THE HONG KONG POLYTECHNIC UNIVERSITY
DEPARTMENT OF COMPUTING

MAKING THE MOST OF BITS: EFFICIENT PROTOCOLS FOR MONITORING LARGE RFID SYSTEMS

By
Kai BU

A thesis submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
January, 2013

CERTIFICATE OF ORIGINALITY

Date: **January, 2013**

Author: **Kai BU**

Title: **Making the Most of Bits: Efficient Protocols for Monitoring Large RFID Systems**

Department: **Department of Computing**

Degree: **Ph.D.** Convocation: **April 11** Year: **2013**

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Signature of Author Kai BU

Abstract

Radio-Frequency Identification (RFID) technology has stimulated many innovative applications in, for example, supply chain management, health care, mobile payment, ticketing, and target tracking. To support these applications, various protocols are necessary for monitoring RFID systems. A recent trend in monitoring protocol design, triggered by the explosion of RFID technology over the last few years, is toward efficient monitoring in large RFID systems. Established efforts for efficient monitoring protocols lie primarily in cardinality estimation, missing-tag detection, and sensed-information collection.

In the thesis, we concentrate on designing efficient protocols for two other monitoring operations, namely misplaced-tag pinpointing and replication attack detection in large RFID systems. We strive for efficiency gains in the protocol design toward making the most of bits. The fewer bits an RFID protocol requires readers and tags to transmit, the more efficiency it promises to large RFID systems.

The major contributions of the thesis to efficient monitoring of large RFID systems are threefold. First, we propose efficient misplaced-tag pinpointing protocols. Misplacement errors fail optimal inventory placement and thus significantly decrease profit. The existing misplaced-tag pinpointing solution needs to collect a large amount

of data from tags. It suffers from time inefficiency and energy inefficiency as well if active tags are in use. The proposed protocols gain time efficiency and energy efficiency by leveraging reader-related data instead of tag-related data and requiring only a fraction of tags to respond. Second, we propose efficient and privacy-preserving replication attack detection protocols. Replication attacks threaten RFID applications but are hard to prevent. Existing detection protocols are limited in efficiency and privacy mainly due to the transmission of tag IDs. The proposed protocols leverage the broadcast nature and collisions, preserving privacy by avoiding ID transmission. They also integrate lightweight operations to save unnecessary execution time and tag responses, and therefore harvest promising gains in both time efficiency and energy efficiency. Third, considering that tag IDs should be protected to enable and secure privacy-sensitive applications in anonymous RFID systems, we further propose a replication attack detection protocol without requiring tag IDs as a priori. More specifically, the anonymity requires that readers cannot query tag IDs from tags or backend servers. The proposed protocol leverages unreconciled collisions to uncover replication attacks. An unreconciled collision is probably due to responses from multiple tags with the same ID, exactly the evidence of replication attacks. Both theoretical analysis and simulation experiments demonstrate that the proposed protocol can detect replication attacks in anonymous RFID systems fairly fast with required accuracy. In summary, we hope that together with established efficient protocols the proposals in the thesis can greatly benefit various monitoring operations in large RFID systems.

Publications

Journal Articles

1. **Kai Bu**, Xuan Liu, Jiaqing Luo, Bin Xiao, and Guiyi Wei, Unreconciled Collisions Uncover Cloning Attacks in Anonymous RFID Systems, *IEEE Transactions on Information Forensics and Security (**TIFS**)*, vol. 8, no. 3, pp. 429-439, 2013.

2. Qingjun Xiao, Bin Xiao, **Kai Bu**, and Jiannong Cao, Iterative Localization of Wireless Sensor Networks: An Accurate and Robust Approach, *IEEE/ACM Transactions on Networking (**ToN**)*, accepted for publication, 2013.

3. Zhixin Sun, Bingqing Luo, Yadang Cheng, and **Kai Bu**, [in Chinese] Secure P2P topology based on a multidimensional DHT space mapping, *SCIENTIA SINICA Informationis*, vol. 43, no. 3, pp. 343-360, 2013.

4. **Kai Bu**, Bin Xiao, Qingjun Xiao, and Shigang Chen, Efficient Misplaced-Tag Pinpointing in Large RFID Systems, *IEEE Transactions on Parallel and Distributed Systems (**TPDS**)*, vol. 23, no. 11, pp. 2094-2106, 2012.

5. **Kai Bu**, Qingjun Xiao, Zhixin Sun, and Bin Xiao, Toward Collinearity-Aware and Conflict-Friendly Localization for Wireless Sensor Networks, *Computer Communications (**COMCOM**)*, vol. 35, no. 13, pp. 1549-1560, 2012.

6. Qingjun Xiao, **Kai Bu**, Zhijun Wang, and Bin Xiao, Robust Localization against Outliers in Wireless Sensor Networks, *ACM Transactions on Sensor Networks (**TOSN**)*, accepted for publication, 2012.

7. Qingjun Xiao, **Kai Bu**, Bin Xiao, and Limin Sun, Efficient Protocol Design for Dynamic Tag Population Monitoring in Large-Scale RFID Systems, *Concurrency and Computation: Practice and Experience*, accepted for publication, 2012.

8. Jiaqing Luo, Bin Xiao, **Kai Bu**, and Shijie Zhou, Understanding and Improving Piece-related Algorithms in the BitTorrent Protocol, *IEEE Transactions on Parallel and Distributed Systems (**TPDS**)*, accepted for publication, 2012.

9. Zhixin Sun, Bingqing Luo, Yadang Chen, and **Kai Bu**, Secure P2P Topology based on a Multidimensional DHT Space Mapping, *SCIENCE CHINA Infomation Science (Science in China Series F)*, November, 2012.

Conference Papers

1. Jiwei Li, **Kai Bu**, Xuan Liu, and Bin Xiao, ENDA: Embracing Network Inconsistency for Dynamic Application Offloading, in *Proc. of The Second Mobile Cloud Computing Workshop (**MCC**)*, Hong Kong, China, August 12, 2013.

2. Fei Wang, Bin Xiao, **Kai Bu**, and Jinshu Su, Detect and Identify Blocker Tags in Tree-based RFID Systems, in *Proc. of IEEE International Conference on Communications (**ICC**)*, Budapest, Hungary, June 9-13, 2013.

3. Xuan Liu, Shigeng Zhang, **Kai Bu**, and Bin Xiao, Complete and Fast Unknown Tag Identification in Large RFID Systems, in *Proc. of the 9th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (**MASS**)*, Las Vegas, Nevada, USA, October 8-11, 2012, pp. 47-55.

4. **Kai Bu**, Xuan Liu, and Bin Xiao, [Poster/Short Paper] Fast Cloned-Tag Identification Protocols for Large-Scale RFID Systems, in *Proc. of the 20th IEEE/ACM International Workshop on Quality of Service (**IWQoS**)*, Coimbra, Portugal, June 4-5, 2012, pp. 1-4.

5. Qingjun Xiao, **Kai Bu**, and Bin Xiao, Efficient Monitoring of Dynamic Tag Populations in RFID Systems, in *Proc. of the 9th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (**EUC**)*, Melbourne, Australia, October 24-26, 2011, pp. 106-113. **Best Paper Award**

6. **Kai Bu**, Bin Xiao, Qingjun Xiao, and Shigang Chen, Efficient Pinpointing of Misplaced Tags in Large RFID Systems, in *Proc. of the 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications*

*and Networks (**SECON**)*, Salt Lake City, Utah, USA, June 27-30, 2011, pp. 287-295.

7. Zhixin Sun, **Kai Bu**, and Ke Ding, A Multidimensional Mapping Mechanism Based Secure Routing Method for DHT, in *Proc. of the 4th International ICST Conference on Scalable Information Systems (**INFOSCALE**)*, Hong Kong, China, June 10-11, 2009, pp. 192-205.

8. **Kai Bu** and Zhixin Sun, A Method Based on AMHI for DDoS Attacks Detection and Defense, in *Proc. of the 9th International Conference for Young Computer Scientists (**ICYCS**)*, Zhang Jia Jie, Hunan, China, November 18-21, 2008, pp. 1571-1576.

Submissions

1. **Kai Bu**, Qingjun Xiao, Xuan Liu, and Bin Xiao, Efficient and Privacy-Preserving Detection of Replication Attacks in Large RFID Systems, under review in *IEEE Transactions on Parallel and Distributed Systems (**TPDS**)*.

2. **Kai Bu**, Qingjun Xiao, Xuan Liu, and Bin Xiao, Efficient Capture of Reader Interference Dynamics in Multi-Reader RFID Systems, under review in *Computer Communications (**COMCOM**)*.

3. **Kai Bu**, Xuan Liu, and Bin Xiao, Approaching the Time Lower Bound on Cloned-Tag Identification for Large RFID Systems, under review in *Ad Hoc Networks*.

4. **Kai Bu**, Xuan Liu, and Bin Xiao, Less is More: Efficient RFID-based 3D Localization, under review in *The Tenth Annual IEEE International Conference on Mobile Ad-hoc and Sensor Systems (**MASS 2013**)*.

5. Xuan Liu, **Kai Bu**, Bin Xiao, and Shigeng Zhang, LOCK: A Fast Tag Scanning Mechanism with Mobile Readers, under review in *The 21st IEEE International Conference on Network Protocols (**ICNP 2013**)*.

Acknowledgements

First and foremost, I am deeply grateful to all the nice and great people I have met or known. Although too many of them may not be mentioned, what I learned from them inspires everything I have accomplished.

Regarding the thesis, I would like to especially thank my supervisor, Prof. Bin Xiao, for giving me encouragement and freedom to pursue interests in research. I would also like to thank my co-supervisor Prof. Hong-Va Leong, Prof. Shigang Chen, Prof. Qixin Wang, Prof. Lei Zhang, Prof. Wei Lou, Prof. Zhijun Wang, and all our group members (Qingjun Xiao, Jiaqing Luo, Guobin Liu, Xuan Liu, Fei Wang, and Jiwei Li) for helping improve preliminary versions of some chapters of the thesis.

"Any author stands on the shoulders of those who have come before." Those listed in the bibliography have contributed excellent research that motivate me to explore new research directions in the thesis.

Finally, and most importantly, I would like to gratefully and sincerely thank my parents, sisters, and relatives for their continuous, unconditional love and support. I owe every step of my progress, intellectual as well as moral, to them. The thesis is dedicated to them all.

Hong Kong S.A.R., China Kai BU

May 28, 2013

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 RFID Explosion

Radio-Frequency Identification (RFID) technology has stimulated innovative applications in various fields, such as supply chain management [Delen et al., 2007, Koh et al., 2003], health care [Janz et al., 2005], ticketing [web, d], mobile payment [web, g], and target tracking [web, h, Zhang et al., 2007, 2010]. To support these applications, researchers dedicate significant efforts to key system monitoring operations. Such monitoring operations include, for example, tag identification [Lee et al., 2005, Myung et al., 2007, Qian et al., 2010, Zanetti et al., 2010a], cardinality estimation [Kodialam and Nandagopal, 2006, Li et al., 2010b, Qian et al., 2011, Shahzad and Liu, 2012], finding popular categories [Sheng et al., 2008], missing-tag detection and identification [Li et al., 2010a, Luo et al., 2012, Tan et al., 2008b, Zhang et al., 2011b], and information collection [Chen et al., 2011, Qiao et al., 2011, Yue et al., 2012]. All the efforts dedicated to RFID application invention and monitoring bring an unprecedented explosion of RFID technology over the last few years—already 1.3 billion tags were in the market in 2005, and even 33 billion were expected in 2010 [web, c].

1.2 Efficiency, Efficiency, Efficiency for Large RFID Systems

Although "Accuracy, Accuracy, Accuracy", the desired spirit of journalism [web, f], maybe the best analogy to the purpose of RFID monitoring protocols with endurable complexity and cost, "Efficiency, Efficiency, Efficiency" must be of primary concern for monitoring protocols in large RFID systems. When an RFID system accommodates up to, for example, hundreds of thousands of tagged objects, it becomes obviously inefficient or even unrealistic to obtain interested system statuses always through identifying all tags. Sometimes it may be necessary to trade a limited amount of accuracy for a leap in efficiency.

