method

The overlap reduced successive cancelation method relies on iterative piecewise elimination of tag correlations to achieve improved detection efficiency. This improved efficiency in detection may produce an undesired loss in detection accuracy. This simulation aims to show the tradeoff that can be achieved between improved computation and detection accuracy with the use of OR-SC. Fig. 4.16 plots the ratio of the number of correctly detected tags as we progress through the blocks of computation. The experiment is run with a set of 100 nodes over a transmission burst of $10^4 \mu s$ in a plane with $d_{max}/d_{min} = 1$ and SNR at 15dB. We divide the received signal into 12 blocks, and compare the performance with increased number of blocks. It can be seen that as we progress with the experiment a small residual error begins to accumulate. This error can be attributed to the wrong estimation of some

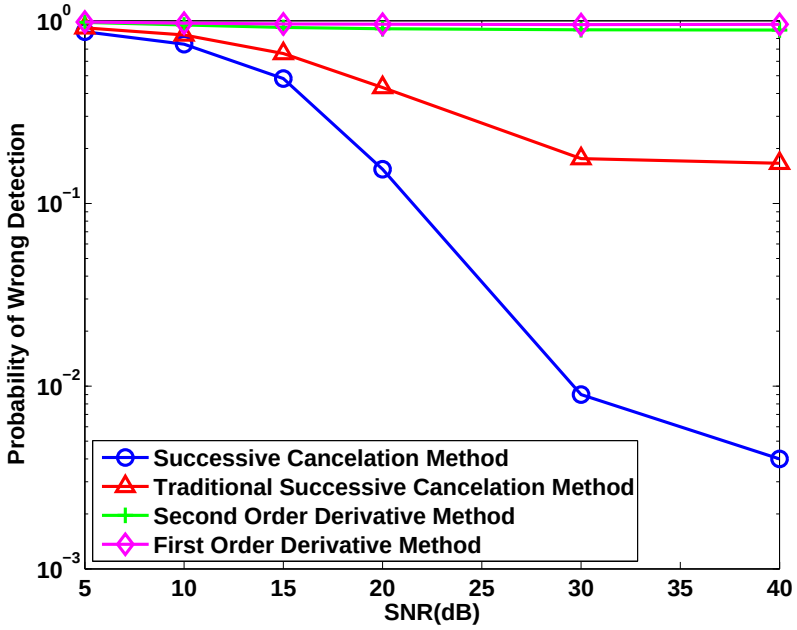


Figure 4.13: The Tag Error Rate versus SNR(dB), with $d_{max}/d_{min} = 5$ and $N = 100$ in $10^4 \mu s$ Window.

tags in the initial block which leads to a wrong estimation of other tags in the consecutive blocks.

4.6 Conclusion

In this chapter, we have proposed new methods to improve the tracking of receiverless transmit-only RFID tags. In a receiverless transmit-only RFID system, it is not possible to perform carrier sensing or collision avoidance and thus the challenge lies in resolving tag collisions. Our basic approach to address this problem is to utilize an enhanced form of multiuser detection at the receiver that can identify overlapping tag signals. We have developed a statistical algorithm to estimate signal amplitude and transmission time that exploits the properties of our tag system, and have integrated these algorithms to achieve an improved successive cancelation algorithm. Our successive cancelation method shows a better performance than the traditional successive cancelation method. Further, by making use of soft tag handoff between tag readers, and updating local tag lists at each reader, we have balanced the computational load across the entire system for improved scalability. In addition, we proposed a new overlap reduced successive cancelation method to further

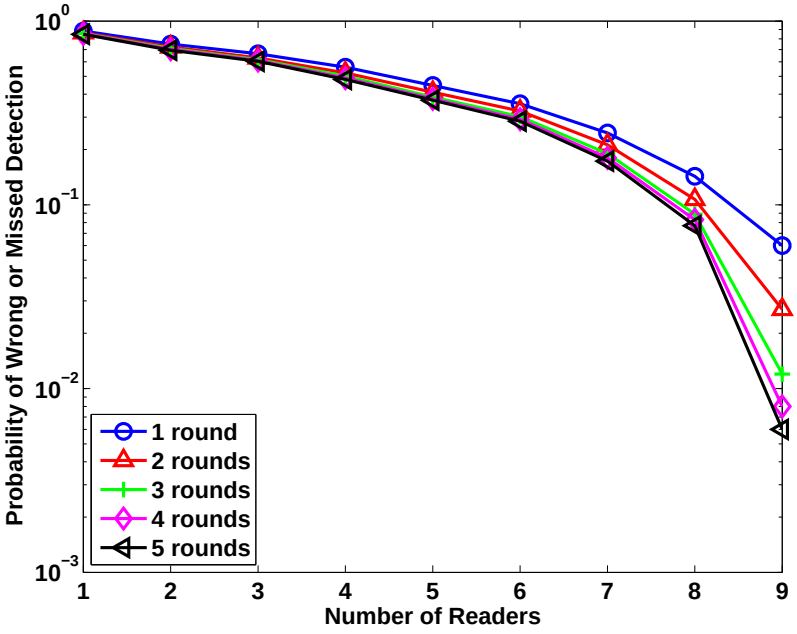


Figure 4.14: Detection Rate in the Initializing Phase, with $d_{max}/d_{min} = 10$, $N = 100$ and noise is 15dB.

reduce the intensive computation and memory costs associated with successive cancelation. The performance under different collision situations, varying levels of near far effects, and noise are examined in simulations, and it is shown that our approach can reliably detect a large number of tags in a realistic inventory-monitoring scenario.

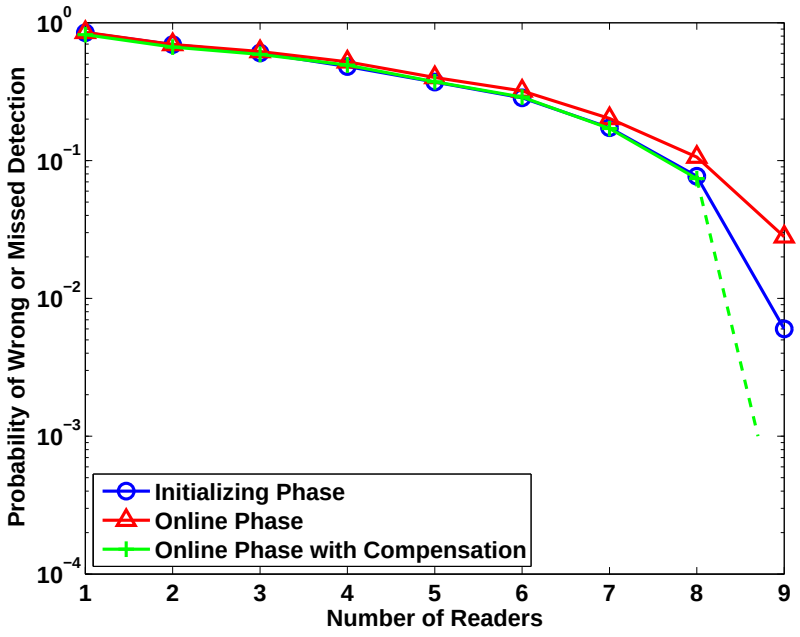


Figure 4.15: Detection Rate Compared between Initializing Phase, Online Phase and Online Phase with Compensation, with $d_{max}/d_{min} = 10$, $N = 100$ in $10^4 \mu s$ Window and SNR is 15dB. Please note that the dotted line was added to represent a number close to zero on a log scale plot.

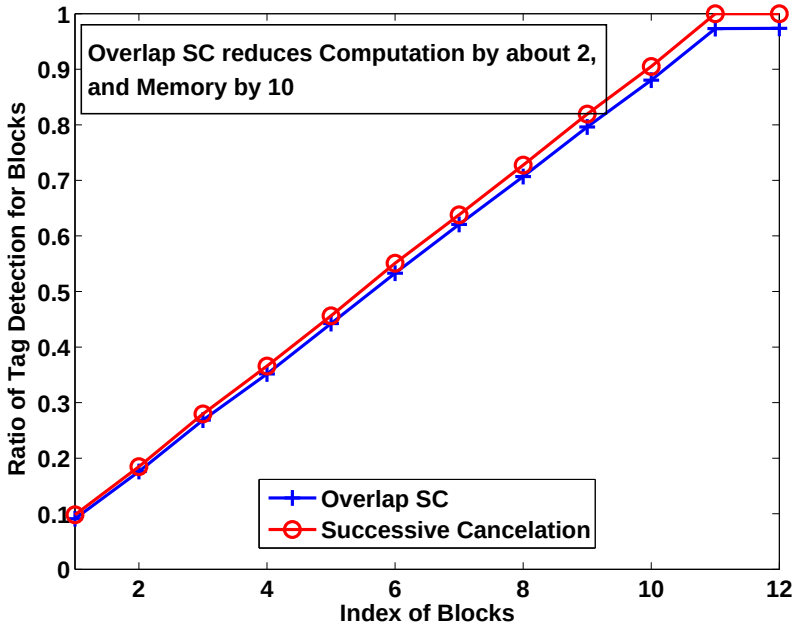


Figure 4.16: Comparison Between Overlap Reduce Successive Cancelation Method and Successive Cancelation Method, with $d_{max}/d_{min} = 1$, $N = 100$ and SNR is 15dB.

Chapter 5

Reactive On-board Regulation of Cognitive Radios based on Link Quality Estimation

5.1 Introduction

The development of improved chip designs and cognitive radio technologies promises to have a significant impact on the way wireless communication is performed, as a node can change its transmission, reception parameters, and communication protocols to communicate efficiently by avoiding interference with licensed or unlicensed users. Ideally, a cognitive radio has the ability to actively monitor several factors in the external and internal radio environment, such as radio frequency spectrum, user behavior and network state, so as to make a smart choice of when and how to transmit data.

As cognitive radios present us a promising picture, the adaptability of the lower layers of the protocol stacks also presents an obvious vulnerability. A node might abuse this freedom and choose a MAC protocol that does not consider the actual link quality, receiver processing speeds, or may even be intended to cause interference to other nodes' transmission. In particular, if this node bypasses higher layer traffic control mechanism and willingly inserts large amount of packets into a specific channel, the data transactions in this channel will be significantly affected.

Both the detection of a malicious or greedy MAC behavior and the defense mechanisms become more complicated in a distributed cognitive radio network, where there is no single, central regulating agent. The diversity of lower layer protocols not only greatly enhances the chances for interference, but also makes it hard to define a universal *legitimate* behavior. An example in point occurs when one transmitting node uses the Aloha MAC [77], while the second transmitting node chooses the CSMA MAC [77]. Even if the second node carrier senses before each transmission, the message will still collide with an immediately subsequent transmission from the first node, as the first node does not "listen" in Aloha.

Further complicating matters is the fact that generally cognitive radios operate in the more dynamic communication environments. As cognitive radios are allowed to adapt their MAC protocols [78, 79] and switch channels [80–89], there will be many traffic scenarios and the communication conditions will tend to vary drastically.

In this chapter, we intend to take an "onboard" approach to regulate a cognitive radio's behaviors in hopes to promote the effectiveness and fairness of data communication in a distributed cognitive radio network. The key idea is to locally observe traffic conditions and enforce the spectrum etiquettes (regulation policies) according to these observations. Our approach does not involve taking time to differentiate the cognitive radio's behavior, such as whether this is a greedy or malicious behavior and then take corresponding reactive responses, but instead to react at the exact moment that a "misbehavior" is detected. Additionally, the regulation takes consideration of the specific link quality and the level of severity of the "misbehavior".