1.2.1 Efficiency Metrics

Two mostly adopted metrics for evaluating RFID monitoring protocols' time efficiency and energy efficiency are *the execution time* and *the number of tag responses*, respectively [Qiao et al., 2011].

First, time efficiency—measured by protocol execution time—is highly important for an RFID monitoring protocol to be scalable as RFID systems grow large. Time efficiency is a primary concern for almost all prior RFID work, such as tag identification [Lee et al., 2005, Liu et al., 2012, Myung et al., 2007, Qian et al., 2010, Zanetti et al., 2010a], cardinality estimation [Kodialam and Nandagopal, 2006, Li et al., 2010b, Qian et al., 2011, Shahzad and Liu, 2012, Xiao et al., 2013], and missing-tag detection and identification [Li et al., 2010a, Tan et al., 2008b, Zhang et al., 2011b].

Second, energy efficiency is measured by the number of tag responses during protocol execution. An energy-efficient RFID monitoring protocol is essential in an RFID

system accommodating many active tags. Active tags can significantly impulse the growth of RFID applications because of their ability of initiating communication and their hundreds-of-feet communication radius, which is much longer than that of passive tags. However, active tags depend on self-carried batteries to enable any operation. Thus, enjoying the improved system performance brought by active tags, we should make the energy cost as low as possible by controlling the number of tag responses [Li et al., 2010b, Luo et al., 2012, Qiao et al., 2011].

1.2.2 Established Efforts

Tag cardinality estimation [Kodialam and Nandagopal, 2006] brings about the first leap in efficient monitoring protocols for large RFID systems. Instead of counting tags through reading all their IDs, Kodialam and Nandagopal propose quickly estimating the number of tags up to a desired level of accuracy. The proposal requires tag responses much shorter than tag IDs and leverages the distribution of the number of tag responses in some time slots. For a given estimation accuracy, the proposal can execute cardinality estimation of any-sized tag sets almost in constant time [Kodialam and Nandagopal, 2006]. Compared with tag identification that has linear time complexity with respect to the number of tags, the proposal promises far more time efficiency to large RFID systems. Qian et al. then propose fast cardinality estimation protocols for RFID systems with multiple readers covering all tags [Qian et al., 2008, 2011]. More recently, Li et al. investigate energy-efficient cardinality estimation protocols for RFID systems with active tags [Li et al., 2010b]; the proposed protocols save energy by requiring only a subset of tags to send responses.

Missing-tag detection [Tan et al., 2008b] is another important monitoring operation that desires efficiency in large RFID systems. A straightforward way to detect

missing tags is first reading all tag IDs and then comparing them against the recorded ones. Since collecting all tag IDs in large RFID systems is inefficient, Tan et al. propose trading detection accuracy for time efficiency [Tan et al., 2008b]. The proposal requires tags to respond in a number of time slots and leverages the fact that, when a time slot supposed to be occupied by tag responses becomes empty, some tag(s) must be missing. The follow-up studies lie in improving energy efficiency by not requiring all active tags to respond [Luo et al., 2012] and improving accuracy by identifying all missing tags [Li et al., 2010a, Zhang et al., 2011b].

More recently, information collection in sensor-augmented RFID systems becomes a new spot where efficient protocols start shining [Chen et al., 2011]. Chen et al. propose a multi-hashing scheme to collect sensed data from all tags without transmitting tag IDs [Chen et al., 2011]. The elimination of ID transmission is critical for guaranteeing time efficiency. Qiao et al. then extend the multi-hashing scheme to applications that collect sensed information from only a subset of tags [Qiao et al., 2011], while Yue et al. dedicate efforts to efficient information collection in large RFID systems with multiple readers [Yue et al., 2012].

1.3 Thesis Contributions

In the thesis, we concentrate on seeking efficient solutions of two other important monitoring operations for large RFID systems, namely *misplaced-tag pinpointing* and *replication attack detection*. We hope that together with established efficient protocols the proposals in the thesis can greatly benefit various monitoring operations in large RFID systems. We next provide an overview of the proposed protocols in the thesis and highlight their major contributions.

1.3.1 Efficient Misplaced-tag Pinpointing

Of great importance to RFID applications in production economics is misplaced-tag pinpointing (MTP), because misplacement errors fail optimal inventory placement and thus significantly decrease profit. Optimal placement can increase profit by up to 8.1% [Bishop, 2003]. Such an increase would yield $1.1 billion more profit for Wal-Mart, the world's largest retailer [Ferreira Chaves et al., 2010]. The existing MTP solution [Ferreira Chaves et al., 2010], originally proposed from a data-processing perspective, collects and processes a large amount of data. It suffers from time-inefficiency (and energy-inefficiency as well if active tags are in use). The problem of finding efficient solutions for the MTP problem from the communication protocol design perspective has never been investigated before.

In Chapter 2, we propose a series of protocols toward efficient MTP solutions in large RFID systems. The proposed protocols detect misplaced tags using reader positions instead of tag positions to guarantee the efficiency and scalability as system scale grows, because RFID readers are much fewer than tags. Considering applications that employ active tags, we further propose a solution requiring responses from only a subset of tags in favor of energy saving. We also design a distributed protocol that enables each reader to independently detect misplaced tags. We then investigate how to apply the proposed protocols in scenarios with tag mobility. To evaluate the proposed protocols, we analyze their optimal performances to demonstrate their efficiency potential and also conduct extensive simulation experiments. The results show that the proposed protocols can significantly increase the time efficiency and the energy efficiency by over 70% on average when compared with the best existing work [Ferreira Chaves et al., 2010].

1.3.2 Efficient Replication Attack Detection

Replication attacks threaten RFID applications but are hard to prevent. Launching a replication attack, an attacker compromises genuine tags and produces their replicas, namely replicated tags. Since replicated tags carry copies of compromised genuine tags' data (e.g., IDs and keys), they behave exactly the same as genuine tags and therefore threaten RFID applications that use the genuineness of tags to validate the authenticity of tagged objects [Bolotnyy and Robins, 2007, Lehtonen et al., 2009b]. For example, attached with replicated tags, products in RFID-enabled supply chains cause financial losses [Koh et al., 2003], healthcare facilities in RFID-aided hospitals jeopardize personal safety [Janz et al., 2005], while RFID-incorporated passport cards may even threaten national security [Koscher et al., 2009]. Existing replication attack detection protocols require intact knowledge of tag IDs and thus are limited in accuracy, efficiency, or even privacy [Koh et al., 2003, Lehtonen et al., 2009a,b, Mirowski and Hartnett, 2007, Zanetti et al., 2010b].

In Chapter 3, we propose a series of protocols toward efficient and privacy-preserving replication attack detection with guaranteed accuracy in large-scale RFID systems. The proposed protocols do not resort to complex cryptography techniques, inefficient tag-wise scanning, or privacy-unaware transmission of tag IDs. Instead, our protocols leverage the broadcast nature and collisions to detect replication attacks, being affordable to off-the-shelf low-cost tags. Toward efficient detection of replication attacks for large-scale RFID systems, we propose introducing two light-weight operations, vector broadcast and slot index recalculation. Armed with these two operations, our protocols can avoid unnecessary execution time and tag responses, and thus harvest significant gains in both time efficiency and energy efficiency. We

evaluate the performance of the proposed protocols through theoretical analysis and extensive simulations. The results show that, when the confidence level is 0.99 and the tolerance ratio of compromised tags is 0.001, our best protocol outperforms the state-of-the-art tag-wise scanning based protocol in time efficiency and energy efficiency by 98.5% and 72.8% on average, respectively.

In Chapter 4, we extend replication attack detection to anonymous RFID systems without tag IDs as a priori. In anonymous RFID systems, tag IDs should be protected to enable and secure privacy-sensitive applications. To this end, we leverage unreconciled collisions to uncover replication attacks. An unreconciled collision is probably due to responses from multiple tags with the same ID, exactly the evidence of replication attacks. This insight inspires GREAT, our pioneer protocol for replication attack detection in anonymous RFID systems. We evaluate the performance of GREAT through theoretical analysis and extensive simulations. The results show that GREAT can detect replication attacks in anonymous RFID systems fairly fast with required accuracy. For example, when only six out of 50,000 tags are replicated, GREAT can detect the replication attack in 75.5 seconds with probability at least 0.99.

1.4 Thesis Outline

The rest of the thesis is organized as follows. Chapter 2 proposes protocols toward efficient MTP solutions in large RFID systems. The proposed protocols can work in a distributed fashion while being robust against tag mobility. Chapter 3 proposes protocols toward efficient and privacy-preserving replication attack detection with

guaranteed accuracy in large RFID systems. The proposed protocols gain time efficiency and energy efficiency through getting rid of complex cryptography techniques, inefficient tag-wise scanning, and privacy-unaware transmission of tag IDs. Chapter 4 takes the first step toward replication attack detection in anonymous RFID systems without requiring tag IDs as a priori and proposes a pioneer protocol. The proposed protocol leverages unreconciled collisions to uncover replication attacks with guaranteed accuracy. Finally, Chapter 5 concludes the thesis and indicates future work.

The primary research outputs emerged from the thesis are as follows:

- Kai Bu, Xuan Liu, Jiaqing Luo, Bin Xiao, and Guiyi Wei, Unreconciled Collisions Uncover Cloning Attacks in Anonymous RFID Systems, *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 8, no. 3, pp. 429-439, 2013.

- Kai Bu, Bin Xiao, Qingjun Xiao, and Shigang Chen, Efficient Misplaced-Tag Pinpointing in Large RFID Systems, *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 23, no. 11, pp. 2094-2106, 2012.

- Kai Bu, Xuan Liu, and Bin Xiao, [Poster/Short Paper] Fast Cloned-Tag Identification Protocols for Large-Scale RFID Systems, in *Proc. of the 20th IEEE/ACM International Workshop on Quality of Service (IWQoS)*, Coimbra, Portugal, June 4-5, 2012, pp. 1-4.

- Kai Bu, Bin Xiao, Qingjun Xiao, and Shigang Chen, Efficient Pinpointing of Misplaced Tags in Large RFID Systems, in *Proc. of the 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, Salt Lake City, Utah, USA, June 27-30, 2011, pp.

287-295.

- Kai Bu, Qingjun Xiao, Xuan Liu, and Bin Xiao, Efficient and Privacy-Preserving Detection of Replication Attacks in Large RFID Systems, under review in *IEEE Transactions on Parallel and Distributed Systems (TPDS)*.

- Kai Bu, Xuan Liu, and Bin Xiao, Approaching the Time Lower Bound on Cloned-Tag Identification for Large RFID Systems, under review in *Ad Hoc Networks*.

Chapter 2

Efficient Pinpointing of Misplaced Tags in Large RFID Systems

Radio-Frequency Identification (RFID) technology brings many innovative applications. Of great importance to RFID applications in production economics is misplaced-tag pinpointing (MTP), because misplacement errors fail optimal inventory placement and thus significantly decrease profit. The existing MTP solution [Ferreira Chaves et al., 2010], originally proposed from a data-processing perspective, collects and processes a large amount of data. It suffers from time-inefficiency (and energy-inefficiency as well if active tags are in use). The problem of finding efficient solutions for the MTP problem from the communication protocol design perspective has never been investigated before. In this chapter, we propose a series of protocols toward efficient MTP solutions in large RFID systems. The proposed protocols detect misplaced tags using reader positions instead of tag positions to guarantee the efficiency and scalability as system scale grows, because RFID readers are much fewer than tags. Considering applications that employ active tags, we further propose a solution requiring responses from only a subset of tags in favor of energy saving. We also design a distributed protocol that enables each reader to independently detect

misplaced tags. We then investigate how to apply the proposed protocols in scenarios with tag mobility. To evaluate the proposed protocols, we analyze their optimal performances to demonstrate their efficiency potential and also conduct extensive simulation experiments. The results show that the proposed protocols can significantly increase the time efficiency and the energy efficiency by over 70% on average when compared with the best existing work.

2.1 Overview

Radio-Frequency Identification (RFID) technology stimulates innovative applications in various fields, such as supply chain management [Delen et al., 2007] and target tracking [Zhang et al., 2007, 2010]. To support these applications, researchers dedicate significant effort to addressing important problems in RFID systems. Such problems include tag identification [Lee et al., 2005, Myung et al., 2007, Qian et al., 2010, Zanetti et al., 2010a], cardinality estimation [Kodialam and Nandagopal, 2006, Li et al., 2010b, Qian et al., 2011], finding popular categories [Sheng et al., 2008], missing-tag detection and identification [Li et al., 2010a, Tan et al., 2008b, Zhang et al., 2011b], and misplaced-tag pinpointing [Ferreira Chaves et al., 2010]. In this chapter, we concentrate on one of these important problems, Misplaced-Tag Pinpointing (MTP) in large RFID systems.

The MTP problem aims to detect and pinpoint tags attached to misplaced inventory items in a large warehouse, retailing store, wharf, or airport. The misplacement error is a notorious foe against optimal inventory placement. Optimal placement can increase profit by up to 8.1% [Bishop, 2003]. Such an increase would yield $1.1 billion more profit for Wal-Mart, the world's largest retailer [Ferreira Chaves et al., 2010].

Misplacement errors, however, hinder us from enjoying this benefit. The statistics in [Raman et al., 2001] show that on average consumers of a leading retailer cannot find 16% of inventory items in the stores because those items are misplaced. Furthermore, Hong Kong International Airport, one of the largest airports in the world, loses more than $20 million annually to relocate misplaced and mis-transported bags [Qian et al., 2010]. Countermeasures against misplacement errors, therefore, represent one of the primary concerns in production economics. Rekik et al.suggest that the RFID technology can be adopted to reduce inventory misplacement errors [Rekik et al., 2008]. One significant application is thus to pinpoint misplaced tags and in return to pinpoint their tagged misplaced inventory items for the purpose of replacement.

The recent solution for MTP is a database-oriented method called RFID Planogram Compliance Verification (RPCV) [Ferreira Chaves et al., 2010]. In RPCV, one reader controls many antennas, each of which is located at one position. RPCV requires inventory items to be placed exactly following a layout plan and uses tag vectors to represent tags attached to inventory items. A tag vector includes the number of readings by the right antenna (i.e., whose corresponding position covers where the tagged item should be placed) and by wrong ones. RPCV finds misplaced tags by processing and classifying all tag vectors. The design of RPCV primarily focuses on computational efficiency (i.e., how fast tag vectors can be processed by the server to find misplaced tags), whereas it is not concerned with communication efficiency (i.e., how fast the information from tags can be collected in order to construct tag vectors and how much energy the tags have to spend in the collection process). In fact, information collection from all tags in a large RFID system is very time-consuming [Chen et al., 2011, Qian et al., 2010]. Time efficiency may outweigh computational

efficiency. In addition, when battery-powered active tags are used, energy efficiency becomes important. RPCV requires each tag to respond dozens of times for finding misplaced items whereas too many tag responses cost active tags a lot of energy [Li et al., 2010b, Qiao et al., 2011]. Another limitation of RPCV is its dependence on layout plans. Practically, it is laborious and challenging to place inventory items exactly in accordance with layout plans.

Unlike RPCV's focus on data processing, this chapter studies the MTP problem from a new angle—the communication protocol design perspective. Particularly, we design efficient communication protocols that collect information from tags to readers for detecting and pinpointing misplaced tags. In order to achieve time efficiency and energy efficiency, we have to abandon the basic approaches in RPCV [Ferreira Chaves et al., 2010], including its underlying tag vectors, and replace them with reader vectors, which take much less time to collect. To save energy, we randomly select only a subset of tags to respond each time when reader vectors are constructed, which compares favorably with RPCV where every tag has to transmit many times. Our protocols do not require tagged items to be placed strictly based on a pre-determined layout plan. To make it more efficient and robust, we further propose a distributed solution that enables each reader to independently detect misplaced tags. We also discuss how to handle scenarios with tag mobility.

In summary, this chapter makes the following contributions to efficient MTP solution in large RFID systems.

- Investigate Basic MTP protocols (*B-MTP*) based on tag-wise positioning and

propose a heuristic reader activation method to accelerate MTP protocols. B-MTP collects less information for tag positioning than RPCV collects for forming tag vectors. Yet the performance analysis of B-MTP demonstrates its inefficiency and unscalability in large RFID systems.

- Propose a Time-efficient MTP protocol (*T-MTP*) through eliminating tag-wise positioning. T-MTP detects misplaced tags using reader vectors instead of tag vectors. Only misplaced tags need to be located. The performance analysis demonstrates that T-MTP also has better energy efficiency in comparison with B-MTP.

- Propose an Energy-Time-efficient MTP protocol (*ET-MTP*) to further enhance energy efficiency of T-MTP by requiring only a subset of tags to send responses.

- Validate the performance of proposed protocols through extensive analysis and simulation. The results show that the proposed protocols can significantly outperform RPCV [Ferreira Chaves et al., 2010] for improving time efficiency and energy efficiency by over 70% on average.

- Propose a Distributed MTP protocol (*D-MTP*), a distributed protocol that enables each reader to independently detect misplaced tags within its coverage. Our analysis shows that D-MTP is more time-efficient than T-MTP and more energy-efficient than ET-MTP in some cases.

- Investigate how to distinguish mobile tags from misplaced tags using multi-round misplaced-tag detection results.

The rest of this chapter is organized as follows. Section 2.2 defines the problem

and the system. Section 2.3 discusses B-MTP and indicates its limitations through performance analysis. Section 2.4 and Section 2.5 present T-MTP and ET-MTP toward high time efficiency and high energy efficiency, respectively. Section 2.6 reports the simulation results. Section 2.7 discusses D-MTP, a distributed MTP protocol that enables each reader to independently detect misplaced tags. Section 2.8 discusses tag mobility and channel reliability. Finally, Section 2.9 concludes the chapter and indicates the future work.

2.2 System Model

2.2.1 Problem Overview

Consider a large RFID system that consists of a set of readers $R = \{r_1, \ldots, r_i, \ldots, r_m\}$ and a set of tags $T = \{t_1, \ldots, t_j, \ldots, t_n\}$. The readers are deployed at known positions to provide position reference for positioning tags [Wang et al., 2007]. We also call them *reference readers*. The tags could be either passive or active according to specific system requirements. Each tag has a unique ID and attaches to an inventory item. (Hereafter in this chapter, we use terms *inventory item* and *tag* interchangeably.) Predefined bit positions on the ID of each tag specify various kinds of information about the inventory item [web, a]. In particular, one section of the tag ID, called *category ID*, specifies the category of the inventory item attached with the tag [web, a, Sheng et al., 2008]. The set of distinct category IDs is denoted as $C = \{c_1, \ldots, c_k, \ldots, c_u\}$. Inventory items are placed by categories—a category of inventory items should be properly placed together in a certain area. But we do not require inventory placement to strictly follow any layout plan, which, however, is a necessary assumption for RPCV [Ferreira Chaves et al., 2010].

In such an RFID system, we formulate the MTP problem as follows. We denote by A_k the area where tags in the category of c_k are placed. If a tag of category c_k locates away from A_k, we regard it as misplaced.[1] The MTP problem is to pinpoint misplaced tags. Once misplaced tags are pinpointed, one can directly walk toward or use navigation methods (e.g., [Matic et al., 2010]) to approach them for replacement. We observe that in practice, although some tags could be misplaced, a majority of tags in each category are still properly placed. This observation provides valuable hints on detecting misplaced tags without using tag positions (Sections 2.4, 2.5, 2.7, and 2.8.1).

2.2.2 Assumptions and Justifications

In this chapter, we concentrate on scenarios not that complicated but general enough to acquire insights for efficient MTP solutions. Several reasonable assumptions we make are as follows.

We assume that an inventory item list in accordance with all present items is maintained on a backend server that executes the MTP protocol and communicates with the readers. The list is updated whenever new items move into the system or existing items move out. Considering that tags may be stolen by misbehaving workers or customers, we can adopt missing-tag detection and identification methods [Li et al., 2010a, Tan et al., 2008b, Zhang et al., 2011b] to timely detect and identify the missing tags. The records corresponding to identified missing tags should be immediately deleted from the list.

To locate tags, we assume that the representative RFID positioning scheme in [Wang et al., 2007] is adopted. In this scheme, a reader has a set of transmission

[1]The exact location of A_k is not pre-determined based on a layout plan. Instead, it is simply where the tags in c_k happen to be. If most tags in c_k are there while one tag is moved to another location, then that tag is misplaced (away from others in the same category).

power levels [web, e]. The communication radius corresponding to a power level can be obtained by a set of reference tags deployed at known positions [Wang et al., 2007] or an RF site survey using a positioning device and radio signal strength measurement device [Zhou et al., 2007]. Without loss of generality, we consider scenarios where readers are deployed on the ceiling of the system, so that communications with tags are relatively free of obstacle. To locate a tag, we need its distance measurements to at least $h \geq 3$ reference readers. Note that theoretically it requires distance measurements to at least four reference readers to locate a tag in 3D space [Hendrickson, 1992]. Combining the implicit constraint that a tag cannot locate higher than the height of the system, $h = 3$ however is the least requirement for the basic positioning scenario in [Wang et al., 2007].

For now, we consider only scenarios where readers follow sequential reading and tags keep stationary. First, it surely will be more complicated when multiple readers read tags in parallel, because of the reader-reader collision problem [Zhou et al., 2007]. The reader-reader collision problem occurs when two readers covering common tags are active to read the tags at the same time. Query messages from the two readers will collide and thus the tags will not send any response. Therefore the reader-reader collision problem may mislead detection of misplaced tags. Although simultaneously activating readers with disjoint covering regions can avoid the reader-reader collision problem [Zhou et al., 2007], it is still very challenging to achieve this with readers frequently adjusting transmission power levels for tag positioning [Wang et al., 2007]. But we can easily apply our MTP protocols to parallel reading scenarios once a more sophisticated reader scheduling protocol is available. We will discuss time efficiency gains by parallel reading in Section 2.6.4.

Second, when mobile tags exist, being away from the supposed area cannot verify a misplaced tag. We need to further verify whether the tag is carried by moving machines or wandering customers. To sidestep this problem, we first investigate efficient MTP protocols with all tags being stationary (Sections 2.3, 2.4, 2.5, 2.6, and 2.7) and then discuss how we can apply them to scenarios with mobile tags (Section 2.8.1).

2.2.3 Performance Metrics

We consider two performance metrics, *the execution time* and *the number of tag responses*, for evaluating MTP protocols' time efficiency and energy efficiency, respectively.

First, time efficiency—measured based on protocol execution time—is highly important for an MTP protocol to be scalable as RFID systems grow large. Time efficiency is a primary concern for almost all prior RFID work, such as tag identification [Lee et al., 2005, Myung et al., 2007, Qian et al., 2010, Zanetti et al., 2010a], cardinality estimation [Kodialam and Nandagopal, 2006, Li et al., 2010b, Qian et al., 2011], and missing-tag detection and identification [Li et al., 2010a, Tan et al., 2008b, Zhang et al., 2011b].

Second, energy efficiency is measured by the number of tag responses during protocol execution. An energy-efficient MTP protocol is essential in an RFID system with many active tags. Active tags can significantly impulse the growth of RFID applications because of their ability of initiating communication and their hundreds-of-feet communication radius, which is much longer than that of passive tags. However, active tags depend on self-carried batteries to enable any operation. Thus, enjoying the improved system performance brought by active tags, we should make the energy

cost as low as possible by controlling the number of tag responses [Li et al., 2010b, Qiao et al., 2011].

2.3 Preliminary and Basic MTP Protocols

This section discusses B-MTP based on tag-wise positioning, which locates each tag. We analyze its performance and limitations, indicating the demand for more efficient MTP protocols in large RFID systems.

2.3.1 B-MTP Design

Intuitively, if all tags have been located, it is straightforward to determine whether any and where tags are misplaced. Using position estimations of tags in the category of c_k, we can easily bound the area A_k. Any tags in the category of c_k but out of A_k are therefore misplaced ones and should be replaced.

Based on the above intuition, we investigate two B-MTP designs, Individual Positioning based B-MTP (*IPB-MTP*) and Collision Arbitration based B-MTP (*CAB-MTP*). They differ from each other in tag-wise positioning process.