We have the following contributions in this chapter. First of all, we propose a reactive onboard regulation approach, analyze its advantages and disadvantages, and the functions of the corresponding onboard regulation module. Second, through extensive experimental studies of two MAC protocols (CSMA and Aloha) with GNU radios on the ORBIT testbed, we explore the relationship between communication effectiveness and packet sending rates. Based on this relationship, we further present our onboard regulation algorithm, which first estimates the link quality and enforces regulations according to the estimation. Last, we implement our regulation approach and evaluate the performance on the ORBIT testbed.

This chapter is organized as follows. First, we will briefly examine the related work in Section 5.2. In Section 5.3, we introduce our system model and our reactive onboard regulation mechanism. After some field studies for different transmission behaviors using GNU radios in Section 5.4, we propose our detailed onboard regulation algorithms in Sections 5.5. In Section 5.6, we evaluate our regulation approach and conclude the chapter in Section 5.7.

5.2 Related Work

MAC protocol research for cognitive radios [78–87] has become very popular in recent years. It is commonly believed that there is no universal MAC protocol that is best suitable

for every communication scenario and nodes should be allowed to choose a proper MAC protocol [78, 79] according to the specific transmission conditions. The resulting diversity therefore brings up more complexity to the enforcement of fair and effective transmissions of the cognitive radio network.

Traditionally, this problem is often addressed by using a trusted third party or distribution of trust. A trusted third party (or a set of nodes) [90] can observe the behaviors of all the wireless nodes, and deny the access requests of nodes who break the spectrum ettiquettes. However, both the detection and the prohibition of a bad behavior are complicated by the breadth of spectrum frequencies (or channels) that these cognitive radios are allowed to access and the various spectrum access protocols that the radios are allowed to use. Therefore, it is easily conceivable that the idea of a monitoring trusted third party is impractical, or just partial solution.

Another method involves building a distribution of trust [91], and using this trust to decide the usage of the spectrum. Such an approach is very inefficient. Above all, the forming of distribution of trust values requires significant overhead. Even if an outlaw node is detected, the prohibition of future malicious and aggressive behaviors is hard to perform, because in a distributed network, there is no admission authority to control the channel access of each cognitive radio.

Traditional methods, which often take a rather long period to detect a malicious or greedy behavior and then take actions according to the detection, lag behind the network changes and thus no longer be the ideal solution to cognitive radio networks. We note that the above methods we discussed are *network-centric*, in which the network authority or the whole network takes the responsibility for detecting which nodes are malicious users and deciding which nodes are allowed to access the spectrum. In contrast, we would like to explore a *host-centric* method or an onboard mechanism, in which the node itself observes the network condition (e.g., how many ACKs it successfully received) and adjusts its behavior accordingly.

This onboard spectrum policy enforcement was first introduced in [92]. Different from the proactive method in [92], which verifies each transmission request according to the spectrum etiquette policies programmed by the user or spectrum owner, our method involves taking actions only when a misbehavior is performed. The regulation is taken by forcing the cognitive radio node to reduce its transmission. Unlike TCP in [93], our regulation

is performed in the MAC layer which cannot be bypassed. In addition, the onboard regulative module (the functional module on the cognitive radio platform that enforces the onboard regulation) does not use a sliding window for traffic control, but summarizes the average performance and takes regulations when the real-time performance deviates from the observed average performance.

5.3 Onboard Regulation

We introduce an onboard regulative approach, whereby an Onboard Regulative Module (ORM) on the cogitive radio node monitors the node's behavior and regulates its transmission according to the observations. This onboard regulative module is an additional functioning module added to the cognitive radio platform. For this platform, we will describe the functional part of a CR that performs the traditional cognitive radio MAC functions using the name 'CR Strategy Reasoner'. More specifically, by CR strategy reasoner, we mean the module by which the node changes its transmission or reception parameters to avoid interference from other users to achieve better performance. The Onboard Regulation Module, on the other hand, will be the functional module or components of a CR node that enforces spectrum etiquettes and regulates the behaviors of the CR strategy reasoner.

Unlike the CR strategy reasoner, we assume the ORM is a module that cannot be modified by developers. Further, it has higher authority than the CR strategy reasoner. When the observations of the CR strategy reasoner's behaviors are not compliant with policy, the ORM can prohibit the transmission instructions of CR strategy reasoner or force the CR strategy resoner to change its transmission or receiving parameters.

5.3.1 Proactive or Reactive Regulations

We begin with a discussion describing our rationale behind choosing between a proactive and an reactive regulation approach.

A proactive regulation scheme involves predicting whether a bad behavior is about to happen, before a transmission actually begins. In this strategy, the CR strategy reasoner is required to send a request to the ORM at every transmission attempt, along with its chosen transmission parameters. After analyzing whether this transmission obeys the spectrum policy (such as whether the chosen transmission power falls into a proper range), the ORM

decides whether to permit this transmission.

Obviously, outlaw behaviors should be prohibited so that the cognitive radio should not do any harm to the wireless network. However, it is challenging to have a thorough understanding and prediction of all the threats in a cognitive radio network, and thus it is not simple to embed an exhaustive specification of restricted actions into the ORM. Further, the strict prohibition of certain behaviors (such as requiring the transmission rate to be below a threshold), is also too restrictive as it implies that the CR cannot fully utilize the link capacity if the radio conditions are good and there is no risk to harming other nodes' performance.

An alternative reactive approach, involves conducting damage control after it happens, as opposed to trying to avoid a damage in the first place. To be more specific, the ORM does not block the transmission attempt initiated by the CR strategy reasoner unless it senses that damage has been done. The damage decision is made by analyzing the feedback the ORM receives.

The reactive approach does not require a detailed spectrum etiquette to be embedded in the ORM, and hence the design for a reactive approach can be relatively simple. Since the ORM takes regulation after an improper behavior is detected, even unforeseen problems can be regulated and the CR strategy reasoner can be allowed to make full use of the channel if conditions allow.

Because communication conditions are highly dynamic and hard to predict in a cognitive radio network, we believe that a reactive approach is the best approach to take.

5.3.2 Onboard Regulative Cognitive Radio Platform

Our basic architecture for the radio platform is presented in Fig. 5.1. In this design, the CR strategy reasoner and onboard regulative module are seperated components on a CR platform. The transmission parameters are chosen by the CR strategy reasoner, with all the data flows going through the ORM. If the "Traffic Analyzer" on the ORM determines that the traffic conditions are not good, the "Punishment Decider" would enforce a "punishment" on transmission attempts issued by the CR strategy reasoner by dropping some data packets or deferring the transmission of the corresponding data.

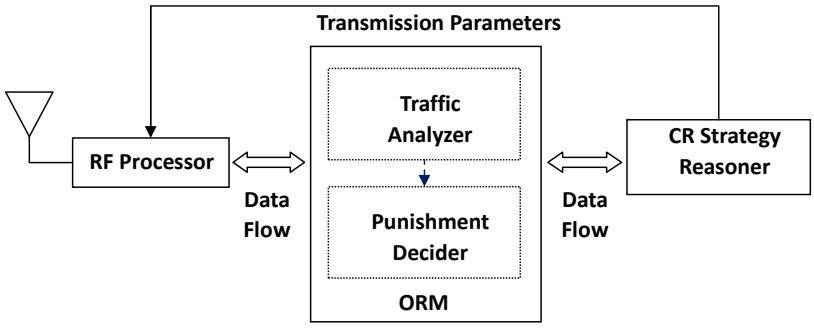


Figure 5.1: Reactive Onboard Regulative Cognitive Radio Platform.

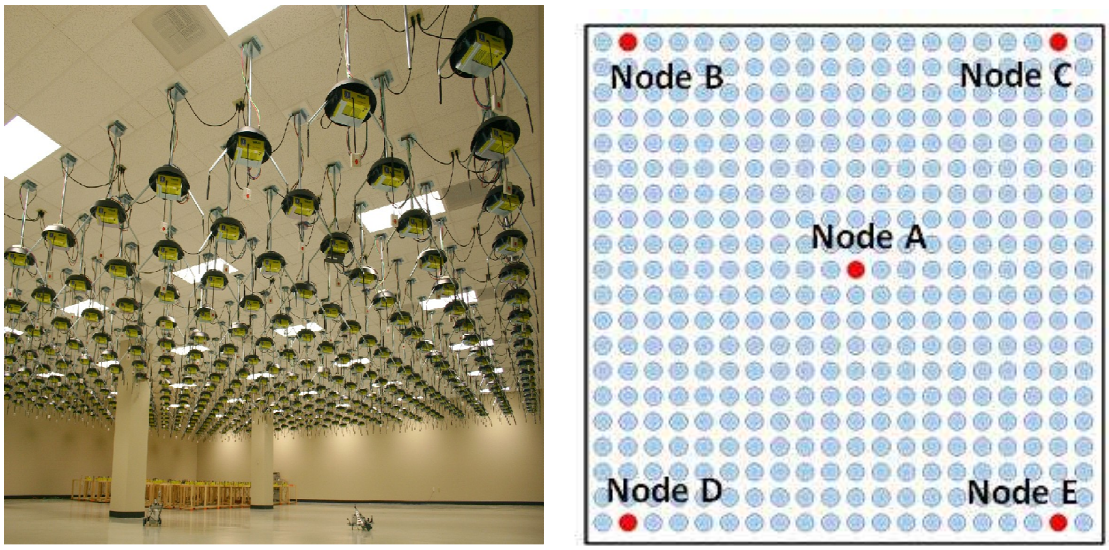
5.4 Traffic Condition Analysis

The analysis of traffic conditions is the foremost important step for reactive onboard regulation because the outcome of the analysis is the sole criterion for later regulation steps. In this section, we describe a suitable method for this analysis.

We start our discussion by discussing the regulation process to ensure fair and effective transmissions. A successful arrival and processing of a packet at the receiver represents an effective transmission, which requires an acceptably good link quality, good processing ability at the receiver, and most importantly, no interference. A fair transmission involves nodes in close vicinity allowing time for others to use the channel, so that each of them has opportunity for its own effective transmission. Therefore, if many packets cannot be received, decoded or processed (many ineffective transmissions), this should be regarded as a bad traffic condition. If a CR could regulate these ineffective transmissions and reduce its own transmissions, this would help improve the overall fairness of transmissions in the network.

If the ORM gets feedback of its transmission from the recipient, it can analyze how effective its own transmissions are and consequently estimate the traffic condtions. To achieve this goal, in our design, it is required that each node responds to every MAC frame it receives with an authenticated ACK. This ACK feedback mechanism is ensured by the ORM on each node. In other words, we require that the ORM responds with an ACK to each data packet it receives, unless the CR strategy reasoner does so.

It is particularly important to understand the performance scenarios for the ACK responses as they are related to different traffic conditions, such as the various packet sending



(a) (b)

Figure 5.2: Layout of Evaluation Setup, (a) Orbit testbed, (b) Positions of GNU radios used for evaluation.

rates and MAC protocols, in order to make good use of the feedback. We introduce a metric, ACK Loss Rate (which will be dicussed in Sec. 5.4.2), and make some field studies of the performance evaluation regarding ACK Loss Rate.

5.4.1 Experiment Setup

Our experiments were performed on the ORBIT testbed at WINLAB, Rutgers University, as shown in Fig. 5.2(a). Among the 400 nodes, five nodes are equipped with a USRP board (GNU Radio) and were chosen for our experiments. They are represented as Node A to Node E, as shown in Fig. 5.2(b). There are two daughterboards and one motherboard on each USRP. For both the transmission and the receiving, we used the RFX400 daughterboard. The modulation method employed was Gaussian Minimum Shift Keying (GMSK) with the carrier frequency at 423MHz. Unless otherwise specified, the transmission power was -10dBM for each antenna.