IPB-MTP Design

IPB-MTP locates tags in n rounds, each of which is dedicated to locating one of n tags. In each round, readers are sequentially activated to broadcast a query message containing a tag ID (recorded in the inventory item list) and wait for the tag's response. Receiving a tag response indicates that the tag is within the reader's covering region corresponding to current transmission power level. The round for a tag ends when h readers are identified to cover the tag and each of the h readers has determined the minimum transmission power level l_{min} for it to cover the tag. Recall

that h is the number of reference readers it needs to locate a tag. The communication radius corresponding to l_{min} for a reader r_i covering a tag t_j is regarded as the distance measurement d_{ij} between r_i and t_j [Wang et al., 2007]. Let (x_{ri}, y_{ri}, z_{ri}) denote the known position of a reference reader r_i, (x_{tj}, y_{tj}, z_{tj}) the position estimation of a tag t_j, and H the height of the system. The representative RFID positioning scheme in [Wang et al., 2007] estimates the position of t_j as follows:

$$\begin{aligned} (x_{tj}, y_{tj}, z_{tj}) = & \underset{(x_{tj}, y_{tj}, z_{tj})}{\arg\min} \sum_{i=1}^{h} (\frac{d_{ij} - \hat{d_{ij}}}{d_{ij}})^2, \\ & \text{subject to } z_{tj} \leq H; \\ & \hat{d_{ij}} = \sqrt{(x_{ri} - x_{tj})^2 + (y_{ri} - y_{tj})^2 + (z_{ri} - z_{tj})^2}. \end{aligned} \tag{2.1}$$

Its best performance can limit the position error to less than 5% of the longest edge of the system [Wang et al., 2007].

To speed up MTP protocols, we can activate readers in an optimized order rather than always following the ordering defined in the set R. The heuristic stems from the truth that if a reader r_i covers a tag t_j (i.e., r_i and t_j can communicate with each other), then

- readers near to r_i are more likely to cover t_j than those far from r_i;
- it is usually faster to find the readers that cover tags in the same category with t_j when starting from readers near to r_i than starting from those far from r_i.

Hence, after we find a reader r_i that covers a tag t_j, we activate the remaining readers in ascending order of their distances to r_i. In a subsequent round for another tag t_k that belongs to the same category of t_j, we will activate r_i first and then activate other readers in the ascending order of their distances to r_i.

CAB-MTP Design

There are different ways to collect information about which readers cover which tags. In IPB-MTP, we iterate through tags. For each tag, readers take turn to broadcast the tag ID to see if they cover the tag. Alternatively, we can iterate through readers. For each reader, it performs a tag-identification protocol [Lee et al., 2005, Myung et al., 2007, Qian et al., 2010, Zanetti et al., 2010a] to identify the IDs of the tags within its coverage. This leads to our second protocol, CAB-MTP, which differs from IPB-MTP only in its way of collecting information about which readers cover each tag. The rest of the protocol is the same.

There are two types of tag-identification protocols, based on slotted Aloha [Lee et al., 2005, Roberts, 1975, Sheng et al., 2010] and tree traversal [Myung et al., 2007, Namboodiri and Gao, 2007], respectively. We design CAB-MTP using slotted Aloha because it can yield higher efficiency in large systems than can Tree-traversal [Qian et al., 2010]. We hereby briefly review the basics of slotted Aloha based collision arbitration protocols to keep the chapter self-contained and refer interested readers to [Lee et al., 2005, Roberts, 1975, Sheng et al., 2010] for more details. Using slotted Aloha, the reader sends a query frame with a certain number of time slots (*frame size*) and each tag picks up a random time slot to respond. A time slot chosen by no tag, only one tag, or multiple tags is usually called an *empty slot*, a *singleton slot*, or a *collision slot*, respectively [Kodialam and Nandagopal, 2006]. The reader can correctly receive the tag response only in a singleton slot; the reader has to continuously send new frames with adjusted frame size until no collision occurs.

2.3.2 Performance Analysis and Limitations

We first derive a performance lower bound for B-MTP and then analyze the optimal performances of IPB-MTP and CAB-MTP to indicate how close they can approach the lower bound.

Remark 2.1. *A lower bound on the number of tag responses N_{B-MTP} and the execution time T_{B-MTP} for B-MTP to pinpoint misplaced tags is as follows:*

$$N_{B-MTP} = hn,$$
$$T_{B-MTP} = hnt_{id},$$

where t_{id} denotes the transmission time of the tag ID.

We derive Remark 2.1 as follows. Because B-MTP requires each tag to respond to at least h reference readers for tag positioning, the total number of tag responses for locating all tags is at least hn. Corresponding to each tag response, the tag ID should be contained either in the query message by IPB-MTP or in the response by CAB-MTP. Then the execution time costed by each tag response is at least t_{id}. Therefore, the total execution time of B-MTP is at least hnt_{id}.

The lower bound in Remark 2.1, however, is hardly achievable, mostly due to two reasons. First, besides transmission of the tag ID, transmission of the tag response when using IPB-MTP or the query message when using CAB-MTP takes additional time, even though such additional time is very small when compared with t_{id}. (We omit this additional time cost in the analysis.) Second, it also incurs additional overhead when reference readers further communicate with the to-be-located tag to determine l_{min}.

We will analyze the optimal performances of IPB-MTP and CAB-MTP by Remark 2.2 and Remark 2.3, respectively, indicating how close they can approach the lower bound.

Remark 2.2. *The optimal number of tag responses $N_{IPB-MTP}$ and the optimal execution time $T_{IPB-MTP}$ for IPB-MTP to pinpoint misplaced tags are as follows:*

$$N_{IPB-MTP} = hn,$$
$$T_{IPB-MTP} = 2hnt_{id}.$$

We derive Remark 2.2 as follows. Suppose that r_i successfully receives t_j's response at transmission power level l_{temp}. The best case for r_i to determine l_{min} is when l_{temp} happens to be l_{min}. In this case, t_j cannot hear r_i when r_i sends a query message at transmission power one level lower than l_{min}. Then t_j will not send any response. As each reference reader covering a to-be-located tag needs to initiate at least one more query message containing the tag ID to determine l_{min}, it is straightforward that at least hnt_{id} more time is needed. Thus the optimal execution time of IPB-MTP is $2hnt_{id}$. The optimal number of tag responses can reach hn as no tag response is induced by l_{min} determination in the best case.

Remark 2.3. *The optimal number of tag responses $N_{CAB-MTP}$ and the optimal execution time $T_{CAB-MTP}$ for CAB-MTP to pinpoint misplaced tags are as follows:*

$$N_{CAB-MTP} = ehn,$$
$$T_{CAB-MTP} = (e+1)hnt_{id}.$$

We derive Remark 2.3 as follows. CAB-MTP using slotted Aloha achieves the best performance when each tag is covered by at least h readers, each of which reads the tag at transmission power level l_{min}. Let n_{ri} denote the number of tags covered by a reader r_i. Because the highest efficiency of slotted Aloha is $\frac{1}{e}$ (i.e., optimally $\frac{1}{e}$ of the tags can be identified within one query frame, where e is the natural constant) [Chen et al., 2011, Qian et al., 2010], it takes r_i at least $en_{ri}t_{id}$ time to identify n_{ri}

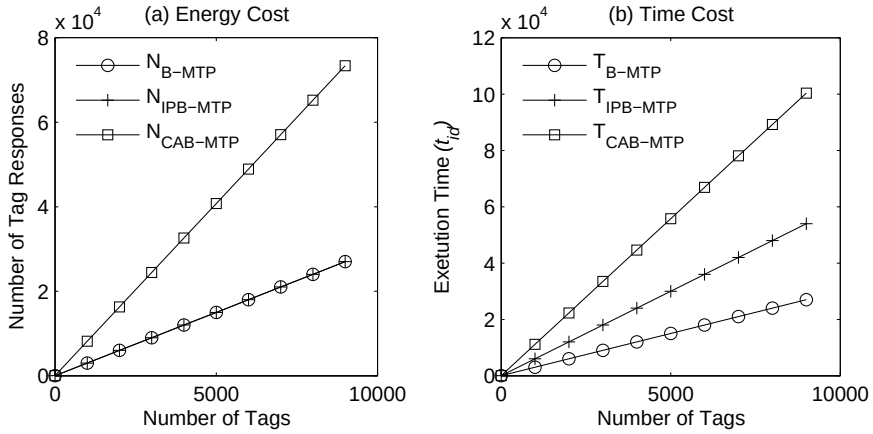


Fig. 2.1: Performance comparison of B-MTP, IPB-MTP, and CAB-MTP.

tags. Then the optimal execution time is derived by

$$\begin{aligned}(\sum_{i=1}^{m} en_{ri}t_{id}) + hnt_{id} &\geq ehnt_{id} + hnt_{id} \\ &= (e+1)hnt_{id}.\end{aligned}$$

The first line uses $\sum_{i=1}^{m} n_{ri} \geq hn$, which is deduced from the condition that each tag is covered by at least h readers. Furthermore, because all unread tags should respond to the query message and optimally $1 - \frac{1}{e}$ of them will continue to respond to the following query message, we can derive the optimal number of tag responses as follows:

$$\begin{aligned}&\sum_{i=1}^{m}(n_{ri} + (1-\frac{1}{e})n_{ri} + (1-\frac{1}{e})^2 n_{ri} + ...) \\ &\approx \sum_{i=1}^{m} en_{ri} \geq ehn.\end{aligned}$$

The second derivation again uses $\sum_{i=1}^{m} n_{ri} \geq hn$.

Figure 2.1 summarizes the performance of B-MTP designs in Remarks 2.1, 2.2, and 2.3, under the basic scenario where $h = 3$. IPB-MTP outperforms CAB-MTP because its

optimal performance (especially the optimal number of tag responses) is closer to the lower bound. A major limitation of B-MTP designs is that their best performances are linear with respect to the system size (i.e., the number of tags in an RFID system). This limitation not only hinders protocol efficiency but also decreases protocol scalability for large RFID systems. Next we will present two more efficient MTP protocols toward increasing time efficiency and energy efficiency, respectively.

2.4 T-MTP: Time-efficient Misplaced-Tag Pinpointing Protocol

This section presents T-MTP and analyzes its performance and limitations. T-MTP enhances time efficiency trough eliminating tag-wise positioning for detecting misplaced tags. Only misplaced tags need to be located for replacement.

2.4.1 Motivation

Reader positions rather than tag positions can also be used to detect misplaced tags. Figure 2.2 illustrates the intuition under a scenario with uniformly deployed readers. After all readers read tags in a category, two separate clusters of readers covering tags in the category are formed. Obviously the majority of this category of tags are covered by readers in the left cluster, which has a much larger cluster size (i.e., the number of included readers) than does the right one. The tag covered by readers in the right smaller cluster is therefore detected as misplaced tags and needs to be located for replacement.

If the category is compactly stored in a small place, only a few nearby readers cover it. They will form a small cluster that is similar to the one for a single misplaced

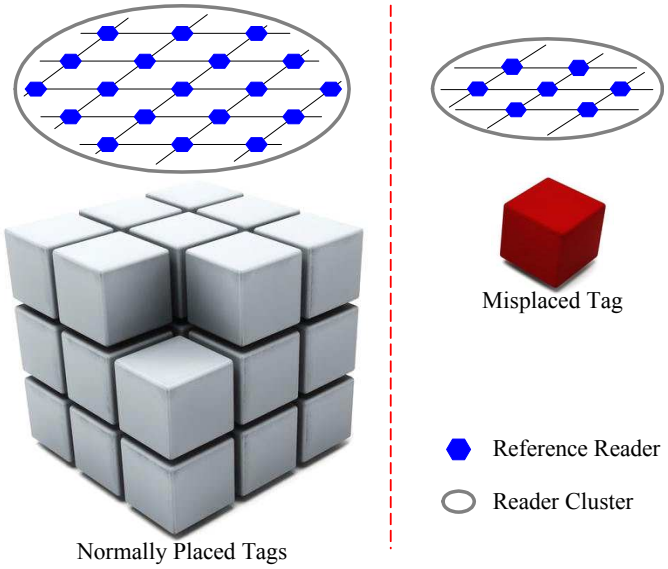


Fig. 2.2: Misplaced-tag detection using reader clusters. The tag covered by readers in the right smaller cluster is misplaced away from those covered by readers in the left larger cluster.

tag. In this case, we can easily detect misplaced tags by estimating the number of tags under either reader cluster. But in a large RFID system a category of tags usually spreads in a large area covered by many readers. Such RFID systems are of our interest in this chapter.

Since tag-wise tag positioning is a major factor limiting the efficiency of B-MTP, we believe that, through eliminating tag-wise positioning, T-MTP can yield promising efficiency gains. Next we will detail the design of T-MTP using the above idea.

2.4.2 T-MTP Design

T-MTP is expected to enhance time efficiency in two respects. First, adopting the aforementioned idea, one time slot is enough for a reader to determine whether it covers any tags. This is because we can ensure that no tag or at least one tag is covered when the only time slot is empty or not. Second, because we no longer require

tag-wise positioning, tag IDs are unnecessary to be included in responses. Indices for distinguishing misplaced tags in the same category could be used if more than one misplaced tags exist. For tag positioning, we need to further distinguish singleton slots from collision slots for determining the number of misplaced tags. In the Philips I-Code system [web, b], 10 bits, which is much shorter than the length of tag ID (usually 96 bits [web, a]), is enough to verify a collision [Kodialam and Nandagopal, 2006].

T-MTP efficiently addresses the MTP problem in two stages, Time-efficient Misplaced-Tag Detection (*T-MTD*) and Pinpointing Information Collection (*PIC*).

Stage I: T-MTD

To detect misplaced tags in a category c_k, T-MTD sequentially activates each reader for one time slot to determine whether it covers tags of category c_k, with all readers using a same transmission power level. Specifically, a reader r_i first broadcasts a query message containing c_k and waits for tag responses. Upon receiving the query message, tags with c_k as the category ID respond by transmitting a 10-bit random bitstring with error-detection (e.g., CRC) embedded. We use 0, 1, or 2 to denote the *slot state* of an empty slot, a singleton slot, or a collision slot, respectively. After each reader being active for one time slot, we form a *reader vector* V with the element $V[i]$ defined by

$$V[i] = \begin{cases} 0, \text{ if } r_i \text{ receives an empty slot,} \\ 1, \text{ if } r_i \text{ receives a singleton slot,} \\ 2, \text{ if } r_i \text{ receives a collision slot.} \end{cases} \tag{2.2}$$

Based on the reader vector V, T-MTD detects misplaced tags through constructing *reader clusters*. A reader cluster consists of readers with $V[i] \neq 0$ surrounded by

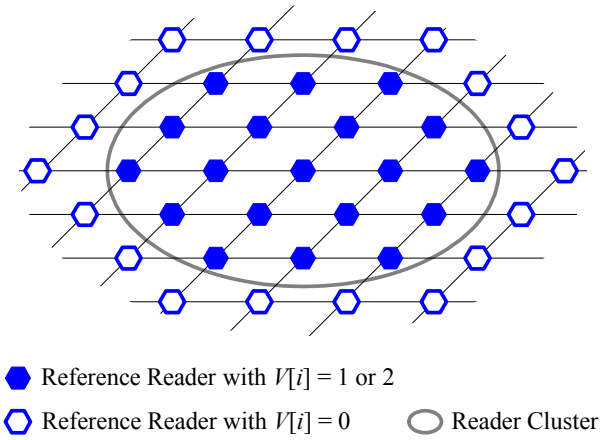


Fig. 2.3: Reader cluster construction using the reader vector V.

readers with $V[i] = 0$. Figure 2.3 illustrates reader cluster construction. The number of readers in a reader cluster indicates the cluster size. Intuitively, the largest reader cluster covers properly placed tags because a majority of tags in a category are supposed to be placed in the right area (Section 2.2.1). Tags covered by readers in other smaller reader clusters are separated away from the right area and therefore are detected as misplaced ones.

Stage II: PIC

PIC further activates readers to collect enough information for positioning misplaced tags, which are detected in stage I. By Equation 2.1, the to-be-collected information is distance measurements between each misplaced tag and at least h reference readers. The active reference reader broadcasts a query message containing category ID c_k by gradually tuning the transmission power level until it determines l_{min}. (Note that we simply use category ID c_k without collecting specific tag ID because the primary goal of MTP is to find where misplaced tags of a certain category locate rather than to differentiate misplaced tags in the same category.) The communication radius

corresponding to l_{min} is used as the distance measurement. Whether a reader covers one or multiple misplaced tags can be determined using the slot states of singleton or collision, respectively. PIC may further activate readers out of the smaller clusters to get enough distance measurements. We still prefer the aforementioned heuristic reader activation method (Section 2.3.1) for accelerating PIC.

2.4.3 Discussion of Misplaced-Tag Detection Accuracy

T-MTD has no false positives but false negatives. T-MTD is false positive free because some misplaced tags must exist if multiple reader clusters exist. Otherwise, if only one reader cluster is constructed, T-MTD reports no misplaced tags. T-MTD, however, may return false negatives if misplaced tags are not far away from the supposed area beyond a distance threshold. Specifically, we deduce that the distance threshold is $2d_r$, where d_r represents the distance interval of uniformly deployed readers. Let $\min |(x_{tj}, y_{tj}) - (x_{ki}, y_{ki})|$ $((x_{ki}, y_{ki}) \in A_k.XY)$ represent the distance between a misplaced tag t_j and its supposed area A_k, where $A_k.XY$ is a set containing all (x, y) coordinates of positions within A_k. T-MTD may fail to detect misplaced tags satisfying the following constraint:

$$\min_{(x_{ki}, y_{ki}) \in A_k.XY} |(x_{tj}, y_{tj}) - (x_{ki}, y_{ki})| \leq 2d_r.$$

2.4.4 Performance Analysis and Limitations

Remark 2.4. *The optimal number of tag responses N_{T-MTP} and the optimal execution time T_{T-MTP} for T-MTP to pinpoint misplaced tags are as follows:*

$$N_{T-MTP} = (\alpha(h-1)+1)n,$$
$$T_{T-MTP} = (um + \alpha hn)(t_{cid} + t_{10b}),$$

where t_{cid} denotes the transmission time of the category ID, t_{10b} the transmission time of a 10-bit bitstring, and α the ratio of the number of misplaced tags to the number of tags.

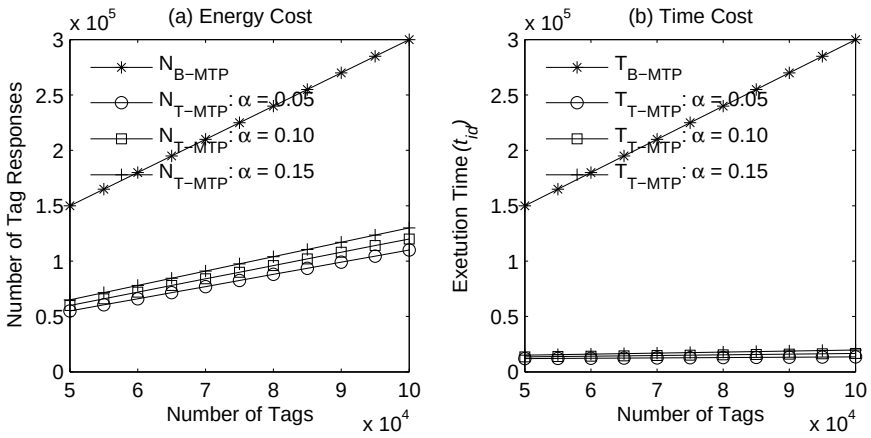


Fig. 2.4: Performance comparison of B-MTP and T-MTP.

We derive Remark 2.4 as follows. Because T-MTD does not need tag-wise positioning, the optimal case is when both the following conditions are satisfied:

- each tag is exactly covered by only one reader;
- for a misplaced tag t_j, given current transmission power level l_{min} for readers that cover t_j, transmission power one level higher than l_{min} is enough for to-be-activated readers to cover t_j.

In the optimal case, PIC induces at least $h(t_{cid}+t_{10b})$ time cost and $h-1$ tag responses for collecting enough information to locate t_j. Combining $um(t_{cid}+t_{10b})$ time cost and n tag responses for T-MTD to form the reader vector V, we therefore derive N_{T-MTP} and T_{T-MTP} in Remark 2.4.

Figure 2.4 plots the optimal performance of T-MTP under the scenario where $m = 50$, $u = 1000$, $h = 3$, and $\alpha = 0.05$, 0.10, 0.15, compared with that of B-MTP. In large RFID systems, it is ordinary that the number of tags in a category is more than the number of readers. Thus Figure 2.4 only shows curves subject to $n \geq um$. The length of the category ID is determined by $\log_2\lceil u \rceil = 10$ bits. Suppose

that 96-bit tag ID is in use. The transmission time t_{cid} and t_{10b} can be approximately represented by $\frac{10}{96}t_{id}$ and $\frac{10}{96}t_{id}$, respectively. As shown in Figure 2.4(b), compared with B-MTP, T-MTP not only significantly decreases the execution time but also exhibits a much better scalability because T_{T-MTP} increases slightly as the system scale grows. Furthermore, T-MTP also outperforms B-MTP in higher energy efficiency because of fewer tag responses as shown in Figure 2.4(a).

Although T-MTP decreases the number of tag responses than B-MTP by a factor of

$$\frac{hn - (\alpha(h-1)+1)n}{hn} = (1-\alpha)(1-\frac{1}{h}),$$

T-MTP still causes a certain amount of unnecessary tag responses. The reason for this limitation is that whenever a reader reads a category of tags, all tags in this category and within the reader's coverage will respond upon receiving the query message. Two or more tag responses, however, make no difference to distinguish a collision slot. We will present a more energy-efficient MTP protocol against this limitation.

2.5 ET-MTP: Energy-Time-efficient Misplaced-Tag Pinpointing Protocol

This section presents ET-MTP, which further enhances the energy efficiency while inheriting the time efficiency of T-MTP. We first discuss the basic idea of energy cost reduction and then detail protocol design and performance analysis.

2.5.1 Motivation

If a reader covering n' tags broadcasts a query message and each tag responds with probability p, we can expect pn' tag responses [Li et al., 2010b]. Figure 2.5 illustrates

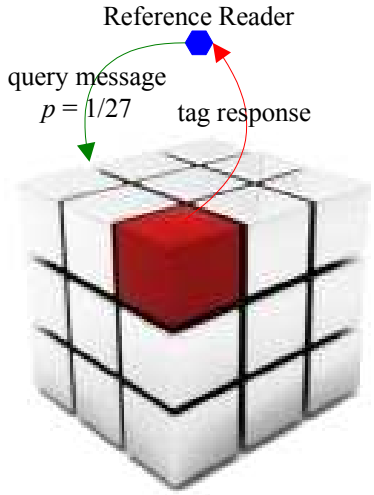


Fig. 2.5: One tag response expected with 27 tags responding with probability $p = \frac{1}{27}$.

an example scenario where $n' = 27$ and $p = \frac{1}{27}$. The reader is therefore likely to receive only one tag response ($n'p = 27 \times \frac{1}{27} = 1$). Similarly, when a reader receives a tag response, we can expect $\frac{1}{p}$ tags being covered. Thus, it is natural to conceive that readers covering the majority of properly placed tags in a category can still receive tag responses even if not all tags respond. The energy cost can be therefore reduced if we design an MTP protocol accordingly.

2.5.2 ET-MTP Design

ET-MTP efficiently addresses the MTP problem in two stages, Energy-Time-efficient MTD (*ET-MTD*) and PIC. Compared with T-MTD, ET-MTD forms the reader vector V more energy-efficiently. Reader cluster construction, misplaced-tag detection, and PIC follow the same processes as that of T-MTP. Next we only expatiate on how ET-MTD forms the reader vector V for the purpose of conciseness.