The GNU Radio is an off-the-shelf software-defined radio with which we could adapt the MAC protocols of the ORBIT node and simulate the MAC behaviors of a cognitive radio node. In our evaluation, we choose two MAC protocols: CSMA and Aloha. The CSMA MAC requires that the node "listens" to the channel before each transmission. A transmission is performed if the channel is vacant, or deferred according to an exponential

backoff rule. In our evaluation, the carrier sensing threshold was 25 dB. In contrast, in the Aloha MAC, a packet is sent out immediately without any carrier sensing.

Further, GNU Radios are transceivers with low reliability and processing speed, causing the performance variations to be much more obvious than would be noticed with other commercial off-the-shelf products in the market. For example, a GNU Radio node is only capable of a packet sending rate of a few packets per second. In addition, in our implementation, because no retransmissions and error-correcting codes are used, the packet error rate is very high. Consequently, using our implementation, we are able to observe a broader range of simulated cognitive radio network behaviors, including those scenarios with poor communication quality.

Throughout our experiments, Node A is a passive data server, which responds to every data frame it receives with an ACK to the sender, while Node B, C, D, E are the transmitters. For the sake of easy analysis, although Node B, C, D, E may change their MAC protocols, Node A always uses CSMA MAC. The statistical information associated with the ACKs at collected in the MAC layer by the corresponding transmitters (Node B, C, D or E).

5.4.2 ACK Loss Rate

A successful arrival of an ACK requires a correct reception of the corresponding packet at the recipient and a safe journey of the ACK back to the sender. The ACK Loss Rate (ALR) is the percentage of the packets sent out without the corresponding ACKs received by the sender. In practice, an ACK is considered lost after a certain amount of time (at timeout). The value of the timeout should be set to a value that is much larger than the average round trip time for a successful data transaction (from the point that a data packet is sent out to the time point that an ACK is received). In our experiments, we set the timeout value at one second (with the measured averange round trip time roughly 150ms).

The ACK Loss Rate is related to the transmission parameters chosen by the CR Strategy Reasoner of the sender, the number of transmitters in the communication environment and the MAC protocols used by the transmitters. Our analysis thus revolves around these elements.

Regarding the transmission parameters, we focus on the packet sending rate (how many

packets that are sent out per second). From the observations of the experiments, which will be presented below, we found that the packet sending rate is an essential parameter that decides the performance. For any other transmission parameters that may cause communication degradation (such as an excessive transmission power), the packet sending rate determines how frequently this degradation is caused. Since determining the exact cause of degradation of communication in a wireless network is an intensive task, we have designed our regulation to reduce or prevent further damages without taking time to analyze the fundamental reasons of the previous damages, and thus we focus on packet sending rate.

Based on this justification, all of our experiments revolve around the packet sending rate. We begin with an experiment examining the ACK Loss Rate versus various packet sending rates for one transmitter, and then discuss the ACK Loss Rate versus packet sending rates with multiple transmitters that use the same MAC protocol. Finally, we evaluate the ACK Loss Rate over a set of packet sending rates with multiple transmitters that use different MAC protocols.

One Transmitter

In the first experiment, only one transmitter is presented and sends data packets to the server (node A). We vary the packet sending rate from 1.25 to 10 packets/second. In addition, the PHY channel rate with GMSK was 32kb/s and the MAC frame size was 80 bytes. The corresponding ACK Loss Rate of CSMA is shown in Fig. 5.3(a), and the ACK Loss Rate of Aloha is shown in Fig. 5.3(b).

In Fig. 5.3, we observe that even with one transmitter, the performance can be very bad (high ACK Loss Rate when the packet sending rate is 10 packets/second), due to the fact that the packet sending rate is beyond the processing speed of the communication nodes, recall that for GNU Radios, packet processing is done on the personal computer via a USB connection. When the packet sending rate falls below the maximum processing ability of the nodes, the performance goes into a relatively steady state regime where the ACK Loss Rate does not monotonically decrease as we decrease the packet sending rate.

Another observation is that when the packet sending rate is high, Aloha MAC has better performance than CSMA MAC. This is because CSMA introduces additional overhead compared to Aloha. This overhead is associated with the time due to the carrier sensing and backoff. Compared with Aloha, this means less time is allowed for the practical packet

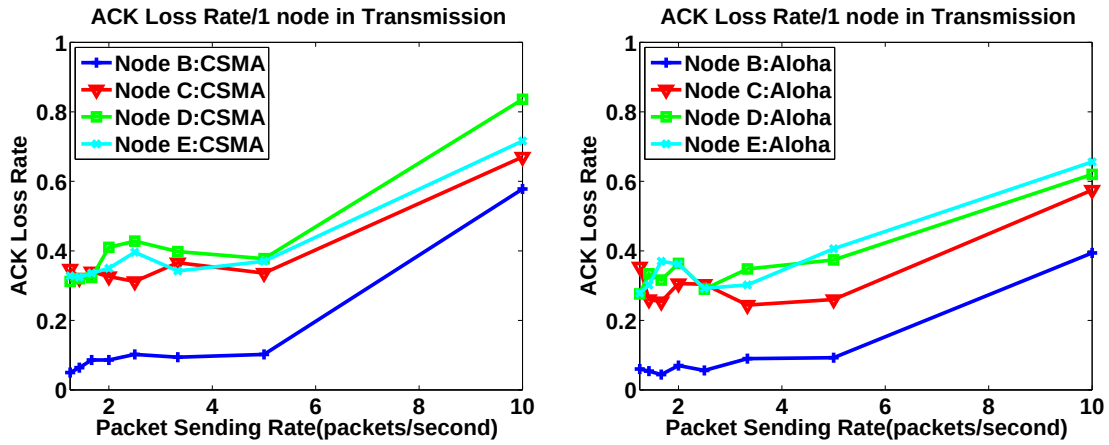


Figure 5.3: ACK Loss Rate with Increasing Packet Sending Rate. (a) One transmitter that follows CSMA, (b) One transmitter that follows Aloha.

transmissions. Therefore, the CSMA MAC only works well when the traffic load is light.

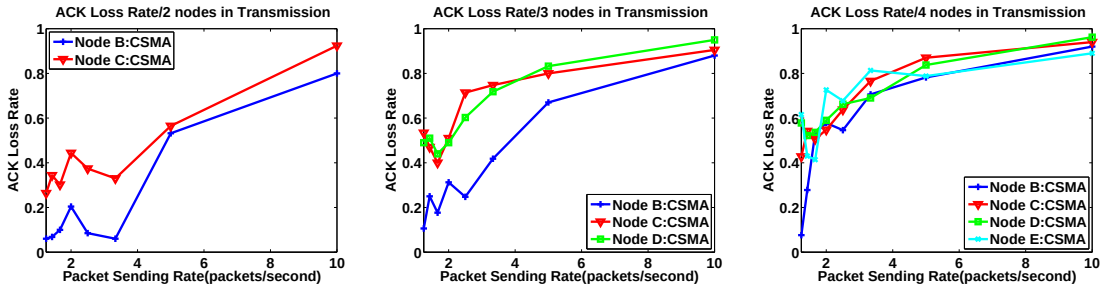
Further, we observe that for different communication links, the values for the ACK Loss Rate at the steady state can vary because the communications from these nodes are experiencing different link qualities.

The observations are helpful in designing the ORM. First of all, the ORM needs to consider the specific link quality and set a specific criterion of when a regulation shall be taken for that link. Further, because the performance at the steady state is statistically stationary, the estimation of the specific link quality can make use of the properties of the steady state.

Multiple Transmitters with the Same MAC

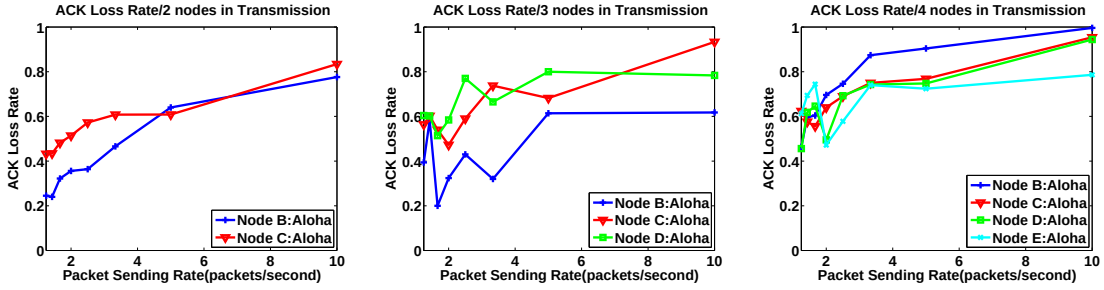
In this experiment, multiple transmitters (two to four transmitters) that use the same MAC protocols send packets to Node A. Fig. 5.4 shows the performance of the nodes that use CSMA, and Fig. 5.5 shows the performance of the nodes that use Aloha.

When more nodes are transmitting, under the same packet sending rate, we observe that the ACK Loss Rate becomes higher (the performance becomes worse). When the packet sending rate decreases to be low enough, different from in the one transmitter scenario, there are two cases. The first case can be observed at the packet sending rate at 3.33 packets/seconds in Fig. 5.4(a), the performance achieves to a steady state, i.e., the performance



(a) (b) (c)

Figure 5.4: ACK Loss Rate with Increasing Packet Sending Rate. (a) Two transmitters that follow CSMA, (b) Three transmitters that follow CSMA, and (a) Four transmitters that follow CSMA.



(a) (b) (c)

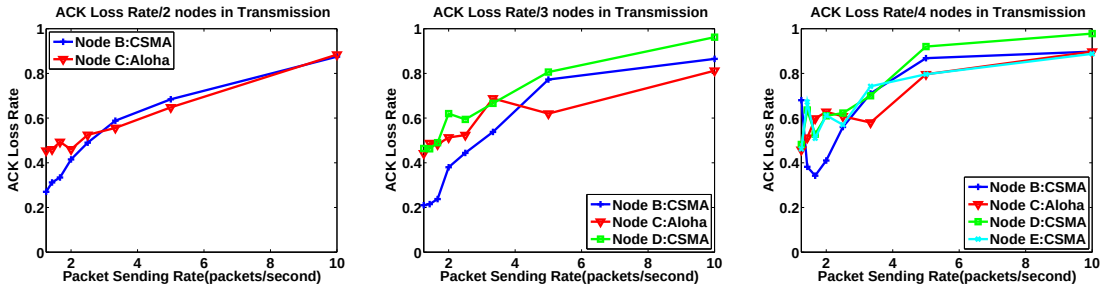
Figure 5.5: ACK Loss Rate with Increasing Packet Sending Rate. (a) Two transmitters that follow Aloha, (b) Three transmitters that follow Aloha, and (a) Four transmitters that follow Aloha.

does not further improve when the packet sending rate continues decreasing. The second case can be observed at the packet sending rate 3.33 packets/second in Fig. 5.5(a), the performance continues improving afterwards, but the improvement is very slight.

Since the performance of the nodes with CSMA or Aloha follows the same trend with only slight differences, the regulations of the ORM can be similar for different MACs, as long as the ORM captures the slight differences resulting from the specific MAC protocol used. In addition, when more nodes are used, the ORM should set a stricter criterion (lower packet sending rate).

Multiple Nodes with Different MACs

In this experiment, multiple nodes transmit simultaneously and use different MAC protocols. Among them, Node C uses Aloha while the other nodes use CSMA. The performance



(a) (b) (c)

Figure 5.6: ACK Loss Rate with Increasing Packet Sending Rate. (a) Two transmitters (Node B follows CSMA and Node C follows Aloha), (b) Three transmitters (Node B and D follow CSMA and Node C follows Aloha), and (a) Four transmitters (Node B, D, E follow CSMA and Node C follows Aloha).

is shown in Fig. 5.6. We observe that the performance still follows the same trend as if they used the same MAC (in Fig. 5.4 and Fig. 5.5) with only slight differences.