ET-MTD forms the reader vector V through two substeps, in which ET-MTD

forms reader vectors V_1 and V_2, respectively. In the first substep, readers are sequentially activated to broadcast a query message comprising a category ID c_k and a probability value p_k. Upon receiving the query message, tags in the category of c_k send a response with probability p_k. We define $p_k \in (0, 1)$ to be

$$p_k = \frac{1}{n'_k}, \tag{2.3}$$

where n'_k denotes the maximum number of category-c_k inventory items that can be covered by a reader with transmission power level l. We estimate n'_k by $n'_k = \lceil \frac{V_l}{V_k} \rceil$, where V_l represents the volume of a reader's covering region at transmission power level l and V_k represents the volume of an inventory item of category c_k. It is easy to estimate V_l and V_k using the communication radius corresponding to l and the dimension information of inventory items, respectively. In the first substep, we only need 1-bit tag response to confirm a tag's presence. Regarding time slots in which a reader receives responses as *non-empty slots*, we form V_1 as

$$V_1[i] = \begin{cases} 0, \text{ if } r_i \text{ receives an empty slot,} \\ 1, \text{ if } r_i \text{ receives a non-empty slot.} \end{cases}$$

Since we form V_1 with tags responding with probability p_k, chances are that some readers with $V_1[i] = 0$ cover tags in the category of c_k but receive no tag response. To avoid this issue, we further introduce the second substep.

In the second substep, ET-MTD forms the reader vector V_2 through activating readers with $V_1[i] = 0$. Each active reader sends a query message comprising only c_k. Tags in the category of c_k must respond to the query message upon receiving it. A 10-bit random bitstring is sent as tag response for the purpose of distinguishing the slot state. Activating each reader with $V_i = 0$ for one time slot, we form the reader

vector V_2 as

$$V_2[i] = \begin{cases} 0, & \text{if } V_1[i] \neq 0, \\ 0, 1, \text{or } 2, & \text{by Equation 2.2 if } V_1[i] = 0. \end{cases}$$

Finally, ET-MTD forms the reader vector V by $V[i] = V_1[i] + V_2[i]$—the same reader vector as T-MTD forms by Equation 2.2. Using V, ET-MTP constructs reader clusters, detects misplaced tags, and conducts PIC for positioning misplaced tags exactly the same with T-MTP. The detection accuracy of ET-MTD is also similar to that of T-MTD, as discussed in Section 2.4.3.

2.5.3 Discussion of Energy Cost Reduction

We consider the scenario that a reader r_i covers n'_k tags and informs each tag to respond with probability p_k. The number of tag responses follows a binomial distribution when forming $V_1[i]$. When all or no tags respond, ET-MTP yields no energy cost reduction. If no tag responds when forming $V_1[i]$, all tags will be enforced to respond when forming $V_2[i]$. In this case ET-MTP degenerates into T-MTP. Otherwise, ET-MTP reduces energy cost through decreasing the number of tag responses. Specifically, we conclude the probability of reducing k' tag responses, denoted as $Pr(k' \mid p_k, n'_k)$, as follows:

$$Pr(k' \mid p_k, n'_k) =$$
$$\begin{cases} p_k^{n'_k} + (1 - p_k)^{n'_k}, & \text{if } k' = 0, \\ \binom{n'_k}{n'_k - k'} p_k^{n'_k - k'} (1 - p_k)^{k'}, & \text{if } 0 < k' < n'_k. \end{cases}$$

Substituting p_k by Equation 2.3, we derive $Pr(k' \mid n'_k)$, the probability of ET-MTP

reducing k' tag responses given n'_k tags covered, as follows:

$$Pr(k' \mid n'_k) =$$
$$\begin{cases} \frac{1}{{n'_k}^{n'_k}} + (1 - \frac{1}{n'_k})^{n'_k}, & \text{if } k' = 0, \\ \binom{n'_k}{n'_k - k'} \frac{1}{{n'_k}^{n'_k - k'}} (1 - \frac{1}{n'_k})^{k'}, & \text{if } 0 < k' < n'_k. \end{cases}$$

Basically, it is highly probable to reduce $n'_k - 1$ tag responses. This is because $n'_k p_k = 1$ (by Equation 2.3) tag response has the highest probability given that the number of tag responses follows a binomial distribution.

Another common approach to conserving energy for active tags is *sleep scheduling* [Chlamtac et al., 1999, Jeong and Jeon, 2006]. Sleep scheduling aims to switch wireless nodes between sleep mode and active mode such that a wireless node keeps in sleep mode as often as possible. The rationale of sleep scheduling is that being active and listening to (or receiving) queries from readers consume energy, while such energy consumption can be conserved if a wireless node switches to sleep mode. However, it is quite challenging to determine the optimal length of sleep period that can minimize energy consumption yet cannot indulge any sleeping wireless node in application performance deterioration [Chlamtac et al., 1999, Jeong and Jeon, 2006]. In, for example, the MTP problem of our concern, it is hard for an MTP protocol to quickly detect misplaced tags that are in the sleep mode.

We choose to detect misplaced tags with all tags in active mode and conserve energy of active tags in two ways. First, we build ET-MTP on top of T-MTP, which is more time-efficient than B-MTP. Through reducing the execution time, we shorten the time for each tag to listen to (or receive) queries from readers; such shortened time in return reduces energy consumption of active tags. Second, we design ET-MTP in

such a way that it enforces only a subset of tags to respond. Since it is well known that transmitting packets costs wireless nodes more energy than listening to (or receiving) queries, we can harvest a significant energy efficiency gain by suppressing a large number of packets, namely tag responses [Li et al., 2010b, Qiao et al., 2011].

2.5.4 Performance Analysis and Limitations

Remark 2.5. *The optimal number of tag responses N_{ET-MTP} and the optimal execution time T_{ET-MTP} for ET-MTP to pinpoint misplaced tags are as follows:*

$$N_{ET-MTP} = \sum_{k=1}^{u} p_k n_k + \alpha hn,$$

$$T_{ET-MTP} = ((\sum_{k=1}^{u} \beta_k + u)m + \alpha hn)t_{cid} + (\sum_{k=1}^{u} \beta_k m + \alpha hn)t_{10b} + um(t_p + t_{1b}),$$

where n_k represents the number of tags in the category of c_k, β_k the percentage of readers with $V_1[i] = 0$ corresponding to the category of c_k, t_p the transmission time of p_k, and t_{1b} the transmission time of 1-bit tag response.

We derive Remark 2.5 as follows. First, to form V_1, the category ID and the probability value are contained in the query message and 1-bit tag response is used. Thus forming V_1 for all u categories costs $um(t_{cid}+t_p+t_{1b})$ time and at least $\sum_{k=1}^{u} p_k n_k$ tag responses. Second, to form V_2, only readers with $V_1[i] = 0$ broadcast query messages containing the category ID and wait for 10-bit tag responses. Thus forming V_2 for all u categories induces $\sum_{k=1}^{u} \beta_k m(t_{cid} + t_{10b})$ time cost and at least αn tag responses. Finally, PIC costs at least $\alpha hn(t_{cid}+t_{10b})$ time and $\alpha(h-1)n$ tag responses, as we discussed in the analysis of Remark 2.4 (Section 2.4.4). Therefore, we can derive N_{ET-MTD} and T_{ET-MTD} claimed in Remark 2.5 through combining related cost in the above three parts.

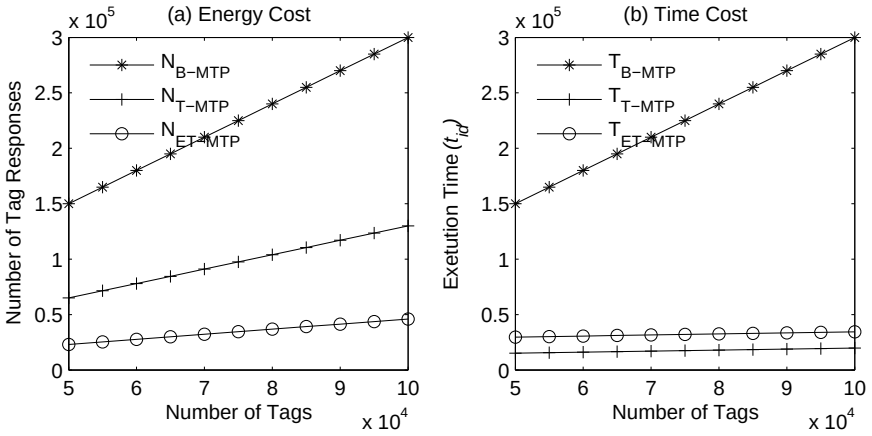


Fig. 2.6: Performance comparison of B-MTP, T-MTP, and ET-MTP.

Figure 2.6 plots the optimal performances of B-MTP, T-MTP, and ET-MTP under the scenario where $m = 50$, $k = 1000$, $h = 3$, $\alpha = 0.15$, 96-bit tag ID, and 7-bit p_k. The transmission time t_p and t_{1b} are approximately estimated by $\frac{7}{96}t_{id}$ and $\frac{1}{96}t_{id}$, respectively. Suppose that a reader can cover at least 100 tags. Then $p_k = \frac{1}{n'_k} \leq 0.01$. For simplicity, we use $p_k = 0.01$ and $\sum_{k=1}^{u} p_k n_k = 0.01 \sum_{k=1}^{u} n_k = 0.01n$ for N_{ET-MTP}, and use u instead of $\sum_{k=1}^{u} \beta_k \leq u$ for T_{ET-MTP}. As we expected, ET-MTP induces fewer tag responses than does T-MTP as shown in Figure 2.6(a). A limitation of ET-MTP is that two rounds of reader activation for forming the reader vector V takes more time than does T-MTP, as shown in Figure 2.6(b).

In summary, it depends on which of time efficiency and energy efficiency is more significant when we choose between T-MTP and ET-MTP. If timely MTP is desired, we prefer T-MTP. If active tags are used and energy saving is desired, we prefer ET-TMP that yields higher energy efficiency than does T-MTP with competitive time efficiency. A hybrid protocol design by adaptively switching between them toward the optimal performance is also worthy of consideration.

2.6 Simulation Evaluation

This section evaluates the efficiency of B-MTP, T-MTP, and ET-MTP by simulations. We compare our protocols with the-state-of-the-art RPCV [Ferreira Chaves et al., 2010]. We use two performance metrics, the execution time and the number of tag responses (Section 2.2.3), to evaluate time efficiency and energy efficiency, respectively. We average the results over 100 trials.

2.6.1 Environment Configuration

We simulate the system as follows. The number of readers and the number of tag categories are $m = 50$ and $u = 1000$, respectively. The number of tags n varies from 50000 to 100000 with $\frac{n}{u}$ per category. The readers are deployed in grid on the ceiling of the simulated system. The number of reference readers for tag positioning by Equation 2.1 is set to $h = 3$. Each reader has 38 tunable transmission power levels as the representative RFID positioning scheme in [Wang et al., 2007]. Each tag has a 96-bit unique ID. The transmission time of the tag ID (i.e., t_{id}), is used as time unit. The transmission time of s bits is estimated by $\frac{s}{96}t_{id}$. The transmission time of the category ID is therefore $t_{cid} = \frac{\lceil \log_2 u \rceil}{96}t_{id} = \frac{10}{96}t_{id}$. Following the system configuration of RPCV, all inventory items are with the same volume. In this case, each reader covers on average $1000 \leq \frac{n}{m} \leq 2000$ tags and 12 bits is enough to express the probability $p_k = \frac{m}{n}$ by Equation 2.3. Thus the transmission time of p_k can be estimated by $\frac{12}{96}t_{id}$.