From the above observations, we can observe that the ACK Loss Rate turns out to be higher when the traffic condition becomes worse. Therefore, the value of ACK loss proves to be an effective indicator of the traffic conditions. The traffic condition can be improved by decreasing the packet sending ratio during bad traffic conditions. In other words, onboard regulation could be accomplished by forcing the CR strategy reasoner to reduce its packet sending rate.

5.5 Reactive Regulation Mechanism

Our onboard regulation is illustrated in Fig. 5.7. The ORM works in two modes on the transmission attempts of the CR Strategy Reasoner: the transparent mode and the punishment mode. The transparent mode (represented by a green light) is characterized by a set of behaviors that are suitable for the current traffic condition. The ORM considers the transmission strategy that is chosen by the CR Strategy Reasoner to be good and takes no regulation to these transmission attempts. On the other hand, if the traffic condition is bad (represented by a yellow light), the ORM considers that the transmission strategy chosen by the CR Strategy Reasoner is not good and applies punishment to its data transmissions. In our design, the punishment involves dropping a certain number of packets (which is equivalent to reducing the packet sending rate). If the traffic condition is extremly bad (represented by a red light), the ORM will drop all the packets.

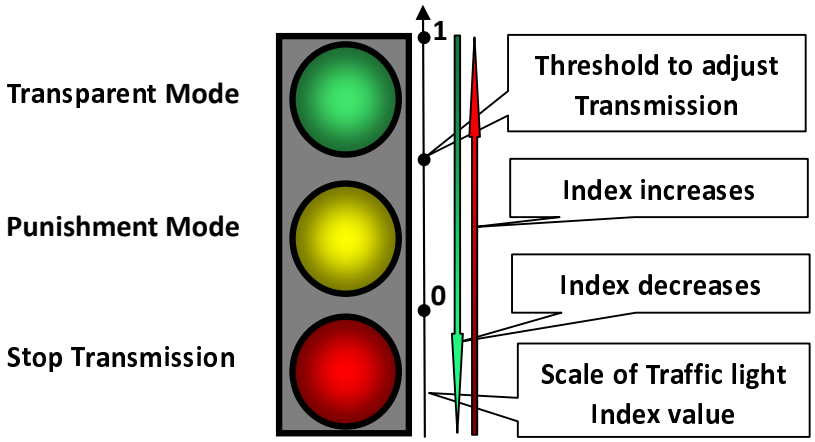


Figure 5.7: Traffic Light Strategy.

The traffic condition is represented by the Traffic Index I in a real-time manner, as shown in Fig. 5.7, with the index value ranging from 0 and 1. When the value is 1, it shows that the link is in a perfect condition, and when the value is 0, the traffic condition is considered extremely bad. The Traffic Threshold τ differentiates a good traffic condition from a bad traffic condition, and behaves as a criterion of whether a regulation shall be taken by the ORM. The ORM sets a proper value for the traffic threshold for each link. In this section, we discuss how to estimate the traffic threshould, how to update the traffic index and describe the detailed regulation method.

5.5.1 Estimation of Traffic Threshold

From the observations of the experimental results in Fig. 5.3, we already know that the performance (ACK Loss Rate) achieves a steady state when the packet sending rate becomes low enough. Since the value of the traffic threshold is correlated with the average performance that a transmitter can achieve in a given channel, we can estimate the traffic threshold by capturing the statistical average performance at the steady state.

According to the experimental results in Fig. 5.3, the performance at the steady state is different among links. Therefore, we do not consider the simple solution of setting a common value for the traffic threshold τ for every link. Instead, an estimating procedure of the traffic threshold is performed when a CR node initially begins its transmission in a channel.

The procedure begins with a decrease of the packet sending rate until a steady state

Algorithm 6 Initial Estimate of Traffic Threshold τ, its Standard Deviation $\sigma(\tau)$ and Traffic Index I

1: Set Packet Sending Ratio as 1
2: Send packets and get the ACK Loss Rate α
3: Keep halving Packet Sending Ratio and repeating Step 2, until the difference between subsequent α is smaller than ϵ
4: Fix this Packet Sending Ratio. Disregard previous results and repeat Step 2 a number of times
5: $\tau = 1 - mean(\alpha)$
6: $\sigma(\tau) = std(\alpha)$
7: $I = \tau$

performance is achieved. As the steady state of the performance is characterized by a stationary ACK Loss Rate, a further decrease of the packet sending rate after the steady state is achieved would not change the ACK Loss Rate significantly. Therefore, when the difference between the ACK Loss Rates at two different packet sending ratios is small, the ORM could regard a steady state is achieved. Next, the ORM obtains the statistical performance by making some measurements at this packet sending rate for which a steady state is believed to be in.

The detailed algorithm for estimating the Traffic Threshold τ and its standard deviation $\sigma(\tau)$ is shown in Algorithm 6. In this algorithm, the packet sending ratio (different from packet sending rate) represents the percentage of packets that are sent out to the channel by the ORM over the total packets that are passed from the upper layer to the MAC layer. In the first step, all the packets (say N packets in total with $N >> 0$) from the upper layer are sent out by the ORM and the corresponding ACK Loss Rate α_0 is summarized. Then the ORM decreases the packet sending ratio by sending out only half of the packets that are passed from the upper layer. For example, if the current packet sending ratio is 0.5, then the ORM randomly sends out half of the packets ($N/2$) that are passed from the upper layer. Again, the ACK Loss Rate α_1 in this step is computed. The ORM keeps halving the packet sending ratio until the difference between subsequent ACK Loss Rate $\alpha_i - \alpha_{i-1}$, $i \geq 1$ is smaller than a preset small positive value ϵ. Finally, the ORM fixes this packet sending ratio and repeats Step 2 and gets the mean and standard deviation of ACK Loss Rate, as shown in Step 5 and 6 in Algorithm 6.

ϵ is a preset value chosen beforehand. If ϵ is small, then it might take a long time to achieve the steady state. On the other hand, if ϵ is big, then the accuracy of computed threshold τ may be inaccurate. Because the value τ and $\sigma(\tau)$ would be updated during the regulation (we will discuss this in the next section), an inaccurate estimate resulting from

an inapproper chosen value ϵ would be revised gradually.

The initial value of traffic index I is set equal to τ, so that a greedy and malicious behavior would result in an immediate deviation of traffic index I from the threshold τ. Consequently, a reative regulation would be performed immediately.

5.5.2 Realtime Update of Traffic Index and Traffic Threshold

As we have mentioned, we intend to using the feedback from the recipient (ACKs) to analyze the traffic condition. Since the traffic index is a time-varying value that represents the real-time traffic condition, the method whereby the traffic index is updated with every computated ACK Loss Rate from a number of ACKs is not used since the change of the traffic index lags behind the individual ACKs, which results in the change of the traffic index. Our alternative is to update the traffic index with every received ACK and timeout.

As the real-time traffic condition (represented by the traffic index) is highly correlated with the ACK Loss Rate, we use an Exponential Weighted Moving Average(EWMA) method [94] to update the traffic Index. Suppose the traffic index value I at the moment t is $I(t)$ and the first feedback f since the moment t arrives at the moment $t + \delta t$, which is formulated as $f(t + \delta t)$. The value of traffic index is updated immediately with this feedback f with a formula $I(t + \delta t) = I(t) + (1 - \lambda_1)(f(t + \delta t) - I(t))$. In this equation, λ_1 is a forgetting factor, with a constant value that is slightly smaller than 1. The value of $f(t + \delta t)$ is 1 if it receives an ACK and 0 if it there is a timeout. In the long run, the value of the traffic index will fluctuate around the average value of the ACK Loss Rate.

Due to reasons, such as ϵ is not properly chosen or a second transmitter is present, the initial estimates of the traffic threshold τ and the standard deviation of the traffix threshold $\sigma(\tau)$ may not be accurate. The bias of the estimates of both the traffic threshold τ and the standard deviation of the traffix threshold $\sigma(\tau)$ can be fixed with a similar method as the update of the traffic index.

Still, we choose to use the Exponential Weighted Moving Average(EWMA) method. From the observations in Sec. 5.4, we know that the ACK Loss Rate does not greatly improve after a steady state is achieved. Because the traffic threshold is directly related to the ACK Loss Rate at the steady state, if the traffic index i is much bigger than the traffic threshold τ, it is abnormal and shows the initial estimate of traffic threshold is inaccurate.

Algorithm 7 Indexes Update

Set a timer for every packet sent out
if Time Out **then**
  $I = I + (1 - \lambda_1)(0 - I)$, where $0 < \lambda_1 < 1$
end if
if An ACK received **then**
  $I = I + (1 - \lambda_1)(1 - I)$
  Disable the corresponding timer
  if $I > \tau + \sigma(\tau)$ **then**
    $\tau = \tau + (1 - \lambda_2)(I - \tau)$, where $0 < \lambda_2 < 1$
    $\sigma^2(\tau) = \sigma^2(\tau) + (1 - \lambda_2)\left((\tau - I)^2 - \sigma^2(\tau)\right)$
  end if
end if

Algorithm 8 Onboard Regulation Algorithm

if $I \geq \tau$ **then**
  Send this packet out
else
  Drop this packet with probability $\left\lfloor a\frac{\tau - I}{\tau}, 1\right\rfloor$, where $a \geq 1$
end if

However, if the traffic index i is smaller than the traffic threshold τ, it only means that the traffic condition has become worse and a regulation shall be taken. In this case, the traffic threshold τ shall not not updated

The detailed algorithm is presented in Algorithm 7. In our algorithm, the traffic threshold and its standard deviation are updated when $I > \tau + \sigma(\tau)$. Suppose the traffic index is normal distributed and a reasonable transmission strategy is used, the traffic index has less than 18% probability that is bigger than $I > \tau + \sigma(\tau)$.

The forgetting factor that is used to update the traffic threshold and its standard deviation is λ_2, which also has a positive value slightly smaller than 1. It is recommended $\lambda_2 > \lambda_1$. In this way, the value of the traffic index changes faster with every feedback so as to capture the real-time change than the traffic threshold and its standard deviation, which represent the long term statistics of the traffic index.

5.5.3 Local Regulation

The onboard regulation is taken when the value of the traffic index I, which represents the real-time performance, is smaller than the value of the traffic threshold τ, which represents the performance in the steady state. The detailed algorithm is shown in Algorithm 8.

If the traffic index is bigger than or equal to the traffic threshold, the ORM would

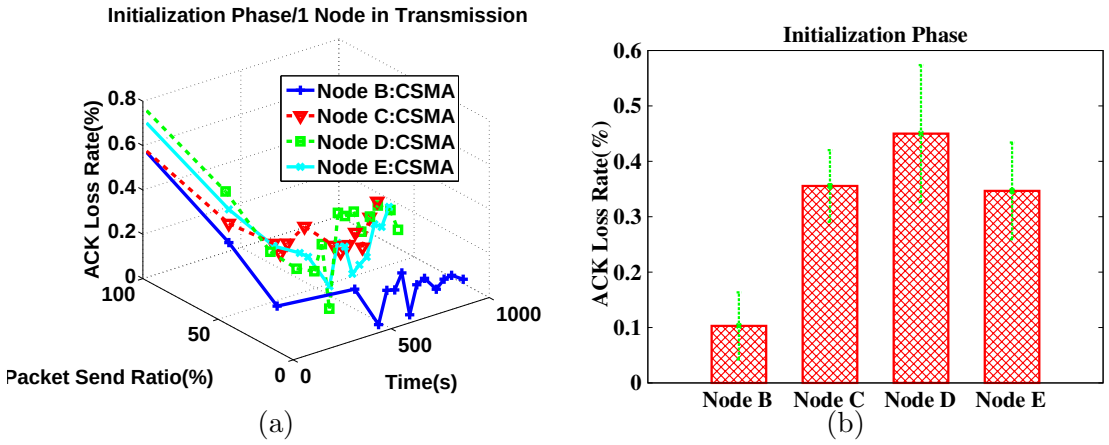


Figure 5.8: Initialization Phase with One Transmitter, (a) ACK Loss Rate evolves with time, (b) The average and the standard deviation of ACK Loss Rate.

consider the transmission strategy that is chosen by the CR Strategy Reasoner to be a proper choice for the current traffic condition. Therefore, it works in the transparent mode and simply sends this packet out, as shown in Step 2 in Algorithm 8.