2.6.2 Comparison Other: RPCV

In RPCV simulation [Ferreira Chaves et al., 2010], each tag needs to be identified dozens of times for RPCV to find misplaced tags. To conduct an objective comparison,

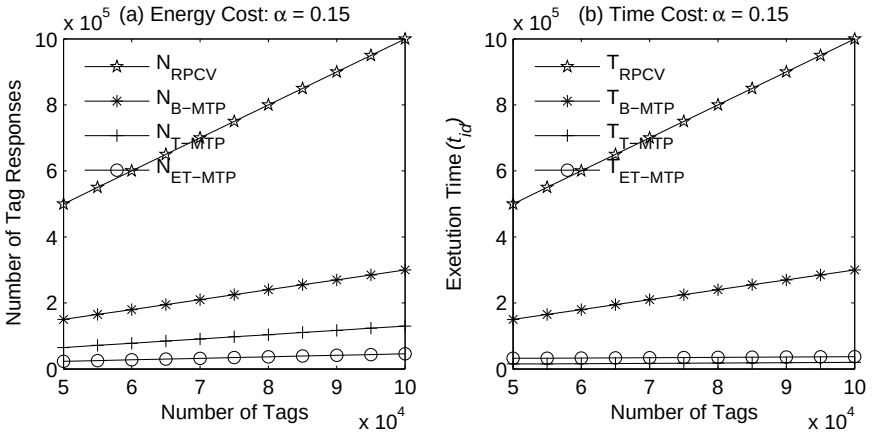


Fig. 2.7: Analytical performance comparison of RPCV, B-MTP, T-MTP, and ET-MTP.

we consider RPCV with each tag being identified 10 times. Both RPCV and our protocols sequentially activate readers to collect information in the simulation. Thus, we can derive a lower bound on the number of tag responses and the execution time of RPCV, denoted as N_{RPCV} and T_{RPCV}, respectively, as the following:

$$N_{RPCV} = 10n,\ T_{RPCV} = 10nt_{id}. \tag{2.4}$$

We hereby compare analyzed optimal performances of RPCV and our protocols in Figure 2.7 to indicate their efficiency potential. Both time cost and energy cost corresponding to the lower bound of RPCV are far beyond that of our protocols. Thus, we directly use the lower bound in Equation 2.4 for comparisons. Similarly, we use the lower bound of B-MTP in Remark 2.1 for the comparison, for B-MTP is neither a wise choice for an efficient MTP solution.

2.6.3 Time Efficiency and Energy Efficiency

We evaluate the performance of proposed protocols with varying number of tags n and misplacement ratio α. In each scenario, we randomly pick αn tags and then

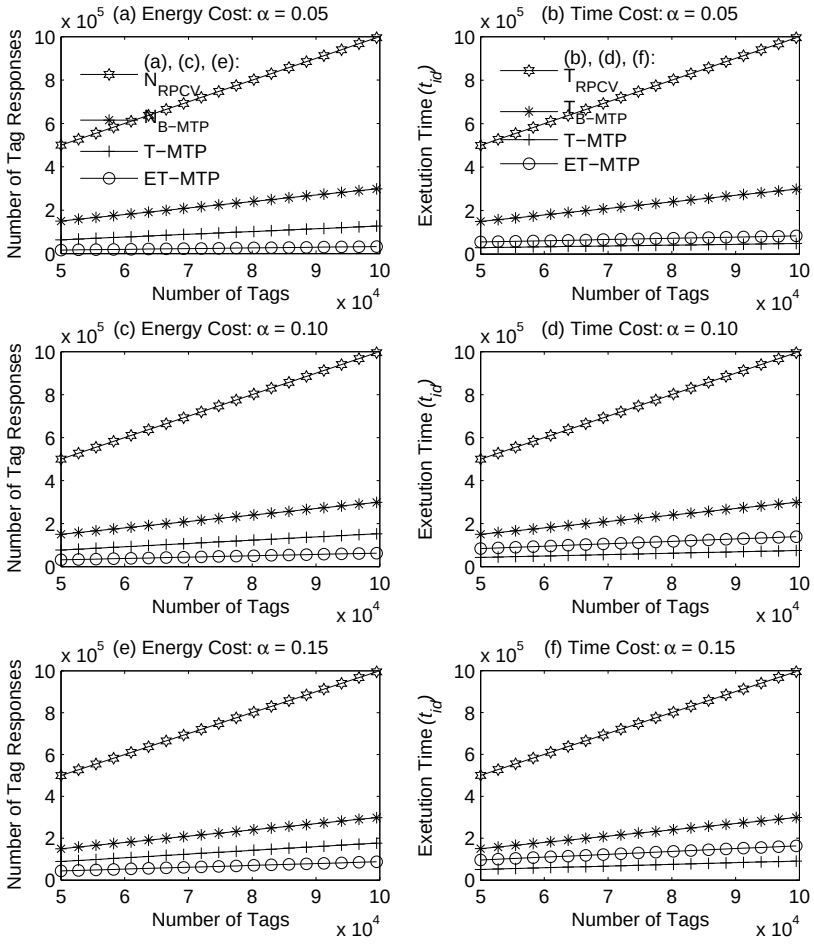


Fig. 2.8: Performance comparison of RPCV, B-MTP, T-MTP, and ET-MTP with varying tag number n and misplacement ratio α.

randomly place them away from the area where they are supposed to be. These tags are therefore the misplaced tags to be pinpointed. Note that we deliberately distribute misplaced tags distant further than the threshold (i.e., $2d_r$, see Section 2.4.3) to avoid false negatives, because we are interested primarily in time efficiency and energy efficiency in this chapter.

Figure 2.8 reports the results under various scenarios in comparison with RPCV. As we expected, all our protocols, namely B-MTP, T-MTP, and ET-MTP, outperform

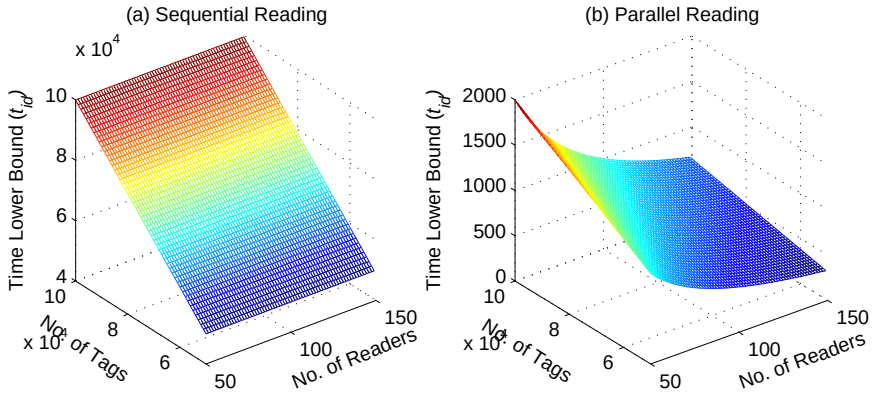


Fig. 2.9: Analytical performance comparison of sequential and parallel reading.

RPCV in both time-efficiency and energy efficiency. Among our protocols, T-MTP yields the highest time efficiency while ET-MTP yields the highest energy efficiency. Both T-MTP and ET-MTP are more time-efficient and energy-efficient than B-MTP. When $\alpha = 0.05$ as shown in Figure 2.8(a)-(b), compared with RPCV, T-MTP can averagely increase the time efficiency by up to 93%, and ET-MTP can averagely increase the energy efficiency by up to 95%. When $\alpha = 0.15$, the time efficiency improvement and the energy efficiency improvement can still be as much as 90% (Figure 2.8(f)) and 91% (Figure 2.8(e)), respectively.

In summary, our efficient MTP protocols, say T-MTP and ET-MTP, can increase both time efficiency and energy efficiency by over 70%, when compared with RPCV [Ferreira Chaves et al., 2010]. This is because T-MTP and ET-MTP are more efficient than B-MTP, which requires 70% lower time and energy cost in the simulation (by Remark 2.1 and Equation 2.4).

2.6.4 Further Discussion of Parallel Reading

We have evaluated protocol efficiency with readers following sequential reading. It is not hard to infer that time efficiency will be further improved if we allow multiple

readers reading in parallel while the number of tag responses will not be affected too much. Without considering the reader collision problem [Zhou et al., 2007], time lower bounds for m readers collecting n tag IDs with sequential and parallel reading are nt_{id} and $\frac{nt_{id}}{m}$, respectively. Figure 2.9 plots the lower bound with varying m and n. Parallel reading provides a promising chance of time efficiency improvement as shown in Figure 2.9.

2.7 D-MTP: Distributed Misplaced-Tag Pinpointing Protocol

This section proposes *D-MTP*, a distributed protocol that enables each reader to independently detect misplaced tags. Our analysis shows that D-MTP is more time-efficient than T-MTP and even more energy-efficient than ET-MTP in some cases.

2.7.1 Motivation and Main Idea

Although T-MTP and ET-MTP activate each reader to query all categories of tags, we observe that not all readers cover all the categories. This observation is especially evident in large RFID systems where a single reader can cover only a small fraction of tags. As for the MTP problem, we even must deliberately make sure that each reader cover only some, but not all, categories of tags. If each reader would cover all categories, we would have not been able to leverage reader positions to quickly detect misplaced tags as we do in T-MTP and ET-MTP.

Motivated by the above observation, we can detect misplaced tags in a more efficient way. The intuition is that if each reader learns which categories it covers, it

can simply broadcast those category IDs and inform tags not in those categories to respond—misplaced tags exist if it receives any response. A desirable side product of the intuition is distributed misplaced-tag detection—a reader can independently verify whether any misplaced tags are within its coverage.

We expect D-MTP to be more efficient than T-MTP and ET-MTP in time efficiency and even in energy efficiency. D-MTP is more time-efficient because it requires each reader to query only categories it covers but T-MTP and ET-MTP require each reader to query all categories. D-MTP could be more energy efficient because it requires only misplaced tags to respond but T-MTP or ET-MTP in addition require all or a fraction of properly placed tags to respond, respectively.

Next we will first investigate how D-MTP learns which categories covered by each reader, and then analyze its performance.

2.7.2 Learning Category Coverage from Tag Monitoring

An RFID system may periodically run various operations to implement important functions, such as missing-tag detection and identification [Li et al., 2010a, Tan et al., 2008b, Zhang et al., 2011b], information collection [Chen et al., 2011], and continuous scanning that monitors a dynamic RFID system with tags frequently moved in or out [Sheng et al., 2010, Xiao et al., 2011, 2012]. These operations monitor tags in real time. We can leverage the information from tag monitoring to deduce which categories covered by each reader. For example, the outputs of continuous scanning are the IDs of all present tags. Examining the IDs of tags covered by a reader, we can easily extract their category IDs.

Misplaced-tag detection using tag monitoring statistics may, however, induce both false negatives and false positives. False negatives arise from tag monitoring statistics

that contain data of misplaced tags. Since it is hard to guarantee that tag monitoring operations only run when all tags are correctly placed, data of misplaced tags may contaminate tag monitoring statistics. Using such contaminated tag monitoring statistics, a reader cannot detect misplaced tags that are already in the statistics, inducing false negatives. We can mitigate the problem of false negative by filtering out misplaced tags related data in the statistics. Learning from past detections is therefore the other way for D-MTP to infer which categories covered by each reader.

False positives arise from obsolete tag monitoring statistics that fail to capture system update after the statistics is obtained. Possible system update operations include rearrangement and replenishment. Rearrangement follows new placement plan and places some categories of tags somewhere else in the system. Replenishment moves more tags into the system and places them next to where stay their same categorized peers. In both cases, some tags may intrude communication regions of readers that do not cover their categories according to the statistics, and therefore trigger false positives. Although it is not practical to prohibit any system update after we obtain latest tag monitoring statistics, it is practical for us to be aware of any system update operation launched ever since. If system update happens between we obtain the statistics and we run MTP protocols, we have to apply our previously proposed protocols (e.g., B-MTP, T-MTP, or ET-MTP) to detect misplaced tags, capturing real-time system status and thus avoiding false positives.

2.7.3 Learning Category Coverage from Past Detections

D-MTP can learn which categories a reader covers from detection results of our previously proposed protocols (e.g., B-MTP, T-MTP, or ET-MTP). In B-MTP (Section 2.3), when a reader has collected all the IDs of tags it covers, it can easily extract

their category IDs. In T-MTP and ET-MTP (Sections 2.4 and 2.5), when a reader has queried all categories, it covers the categories that respond to the queries. Let C_i represent the set of categories covered by a reader r_i, and M_i the set of categories of detected misplaced tags within r_i's coverage. Then we can derive $\hat{C}_i$, the set of categories that r_i correctly covers, by

$$\hat{C}_i = C_i - M_i. \tag{2.5}$$

After getting the set $\hat{C}_i$ for every reader $r_i \in R$, the reader r_i can detect misplaced tags in any categories other than those in $\hat{C}_i$.