When the ORM works in the punishment mode, the punishment is performed according to the real-time traffic condition, i.e. how worse the real-time performance (represented by the traffic index I) than the steady state performance (represented by the traffic threshold τ). In Step 4, the ORM drops the packet with a probability $\lfloor a\frac{\tau-I}{\tau}, 1\rfloor$, where $a \geq 1$. When $a = 1$, $0 \leq \frac{\tau-I}{\tau} \leq 1$. If $\tau = I$, then there is no punishment. If $I = 0$, then all the ACKs are lost and the later transmissions are prohibited. A value of a that is larger than 1 performs the adaptive regulation more effective.

5.6 Evaluation

In this section, we will evaluate the performance of our proposed algorithms with the experimental setup defined in Fig. 5.2 on the ORBIT testbed. The purpose of our experiments is to verify whether an initial estimate of the traffic threshold could be accurately made, whether the traffic index could track the traffic condition in a real-time manner, and whether our proposed regulation can enhance effective communications.

5.6.1 Initial Estimate of the Traffic Threshold

The estimate of the traffic threshold for each of the four transmission links (Node B and A, Node C and A, Node D and A, Node E and A) is made in the ORBIT testbed. The result of the estimation process is illustrated in Fig. 5.8. Both the changes of the ACK Loss Rate and the packet sending ratio related to time are recorded in Fig. 5.8(a). As discussed in the Algorithm 6, the packet sending ratio keeps halving until the difference between the subsequent ACK Loss Rate (a statistics with 100 packets sent out) is smaller than 0.05 (a chosen value for ϵ). We then fix that packet sending ratio and recalculate the ACK Loss Rate over ten iterations.

The means and the standard deviations of the ACK Loss Rate are shown in Fig. 5.8(b). After a comparison of these values with the ACK Loss Rates in the steady state in Fig. 5.3(a), we find that these estimates are rather accurate. As shown in Algorithm 6, the estimate of the traffic threshold is 1 minus the mean of the ACK Loss Rate and the estimate of the standard deviation of the traffic threshold is simply the standard deviation of the ACK Loss Rate.

We note that this estimation process takes a long time (less than 1000 seconds) in our experiment. This is because of the low processing and communication speed of the GNU Radio, where there was only a maximum packet sending rate of a few bytes per second. For a commercially-used transmitter on the market, because of a better processing speed, the intial estimation would need less time. For example, if a node can transmit a few hundreds packets per second, then the time for this initial estimate takes only a few seconds.

5.6.2 Index Update for a Single Transmitter

In this experiment, we verify the update of traffic index and traffic threshold when there is only one transmitter. The experiment includes three parts with the results shown in Fig. 5.9. In addition, we use a as 2 in Algorithm 8.

In the first part, Node B transmits with a proper packet sending rate (5 packets/second) and begins with an accurate estimate of the traffic threshold. The real-time update of the traffic index I, the traffic threshold τ, the standard deviation of the traffic threshold $\sigma(\tau)$ are shown in Fig. 5.9(a)(b). The threshold τ and its standard deviation $\sigma(\tau)$ are fixed during the test and the traffix index I fluctuates around the traffic threshold τ. This shows that

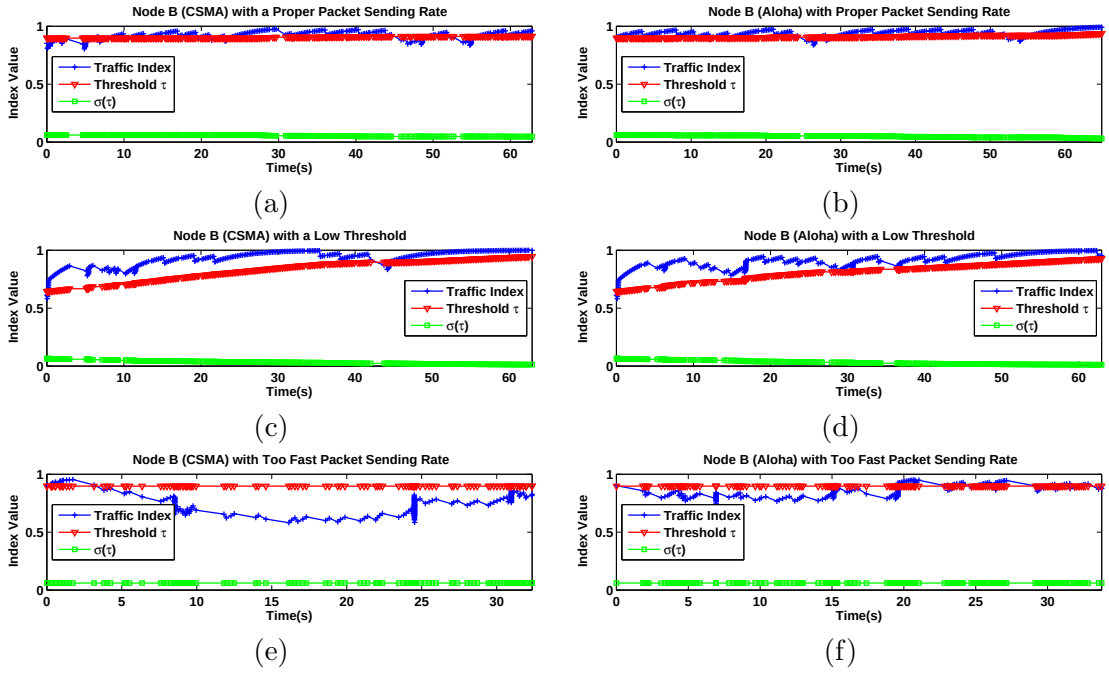


Figure 5.9: Traffic Index versus Traffic Index Threshold τ and its Standard Deviation $\sigma(\tau)$. Node B transmitting without Interference. (a) Node B transmits with a proper Packet Sending Rate and follows CSMA, (b) Node B transmits with a proper Packet Sending Rate and follows Aloha, (c) Node B transmits with a proper Packet Sending Rate and follows CSMA, but begins with a low Traffic Index Threshold τ, (d) Node B transmits with a proper Packet Sending Rate and follows Aloha, but begins with a low Traffic Index Threshold τ (e) Node B transmits with too fast Packet Sending Rate and follows CSMA, (f) Node B transmits with too fast Packet Sending Rate and follows Aloha.

the ORM works in the transparent mode and does not take regulation when the cognitive radio works properly.

In the second part, the intial value of the traffic threshold is set below its proper value and Node B still uses a proper packet sending rate. We observe that the traffic threshold τ increases to its proper value in in Fig. 5.9(c)(d). Even if the estimation of the traffic threshold is not accurate (due to an improper choice of ϵ or collisions), our algorithm still works well.

In the third part, the traffic threshold is set correctly, but the transmitter uses a packet sending rate faster than the communication parties could handle. In Fig. 5.9(e)(f), we observe that the traffic index is below the traffic threshold (which indicates a regulation and packet dropping). Also, we observe that the traffic index in Fig. 5.9(e) is lower than that in Fig. 5.9(f). It shows that the Aloha MAC has better performance over the CSMA

MAC when the traffic load is heavy.

5.6.3 Index Update with an Interruption of Burst Traffic

As we mentioned, the cognitive radio network is characterized by dynamic traffic conditions. In this experiment, we evaluate the performance of onboard regulations with bursty traffic interference. To be more specific, we would like to know whether a regulation is performed when a bursty traffic interference is introduced and whether the regulation stops when the bursty traffic interference is gone.

The experiment setup is as follows. The first transmitter (Node B) uses a proper packet sending rate (5 packets/second) and a correct traffic threshold (0.89). In the middle of its transmission, a second transmitter (Node D) begins its transmission for approximately 20 seconds.

To get a comprehensive study of the multi-MAC cognitive radio environment, this experiment includes four parts with variously-chosen MAC protocols. In the first part, both Node B and D use CSMA MAC. In the second part, both Node B and D use Aloha MAC. In the third part, Node B uses CSMA and Node D uses Aloha. In the last part, Node B uses Aloha and Node D uses CSMA.

We record the index changes in Fig. 5.10. In all the parts of the experiment, the traffic index drops below the traffic threshold shortly after the appearance of the burst traffic. This indicates a regulation and a responsive action of the onboard regulation towards a burst traffic interference. Further, after the burst traffic is gone, the traffic index gradually increases to the same level of the traffic threshold (in the transparent mode). It shows that the transmitter could recover to a normal transmission level after the traffic condition becomes good.

5.6.4 Index Update for Two Transmitters

In the cognitive radio network, to ensure a fair transmission, when two or more nodes have transmission attempts, all of them shall reduce their own transmissions to allow some time for other nodes. In this experiment set, we explore the onboard regulative behaviors when two nodes transmit simultaneously.

Both nodes (Node B and D) send packets to Node A. They both choose a proper packet

sending rate that is only suitable for the case of only one transmitter (5 packets/second). In addition, they both begin with a correctly set traffic threshold (roughly 0.89 for Node B and 0.55 for Node D).

Similar to the burst traffic experiment set, this experiment set includes three parts, where the performance is shown in Fig. 5.11. Fig. 5.11(a)(b) records the index update when both nodes use CSMA. The performance of the case when both nodes use Aloha is shown in Fig. 5.11(c)(d). In Fig. 5.11(e)(f), Node B uses CSMA and Node D uses Aloha.

In all the three experiments, we observe that the traffic indexes go below the traffic threshold for the most of the time. This shows that the ORMs in both Node B and D are in the punishment mode and taking regulations on their own transmission.

We observe a slight difference between the regulations. The traffic indexes in Fig. 5.11(a)(b) are obviously smaller than those in Fig. 5.11(c)(d). Since a low traffic index means that more packets are dropped, this justifies that Aloha MAC has better performance in a heavy traffic load condition.

5.6.5 ACK Loss Rate

Expression	*Meaning*
$C(P)$	Node B with proper Packet Sending Rate and CSMA
$C(F)$	Node B with too fast Packet Sending Rate and CSMA
$A(P)$	Node B with proper Packet Sending Rate and Aloha
$A(F)$	Node B with proper Packet Sending Rate and Aloha
$CC(B)$	Node B (Both Node B and D with CSMA)
$CC(D)$	Node D (Both Node B and D with CSMA)
$AA(B)$	Node B (Both Node B and D with Aloha)
$AA(D)$	Node D (Both Node B and D with Aloha)
$CA(B)$	Node B (Node B with CSMA and Node D with Aloha)
$CA(D)$	Node D (Node B with CSMA and Node D with Aloha)
$AC(B)$	Node B (Node B with Aloha and Node D with CSMA)
$AC(D)$	Node D (Node B with Aloha and Node D with CSMA)

Table 5.1: Indexes

In the above experiments, by reading the traffic index update, we can observe that the ORM takes actions when the traffic condition is not good. In this experiment set, we explore the resulting performance due to the regulations by analyzing three numerical metrics.

The first metric is the number of packets that are sent out. Because some packets are dropped in the punishment mode, the number of packets that are sent out indicates the severity of the punishment being performed. For a given number of packet transmission

attempts, the lower the number of packets sent out, the harsher the punishment is.

The second metric is the number of ACKs that are received by the transmitter. As we have mentioned, a successfully receipt of an ACK indicates an effective data transaction, the number of ACKs received shows the effectiveness (throughput) of the data transactions.

The third metric is ACK Loss Rate. It is the ratio of the number of ACKs received over the number of packets actually sent out, which indicates the actual transmission efficiency. Since cognitive radios can be resource-limited systems, efficient transmissions should make good use of their limited battery.

We performed eight pairs of experiments. In each experiment, the higher layer (UDP protocol) passes 100 packets to the MAC layer. Every pair of experiments have the same experiment setup, except that one of them has the ORM and the other doesn't. We explain the context of these experiments in Table.5.1.

The performance is shown in Fig. 5.12. In the figure, the red solid columns show the performance of nodes without regulation and the green dotted columns show the performance of nodes with regulation. The lengths of lines crossing the top of columns represent the values of the standard deviations.

Fig. 5.12(a) shows the number of packet sent out. Because the MAC layers without regulations convey every packet that is passed from its upper layer to the recipient, the number of packets that are sent out is 100. The behaviors are quite different for nodes with regulation. First of all, if there is only one transmitter with proper packet sending rate (the cases of C(P) and A(P)), the number of packets sent out approaches 100, which indicates that the ORM seldom applies regulation. Secondly, if there is only one transmitter with too fast of a packet sending rate (the cases of C(F) and A(F)), about twenty percent of packets are dropped due to the regulation. When there are two transmitters (the cases of CC, AA, CA and AC), more packets are dropped than in one transmitter case.

Fig. 5.12(b) shows the number of ACKs received. In the cases of one transmitter with a proper packet sending rate (C(P) and A(P)), the number of ACKs received by nodes with regulations is closely approaching that by nodes without regulations. In the cases of one transmitter with too fast of a packet sending rate (C(F) and A(F)), the performance of the node with regulation is obviously better than that of the node without regulation. In the cases of two transmitters (CC, AA, CA, and AC), we observe that the number of ACKs received are generally similar for both cases (with and without regulations).

The ACK Loss Rate for data transmission is shown in Fig. 5.12(c). Clearly, for most of the cases, the ACK Loss Rate for the node with regulation is smaller than that without regulation. This shows that a better transmission efficiency can be achieved when the ORM is used.

5.7 Conclusions

In this chapter, we have proposed a reactive onboard regulative mechanism to ensure fair and effective transmission in the cognitive radio network. In this mechanism, an onboard regulative module is embedded into the platform of a cognitive radio node, which monitors the traffic condition in a particular channel and takes regulations when excessive packets are considered to be injected into the networks. Specifically, after the analysis of the ACK Loss Rate for GNU Radio nodes with CSMA or Aloha MAC in the ORBIT testbed at various packet sending rates, we proposed an adaptive onboard regulative algorithm, that is able to estimate the average performance for a specific link and take regulatory action according to the difference between the real-time performance and the observed average performance. We evaluated our algorithms on the ORBIT testbed through extensive experiments. The results showed that our proposed algorithms achieve a more effective transmission than a node without regulation.

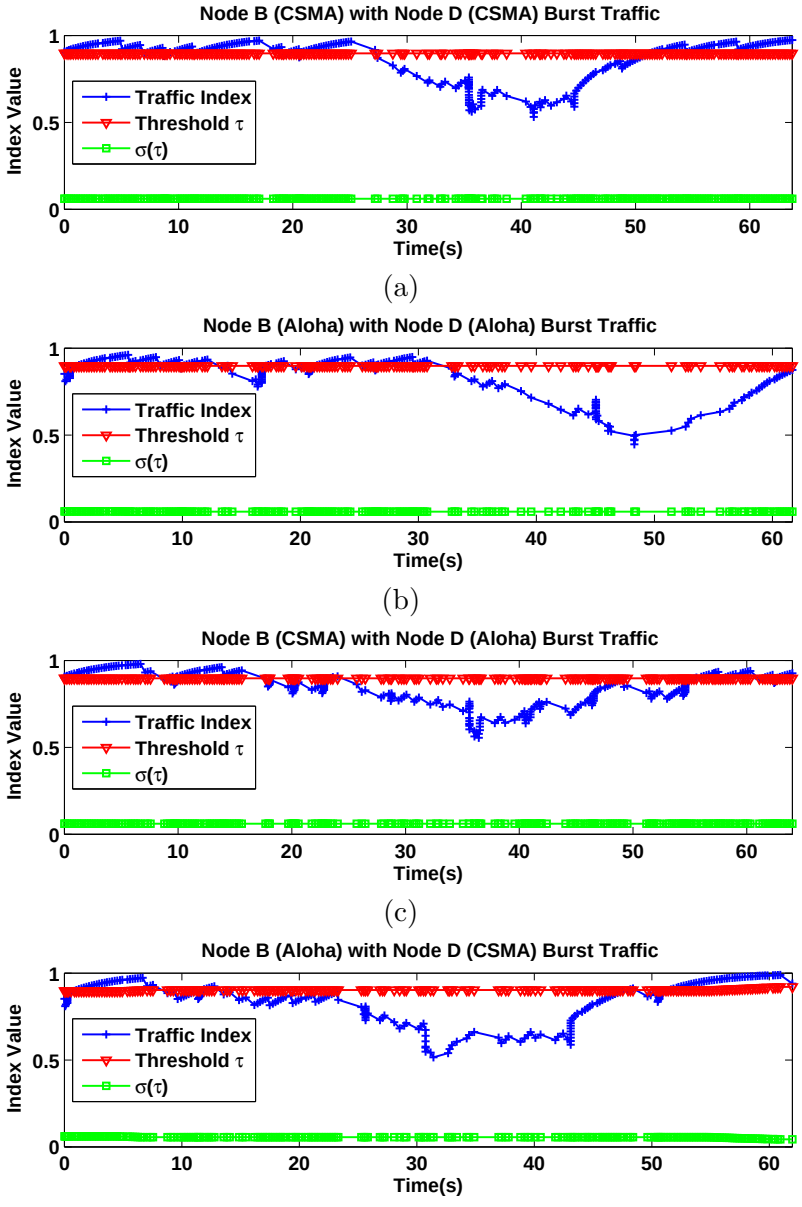


Figure 5.10: Traffic Index versus Traffic Index Threshold τ and its Standard Deviation $\sigma(\tau)$. Node B transmitting with Burst Traffic Interference from Node D. (a) Both Node B and D follow CSMA, (b) Both Node B and D follow Aloha, (c) Node B follows CSMA and Node D follows Aloha, (d) Node B follows Aloha and Node D follows CSMA.

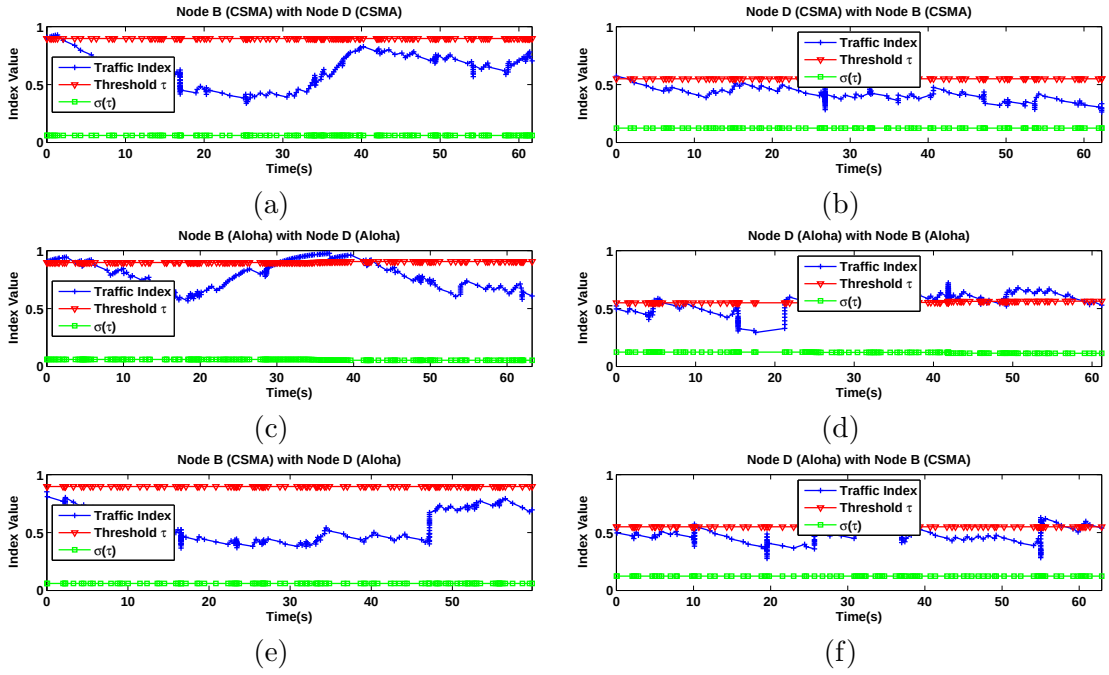
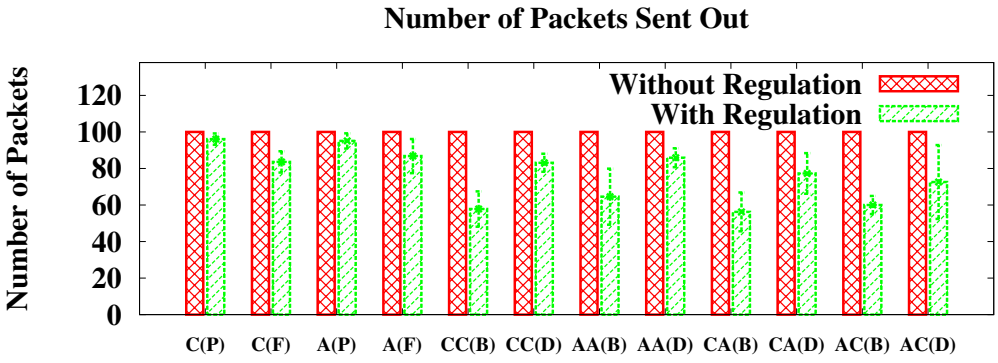
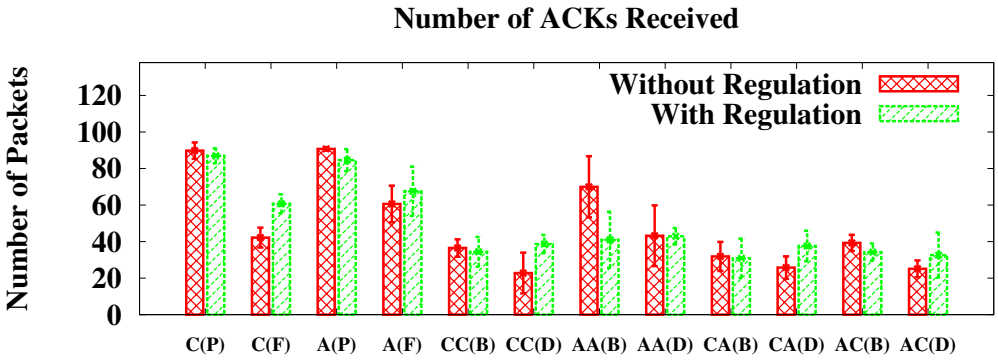


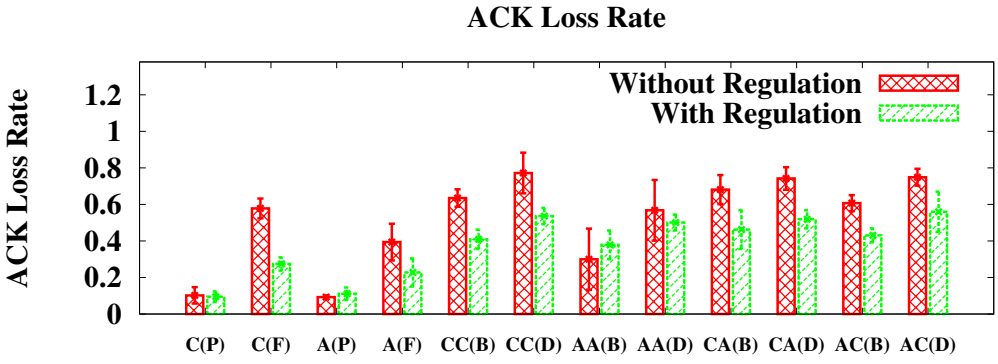
Figure 5.11: Traffic Index versus Traffic Index Threshold τ and its Standard Deviation $\sigma(\tau)$. Both Node B and D transmitting. (a) Node B indexes change (both Node B and D follow CSMA), (b) Node D indexes change (both Node B and D follow CSMA), (c) Node B indexes change (both Node B and D follow Aloha), (d) Node D indexes change (both Node B and D follow Aloha), (e) Node B indexes change (Node B follows CSMA and Node D follows Aloha), (f) Node D indexes change (Node B follows CSMA and Node D follows Aloha).



(a)



(b)



(c)

Figure 5.12: Performance with and without Regulation (refer to Tab.5.1 for the detailed index meanings of the horizontal axis). (a) Number of packets sent out, (b) Number of ACKs received at the sender, (d) ACK Loss Rate.

Chapter 6
Conlusion and Future Work

6.1 Conclusion

In this thesis, we discussed power modulated challenge-response to verify the claimed locations of wireless nodes, adaptive delay-sensitive location-oriented content delivery, facilitating active transmit-only RFID systems through receiver-based processing, and reactive on-board regulation of cognitive radios based on link quality estimation.

To ensure the secure verification in location-based systems, we have proposed the technique of modulating the transmission powers in a challenge-response mechanism to verify the truthfulness of an entity's claimed location. We evaluated the effectiveness under naive and smart adversarial models of three presented strategies, which are direct PMCR, indirect PMCR and signal strength PMCR. In particular, we looked at the probability of falsely declaring a claimant is at a valid position for these three schemes versus the distance between the true and claimed position of the claimant. Additionally, we also showed that these methods are susceptible to collusion of multiple adversaries in the presence of naive and smart colluders. We also proposed the use of rotational directional power-modulated challenge response, where directional antennas are used to defend against collusion attacks.

Secondly, we examined the efficacy of an on-demand multimedia service in location-based systems where mobile nodes access data according to their locations. Given limited number of access points and an abundance of service requests that result from the nodes moving around, the transmission time is not negligible and will introduce tremendous delays when the system supports many users simultaneously. We used the movement pattern of a mobile node, which is modeled as a semi-Markov process, and formulated a criterion for optimizing the service strategy. We first proposed an improved multicast strategy for an AP-centric method, where all the data is transmitted by APs, and then we presented the Deputy&Forward method, in which nodes who have previously received location-based

data can assist the system by serving nodes that newly arrive at the location. We discuss two Deputy&Forward methods, single channel and multiple channel Deputy&Forward and analyze their serving strategies. Based on simulation studies, we have shown that the improved AP-centric method has better performance than three baseline strategies and that the Deputy&Forward method can achieve better latency and throughput than the AP-centric method.

Thirdly, we have proposed new methods to improve the tracking efficacy of receiverless transmit-only RFID tags in RFID systems by utilizing an enhanced form of multiuser detection at the receiver that can identify overlapping tag signals. We have developed an improved successive cancelation algorithm to statistically estimate signal amplitude and transmission time that exploits the properties of our tag system. Our successive cancelation method has a better performance than the traditional successive cancelation method. Further, we have balanced the computational load across the entire system for improved scalability by making use of soft tag handoff between tag readers, and updating local tag lists at each reader. In addition, we proposed a new overlap-reduced successive cancelation method to further reduce the intensive computation and memory costs associated with successive cancelation. Finally, we evaluated the performance under different collision situations, where varying levels of near far effects, and noise were examined in simulations and showed a better detection ratio in a realistic inventory-monitoring scenario.

Lastly, we have proposed a reactive, onboard regulative mechanism to ensure secure and effective transmission in cognitive radio systems. The reactive onboard regulative mechanism is performed by embedding an onboard regulative module into the platform of a cognitive radio node, to monitor the traffic condition in a channel and applies regulation when excessive packets are considered to be injected into the network. We analyzed the ACK Loss Rate of GNU Radio nodes with CSMA or Aloha MAC on the ORBIT testbed at various packet sending rates, and proposed an adaptive onboard regulative algorithm, that is able to estimate the average performance for a specific link and apply regulation according to the difference between the real-time performance and the observed average performance. We evaluated our algorithms on the ORBIT testbed and achieved better transmission efficiency than that of a node without regulation.

6.2 Future Work

We have evaluated the performance of our proposed methods under real system settings. In the future, we would like to re-evaluate our proposed research methodology in more practical systems and applications. First of all, a practical location-based system needs to be implemented on a WiFi or a cellular network, so that our power-modulated challenge response techniques and adaptive location-oriented content delivery could be verified in a real system.

As to the power modulation, we plan to put the data that are collected from the system into our security analysis. In particular, we would like to examine the formulation of the path loss and fading model and the modulation of the transmission powers of the APs. Further, the probabilities of false positive and false negative would be recalculated, which would be compared to the results in our thesis.

Similarly, since our improved multicast strategy makes use of the movement pattern of users in a location-based system, it is important to make an accurate description of the moving patterns. Moreover, we note that our evaluation is performed with a single layout (one AP and nine locations), different scenarios should be considered to get a comprehensive evaluation. Additionally, WINLAB has produced practical transmit-only low-cost tags and we would like to implement our detection algorithms into a real RFID reader.

In this thesis, we introduced three wireless systems (location-based systems, RFID systems and cognitive radio systems) and focused on improving the efficacy and security of these systems. We believe that more problems need to be addressed in these three systems. In particular, our onboard regulation currently is used to address problems associated with a node injecting too many packets into a network. To address other vulnerabilities in a cognitive radio network, we would extend our onboard regulation techniques and evaluate these algorithms on the ORBIT testbed. In addition, we would evaluate other forms of feeback besides the ACK Loss Rate, and use these to formulate a more powerful onboard regulation module that can address a broader variety of threats facing a cognitive radio network.

References

[1] E. Bertino, B. Catania, M. Damiani, and P. Perlasca. Geo-rbac: a spatially aware rbac. In *Proceedings of the Tenth ACM Symposium on Access Control Models and Technologies(SACMAT 05)*, 2005.

[2] S. Chen, Y. Zhang, and W. Trappe. Inverting sensor networks and actuating the environment for spatio-temporal access control. In *The Fourth ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2006)*, 2006.

[3] K. Langendoen and N. Reijers. Distributed localization in wireless sensor networks: a quantitative comparison. *Comput. Networks*, 43(4):499–518, 2003.

[4] N. Priyantha, A. Chakraborty, and H. Balakrishnan. The CRICKET location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking (Mobicom 2000)*, pages 32–43, 2000.

[5] D. Nicelescu and B. Nath. Ad hoc positioning (APS) using AOA. In *Proceedings of IEEE Infocom 2003*, pages 1734 – 1743, 2003.

[6] Z. Li, W. Trappe, Y. Zhang, and B. Nath. Robust statistical methods for securing wireless localization in sensor networks. In *Proceedints of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pages 91–98, 2005.

[7] S. Capkun and J.P. Hubaux. Secure positioning in sensor networks. Technical report EPFL/IC/200444, May 2004.

[8] S. Capkun and J. P. Hubaux. Secure positioning of wireless devices with application to sensor networks. In *Proceedints of IEEE INFOCOM 2005*, 2005.

[9] L. Lazos and R. Poovendran. SeRLoc: Secure range-independent localization for wireless sensor networks. In *Proceedings of the 2004 ACM Workshop on Wireless Security*, pages 21–30, 2004.

[10] L. Lazos, R. Poovendran, and S. Capkun. Rope: robust position estimation in wireless sensor networks. In *Proceedints of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pages 324–331, 2005.

[11] Patrick Chiu Paul Castro and etc. A probabilistic room location service for wireless networked environments. In *Ubicomp*, 2001.

[12] A. Udaya Shankar Moustafa A. Youssef, Ashok Agrawala. Wlan location determination via clustering and probability distributions. In *IEEE International Conference on Pervasive Computing and Communications*, 2003.

[13] William Stallings. *Network Security Essentials*. 2002.

[14] W. Xu, W. Trappe, Y. Zhang, and T. Wood. The feasibility of launching and detecting jamming attacks in wireless networks. In *MobiHoc '05: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, pages 46–57, New York, NY, USA, 2005. ACM Press.

[15] D. Niculescu and B. Nath. Trajectory based forwarding and its applications. In *Proceedings of Mobicom '03*, pages 260–272, 2003.

[16] S. Capkun and M. Cagalj. Integrity regions: Authentication through presence in wireless networks. In *Proceedings of the 2006 ACM workshop on Wireless security*, 2006.

[17] N. Sastry, U. Shankar, and D. Wagner. Secure verification of location claims. In *Proceedings of the 2003 ACM workshop on Wireless security*, pages 1–10, 2003.

[18] S. Brands and D. Chaum. Distance bounding protocols. In *Workshop on the theory and application of cryptographic techniques on Advances in cryptology*, 1994.

[19] A. Goldsmith. *Wireless Communications*. Cambridge University Press, Stanford University, 2004.

[20] T.S. Rappaport. *Wireless Communications- Principles and Practice*. Prentice Hall, 2001.

[21] W.D. Rummler. More on the multipath fading channel model. *IEEE Trans. Commun.*, 29:346–352, 1981.

[22] S. S. Ghassemzadeh, R. Jana, W. Rice, W. Turin, and V. Tarokh. Measurement and modeling of an indoor UWB channel. *IEEE Trans. Commun.*, 52:1786–1796, 2004.

[23] C. Kaufman, R. Perlman, and M. Speciner. *Network Security: Private Communication in a Public World*. Prentice Hall, 1995.

[24] S. Northcutt. *Network Intrusion Detection: An Analyst's Handbook*. New Riders, 1999.

[25] W. Trappe and L.C. Washington. *Introduction to Cryptography with Coding Theory*. Prentice Hall, 2002.

[26] Joseph J. Carr. *Practical Antenna Handbook (Paperback)*. McGraw-Hill/TAB Electronics; 4 edition, 2001.

[27] P. Bahl and V.N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of IEEE Infocom 2003*, pages 775–784, 2000.

[28] P. Bahl, V.N. Padmanabhan, and A. Balachandran. Enhancements to the RADAR User Location and Tracking System. Technical Report Technical Report MSR-TR-2000-12, Microsoft Research, February 2000.

[29] P. Enge and P. Misra. *Global Positioning System: Signals, Measurements and Performance*. Ganga-Jamuna Pr, 2001.

[30] Shang Y., Ruml W., and Zhang Y. Localization from mere connectivity. In *Proceedings of the Fourth ACM International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc)*, 2003.

[31] D. Nicelescu and B. Nath. DV based positioning in ad hoc networks. *Telecommunication Systems*, 22(1-4):267–280, 2003.

[32] F. Anjum. Location dependent key management using random key-predistribution in sensor networks. In *Proceedings of the 2006 ACM workshop on Wireless security*, 2006.

[33] D. Liu, P. Ning, and W. Du. Attack-resistant location estimation in sensor networks. In *Proceedints of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, 2005.

[34] Y. Zhang, Z. Li, and W. Trappe. Power-modulated challenge-response schemes for verifying location claims. In *Globecom '07: Proceedings of the IEEE Global Telecommunications Conference*, 2007.

[35] D. Liu and P. Ning. Location-based pairwise key establishments for static sensor networks. In *The First ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2003)*, 2003.

[36] L. Aalto, N. Gthlin, J. Korhonen, and T. Ojala. Bluetooth and wap push based location-aware mobile advertising system. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services MobiSys '04*, 2004.

[37] J. Ortiz, C. Baker, D. Moon, R. Fonseca, and I. Stoica. Beacon location service: a location service for point-to-point routing in wireless sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks IPSN '07*, 2007.

[38] J. Lil, A. Kubota, and H. Kameda. Location management for pcs networks with consideration of mobility patterns. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, 2005.

[39] J. Lin and S. Paul. Rmtp: Reliable multicast transport protocol. In *Proceedings of IEEE INFOCOM '96*, 1996.

[40] R. Bhatia and L. Li. Characterizing achievable multicast rates in multi-hop wireless networks. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing MobiHoc '05*, 2003.

[41] P. Chaporkar, A. Bhat, and S. Sarkar. An adaptive strategy for maximizing throughput in mac layer wireless multicast. In *Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing MobiHoc '04*, 2004.

[42] Q. Cao1, T. Abdelzaher1, T. He2, and R. Kravets1. Cluster-based forwarding for reliable end-to-end delivery in wireless sensor networks. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, 2007.

[43] W. Yuen, R. Yates, and S. Mau. Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks. In *Proceedings of IEEE INFOCOM '03*, 2003.

[44] U. Kubach and K. Rothermel. Exploiting location information for infostation-based hoarding. In *Proceedings of the 7th annual international conference on Mobile computing and networking MobiCom '01*, 2001.

[45] T. Small and Z. Haas. The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way). In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking and computing MobiHoc '03*, 2003.

[46] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 2003.

[47] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, 2004.

[48] A. Kate, G. Zaverucha, and U. Hengartner. Anonymity and security in delay tolerant networks. In *Proceedings of the 3rd International Conference on Security and Privacy in Communication Networks (SecureComm 2007)*, 2007.

[49] J. Proakis. *Digital Communications(4e)*. McGraw-Hill Companies, Inc, 2001.

[50] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Hall, 2005.

[51] R. Howard. *Dynamic Probabilistic Systems*. John Wiley & Sons, Inc, 1971.

[52] R. Yates and D. Goodman. *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*. John Wiley & Sons, Inc, 2003.

[53] A. Papoulis and S. Pillai. *Probability, Random Variables, and Stochastic Processes*. Mc Graw Hill, 2002.

[54] T. Cormen, C. Leiserson, and R. Rivest. *Introduction to Algorithms*. Mit Press, 2001.

[55] A. Juels. Rfid security and privacy: a research survey. *Selected Areas in Communications, IEEE Journal on*, 24(2):456–460, April 2006.

[56] R. Weinstein. Rfid: A technical overview and its application to the enterprise. In *IEEE Computer Technology*, pages 27–33, May 2005.

[57] G. Roussos. Enabling RFID in retail. In *IEEE Computer Technology*, pages 27–33, March 2006.

[58] G. Bhanage, Y. Zhang, Y.Y. Zhang, W. Trappe, and R. Howard. Rollcall: The design for a low-power and highly-robust asset tracking system. Technical report, Technical Report WINLAB-TR-291, WINLAB, Rutgers University, December 2006.

[59] A. Jones, R. Hoare, S. Dontharaju, and S. Tung. An automated,reconfigurable, low-power RFID tag. In *Design Automation Conference, 2006 43rd ACM/IEEE*, 2006.

[60] S. Jain and S. Das. Collision avoidance in a dense RFID network. In *WiNTECH*, pages 49–56, September 2006.

[61] F. Schoute. Dynamic frame length ALOHA. In *IEEE Transactions on Communications*, pages 565–568, April 1983.

[62] M. Kodialam and T. Nandagopal. Fast and reliable estimation schemes in rfid systems. In *MobiCom*, pages 322–333, September 2006.

[63] B. Zhen, M. Kobayashi, and M. Shimizu. To read transmitter-only rfid tags with confidence. In *PIMRC'04*, September 2004.

[64] J. Cha and J. Kim. Novel anti-collision algorithms for fast object identification in rfid system. In *Proceedings of the 11th International Conference on Parallel and Distributed Systems - Workshops (ICPADS'05)*, pages 63–67, September 2005.

[65] C. Floerkemeier and M. Wille. Comparison of transmission schemes for framed aloha based rfid protocols. In *Applications and the Internet Workshops, International Symposium on*, January 2006.

[66] D. Hush and C. Wood. Analysis of tree algorithms for rfid arbitration. In *IEEE International Symposium on Information Theory*, August 1998.

[67] C. Law, K. Lee, and K. Siu. Efficient memoryless protocol for tag identification. In *Proceedings of the 4th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, August 2000.

[68] F. Zhou, C. Chen, D. Jin, C. Huang, and H. Min. Evaluating and optimizing power consumption of anti-collision protocols for applications in RFID systems. In *ISLPED'04*, August 2004.

[69] J. Myung and W. Lee. An adaptive memoryless tag anti-collision protocol for RFID networks. In *IEEE ICC*, March 2005.

[70] H. Vogt. Efficient object identification with passive RFID tags. In *International Conference on Pervasive Computing*, pages 98–113, 2002.

[71] S. Verdu. *Multiuser Detection*. Cambridge University Press, Pitt Building, Trumpington Str., Cambridge, UK, first edition, 1998.

[72] J. Proakis. *Digital Communications*. McGraw-Hill Companies, Inc., fourth edition, 2001.

[73] A. Almutairi, S. Miller, H. Latchman, and T. Wong. Power control algorithm for mmse receiver based cdma systems. In *Communications Letters, IEEE*, pages 346–348, November 2000.

[74] N. Benvenuto, B. Carnevale, and S. Tomasin. Energy optimization of cdma transceivers using successive interference cancellation. In *Globecom*, pages 2644–2648, 2004.

[75] J.G. Andrews and T.H.Y. Meng. Amplitude and phase estimation considerations for asynchronous cdma with successive interference cancellation. In *VTC*, pages 1211–1215, 2000.

[76] P. Patel and J. Holtzman. Analysis of a ds/cdma successive interference cancellation scheme using correlations. In *Globecom*, pages 76–80, 1993.

[77] D. Bertsekas and R.Gallager. *Data Networks*. Prentice Hall, NJ, 1992.

[78] Christian Doerr, Michael Neufeld, Jeff Fifield, Troy Weingart, Douglas C. Sicker, and Dirk Grunwald. Multimac - an adaptive mac framework for dynamic radio networking. In *2005 First IEEE International Symposium on Dynamic Spectrum Access Networks(DySPAN)*, 2005.

[79] Byungjoo Lee and Seung Hyong Rhee. Adaptive mac protocol for throughput enhancement in cognitive radio networks. In *International Conference on Information Networking(ICOIN)*, 2008.

[80] Xiangpeng Jing and Dipankar Raychaudhuri. Spectrum co-existence of ieee 802.11b and 802.16a networks using the cscc etiquette protocol. In *Proceedings of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks(Dyspan)*, 2005.

[81] Xiangpeng Jing and Dipankar Raychaudhuri. Spectrum co-existence of ieee 802.11b and 802.16a networks using reactive and proactive etiquette policies. In *ACM Journal Mob. Netw. Appl.*, 2006.

[82] Xiangpeng Jing and Dipankar Raychaudhuri. Global control plane architecture for cognitive radio networks. In *Proceedings of IEEE CogNets 2007 Workshop - Towards Cognition in Wireless Networks (in conjunction with IEEE ICC)*, 2007.

[83] Li Chun Wang, Yin Chih Lu, Chung Wei Wang, and David S. L. Wei. Latency analysis for dynamic spectrum access in cognitive radio: Dedicated or embedded control channel? In *The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2007.

[84] Mansi Thoppian, S. Venkatesan, and Ravi Prakash. Csma-based mac protocol for cognitive radio networks. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks(WoWMoM)*, 2007.

[85] Alex Chia chun Hsu, David S. L. Wei, and C.-C. Jay Kuo. A cognitive mac protocol using statistical channel allocation for wireless ad-hoc networks. In *IEEE Wireless Communications and Networking Conference(WCNC)*, 2007.

[86] Hang Su and Xi Zhang. Cream-mac: An efficient cognitive radio-enabled multi-channel mac protocol for wireless networks. In *The IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)*, 2008.

[87] Cordeiro Carlos and Challapali Kiran. C-mac: A cognitive mac protocol for multi-channel wireless networks. In *IEEE International Symposium on Dynamic Spectrum Access Networks(DySPAN)*, 2007.

[88] Hang Su and Xi Zhang. Channel-hopping based single transceiver mac for cognitive radio networks. In *IEEE Information Theory Society, the 42th Conference on Information Sciences and Systems (CISS)*, 2008.

[89] Yogesh R Kondareddy and Prathima Agrawal. Synchronized mac protocol for multi-hop cognitive radio networks. In *IEEE International Conference on Communications(ICC)*, 2008.

[90] L. Maccari, L. Mainardi, M. Marchitti, N. Prasad, and R. Fantacci. Lightweight, distributed access control for wireless sensor networks supporting mobility. In *IEEE International Conference on Communications (ICC)*, May 2008.

[91] K. Chang and K. Shin. Distributed authentication of program integrity verification in wireless sensor networks. *ACM Transactions on Information and System Security (TISSEC)*, 11(3), March 2008.

[92] Wenyuan Xu, Pandurang Kamat, and Wade Trappe. Trieste: A trusted radio infrastructure for enforcing spectrum etiquettes. In *IEEE Workshop on Networking Technologies for Software Defined Radio (SDR) Networks*, 2006.

[93] Transmission control protocol. http://en.wikipedia.org/wiki/TCP.

[94] S. Orfanidis. *Optimum Signal Processing.* McGraw-Hill Publishing Company, NY, 2007.

Curriculum Vita

Yu Zhang

2009 Ph. D. in Electrical Engineering, Rutgers University

2002 M. E. in Electrical Engineering, Wuhan University, Wuhan, P.R. China

1999 B. S. in Electrical Engineering, Wuhan University, Wuhan, P.R. China

2004-2009 Graduate Assistant, WINLAB, Rutgers University

2002-2003 Software Engineer, Beijing Fiberhome Technology

2008 Yu Zhang, Zang Li, Wade Trappe, Evaluation of Localization Attacks on Power-Modulated Challenge-Response Systems, in IEEE Transactions on Information Forensics and Security, June 2008.

2007 Yu Zhang, Zang Li, Wade Trappe, Power-Modulated Challenge-Response Schemes for Verifying Location Claims, in Proceedings of IEEE Global Communications Conference, November 2007.

2007 Gautam Bhanage, Yu Zhang, Yanyong Zhang, Wade Trappe, and Rich Howard, RollCall : The Design For A Low Cost And Power Efficient Active RFID Asset Tracking System, in Proceedings of Eurocon 07, September 2007.

2007 Yu Zhang, Gautam Bhanage, Wade Trappe, Yanyong Zhang, and Rich Howard, Facilitating an Active Transmit-only RFID System Through Receiver-based Processing, in Proceedings of 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, June 2007.

2006 Shu Chen, Yu Zhang, Wade Trappe, Inverting sensor networks and actuating the environment for spatio-temporal access control, in Proceedings of the fourth ACM workshop on Security of ad hoc and sensor networks, October 2006.