We now discuss misplaced-tag detection accuracy of MTP using $\hat{C}_i$ learned from detection results of B-MTP, T-MTP, or ET-MTP. As we discussed in Section 2.7.2, false positives will arise from the statistics that fail to capture latest system update. To avoid them, we still need to verify that no such system update occurs after we collect $\hat{C}_i$. Whether we could avoid false negatives due to contaminated statistics in Section 2.7.2 depends on which of B-MTP, T-MTP, and ET-MTP we use to generate $\hat{C}_i$. B-MTP can successfully avoid them because B-MTP uses tag-wise positioning and induces no false negatives. T-MTP and ET-MTP can eliminate misplaced categories that they can detect, by Equation 2.5. Since T-MTP and ET-MTP may induce false negatives, we cannot rule out misplaced categories that they fail to detect. But compared with tag monitoring statistics, B-MTP, T-MTP, and ET-MTP still try the best to generate less contaminated statistics that induce fewer false negatives.

2.7.4 Performance Analysis

We analyze time efficiency and energy efficiency of D-MTP compared with that of T-MTP and ET-MTP. Since among the proposed protocols [Bu et al., 2011], T-MTP

and ET-MTP are the most time-efficient and the most energy efficient, respectively, we compare D-MTP's time efficiency with T-MTP's and D-MTP's energy efficiency with ET-MTP's. Our analysis is convincing to show that D-MTP is more time-efficient than T-MTP and has chances of being more energy-efficient than ET-MTP.

We first analyze D-MTP's time efficiency with respect to the execution time. Recall that we use D-MTP when we ensure that no system update could induce false positives. Under the same scenario with a certain amount of misplaced tags, D-MTP and T-MTP take similar time to locate them. We only need to analyze the time for D-MTP to detect misplaced tags if any in each category—the same result as that of T-MTD (Section 2.4.2). We denote the time for D-MTP and T-MTP to detect misplaced categories as T'_{D-MTP} and T'_{T-MTP}, respectively. We have derived in Section 2.4.4 that

$$T'_{T-MTP} = \sum_{k=1}^{u} \sum_{i=1}^{m} (t_{cid} + t_{10b}) = um(t_{cid} + t_{10b}).$$

Next we will derive T'_{D-MTP} and prove it less than T'_{T-MTP}.

The time T'_{D-MTP} consists of two parts, the time for detecting whether misplaced categories exist and the time for verifying which categories they are if any. To detect whether misplace categories exist within a reader r_i's coverage, the reader r_i queries by broadcasting all category IDs in $\hat{C}_i$ (Equation 2.5). Upon receiving the query message, tags not in the broadcasted categories respond with a 10-bit random string. The reader r_i then determines that it covers no misplaced tag if it receives an empty slot, one misplaced tag if a singleton slot, or multiple misplaced tags if a collision slot. Then in the first step, the time for r_i to detect the existence of misplaced tags is therefore $|\hat{C}_i|t_{cid} + t_{10b}$. To facilitate the analysis of the time for distinguishing the

categories of misplaced tags, we introduce p_i and define it as follows:

$$p_i = \begin{cases} 0, \text{ if } r_i \text{ receives an empty slot,} \\ 1, \text{ if } r_i \text{ receives a non-empty slot.} \end{cases}$$

In the second step, only the readers with $p_i = 1$ continue to determine the exact categories that misplaced tags belong to. A simple method is that r_i queries through broadcasting one category ID in $C - \hat{C}_i$ after another. Responses corresponding to any category ID reveals that some tags in this category are misplaced. For ease of determining the number of misplaced tags, we still use 10-bit responses. Using this method, it takes $|C - \hat{C}_i|(t_{cid} + t_{10b})$ for r_i to determine all the misplaced categories covered by r_i.

Combining the time in the above two steps, we therefore have

$$\begin{aligned} T'_{D-MTP} &= \sum_{i=1}^{m}(|\hat{C}_i|t_{cid} + t_{10b} + p_i|C - \hat{C}_i|(t_{cid} + t_{10b})) \\ &\leq \sum_{i=1}^{m}(|\hat{C}_i|t_{cid} + t_{10b} + |C - \hat{C}_i|(t_{cid} + t_{10b})) \\ &= \sum_{i=1}^{m}(|C|(t_{cid} + t_{10b}) - (|\hat{C}_i| - 1)t_{10b}) \\ &< \sum_{i=1}^{m}(|C|(t_{cid} + t_{10b})) = T'_{T-MTP}. \end{aligned}$$

We then analyze D-MTP's energy efficiency with respect to the number of tag responses. Under the same scenario with a certain misplaced tags, D-MTP, T-MTP, and ET-MTP require similar number of tag responses to locate them. Therefore, similar to the analysis of D-MTP's time efficiency, we only consider N'_{D-MTP}, the number of tag responses for D-MTP to detect misplaced categories. Given the misplacement ratio α, it is straightforward that

$$N'_{D-MTP} = 2\alpha n,$$

because each misplaced tag needs to respond twice—one to the query of existence of misplaced tags and the other to the query of which are misplaced categories. Let N'_{T-MTP} and N'_{ET-MTP} denote the number of tag responses for T-TMP and ET-MTP to detect misplaced categories, respectively. Retrospecting to the performance analysis of T-MTP (Section 2.4.4) and ET-MTP (Section 2.5.4), we have

$$N'_{T-MTP} > n,$$
$$N'_{ET-MTP} = \sum_{k=1}^{u} p_k n_k > \frac{n}{m} \sum_{k=1}^{u} p_k.$$

Therefore, by solving $N'_{D-MTP} < N'_{T-MTP}$ and $N'_{D-MTP} < N'_{ET-MTP}$, we conclude that

- if $\alpha < 0.5$, D-MTP is more energy-efficient than T-MTP;
- if $\alpha < \frac{1}{2m} \sum_{k-1}^{u} p_k$, D-MTP is even more energy-efficient than ET-MTP.

2.8 Discussion

2.8.1 Tag Mobility

This section discusses the MTP problem in the presence of tag mobility. If mobile tags exist, we cannot simply apply the preceding protocols but need to distinguish misplaced tags from mobile tags, which could be carried by roaming machines or wandering customers. The intuition is to first track detected misplaced tags for a while and then analyze their location traces. If a detected misplaced tag hovers around the same location within a certain time interval, it is likely to be misplaced. If a detected misplaced tag moves from place to place within the time interval, it is likely to be mobile.

One natural approach to obtaining detected misplaced tags' traces is using multi-round detection. Using one of the preceding protocols (e.g., B-MTP, T-MTP, ET-MTP, or D-MTP), we run it several times to obtain location traces of detected misplaced tags. Assume that we need x locations to verify whether a detected misplaced tag is really misplaced or mobile. Let $l_i(t_j)$ ($i \in [0, x-1]$) denote the $(i+1)$th location in a tag t_j's location trace. Then we can distinguish misplaced tags from mobile tags using the following condition:

$$\frac{\sum_{i=0}^{x-1} |l_i(t_j) - l_0(t_j)|}{x} < \delta, \tag{2.6}$$

where $|l_i(t_j) - l_0(t_j)|$ is the distance between locations $l_i(t_j)$ and $l_0(t_j)$, and δ is a threshold that can be set according to localization accuracy. If the condition in Equation 2.6 is satisfied, t_j is likely to be stationary and therefore really misplaced. Otherwise, t_j is mobile and could be carried by machines or people roaming through the system.

Certainly tag mobility places a heavy overhead burden on MTP solutions, especially when many mobile tags move frequently. We therefore do not encourage pinpointing misplaced tags with mobile tags available unless the necessity outweighs the cost. More thorough investigation of tag mobility is left for our future work.

2.8.2 Channel Reliability

This section discusses the impacts of channel errors, packet loss on the proposed protocols, and suggests countermeasures against the potential impacts.

Both channel errors and packet loss may induce false negatives to the proposed protocols. For B-MTP, false positives due to channel errors occur when (1) queries from readers to tags are interfered and tags cannot determine whether to respond; and

(2) tags normally respond but responses from tags to readers are interfered. In both cases, some tags cannot successfully communicate with readers and thus avoid being detected if they are misplaced. For T-MTP, ET-MTP, and D-MTP, false negatives due to channel errors occur only when queries from readers to tags are interfered and tags cannot determine whether to respond. False negatives induced by packet loss are easier to infer—if queries from readers to misplaced tags or responses from misplaced tags to readers lost, we could hardly detect those misplaced tags and thus encounter false negatives.

To guarantee channel reliability for RFID communication, it is common to use a transmission power level high enough to drown the background noise. Certainly higher transmission power causes more energy consumption. Yet, packet loss is more challenging and may not be addressed by solely using a high transmission power level. We may have to resort to multi-round detection for guaranteeing detection accuracy at the cost of reduced time efficiency. When adopting the above countermeasures, the proposed energy- and time-efficient protocols (e.g., T-MTP, ET-MTP, and D-MTP) become more favorable than B-MTP and the state-of-the-art RPCV [Ferreira Chaves et al., 2010].

2.9 Summary

We have studied efficient MTP solutions against misplacement errors, a major concern in production economics due to their serious impact on profit. Departing from previous research that collects a large amount of data, this chapter investigates efficient MTP solutions from the perspective of communication protocol design.

We propose a series of protocols toward efficient MTP solution in large RFID systems, even in a distributed manner and robust against tag mobility. T-MTP detects misplaced tags based on reader vectors instead of tag vectors. It yields significantly increased time efficiency and energy efficiency, compared with basic solutions based on tag-wise positioning. ET-MTP caters for the trend of applying more and more popular active tags with self-equipped batteries. In favor of energy saving, ET-MTP requires only a fraction of tags to respond. To address the MTP problem in a distributed manner, D-MTP enables each reader to independently detect misplaced tags. D-MTP is more time-efficient than T-MTP and even could be more energy-efficient than ET-MTP. Analysis and experiments validate that the proposed protocols outperform the state of the art in both time efficiency and energy efficiency, which are important to guarantee protocol scalability in large RFID systems. Finally, we further discuss how to apply the proposed protocols in scenarios with mobile tags.

Our future work lies in the following three directions. First, we now only consider sequential reading. As we discussed in Section 2.6.4, parallel reading can yield higher time efficiency than can sequential reading. A promising topic is thus to adjust existing multi-reader scheduling protocols (e.g., in [Zhou et al., 2007]) to the MTP problem or even to design a new scheduling method that fits in better. Second, the positioning accuracy of the scheme in [Wang et al., 2007] may not satisfy requirements of certain applications. Inspired by the proliferation of sensor network localization [Bu et al., 2012b, Liu et al., 2010], we could borrow some ideas therein to improve tag positioning accuracy. Third, although evaluating research on large-scale RFID systems depends primarily on simulation nowadays, we urge our future work to evaluate and refine the proposed protocols in real RFID systems.

Chapter 3

Efficient and Privacy-Preserving Detection of Replication Attacks in Large RFID Systems

Replication attacks threaten Radio-Frequency Identification (RFID) applications and are hard to prevent. Existing replication attack detection protocols are limited in accuracy, efficiency, or even privacy. In this chapter, we propose a series of protocols toward efficient and privacy-preserving replication attack detection with guaranteed accuracy in large-scale RFID systems. The proposed protocols do not resort to complex cryptography techniques, inefficient tag-wise scanning, or privacy-unaware transmission of tag IDs. Instead, our protocols leverage the broadcast nature and collisions to detect replication attacks, being affordable to off-the-shelf low-cost tags. Toward efficient detection of replication attacks for large-scale RFID systems, we propose introducing two light-weight operations, vector broadcast and slot index recalculation. Armed with these two operations, our protocols can avoid unnecessary execution time and tag responses, and thus harvest significant gains in both time efficiency and energy efficiency. We evaluate the performance of the proposed protocols through theoretical analysis and extensive simulations. The results show that, when

the confidence level is 0.99 and the tolerance ratio of compromised tags is 0.001, our best protocol outperforms the state-of-the-art tag-wise scanning based protocol in time efficiency and energy efficiency by 98.5% and 72.8% on average, respectively.

3.1 Overview

Radio Frequency Identification (RFID) systems are vulnerable to various security attacks, mostly due to the hardware constraints of low-cost tags and the broadcast nature of wireless communication [Huang and Kapoor, 2009, Juels, 2006, Weis et al., 2004]. The *replication attack*, also known as cloning attack, is one of the most challenging security threats to RFID applications [Juels, 2006, Weis et al., 2004]. Launching a replication attack, an attacker compromises genuine tags and produces their replicas, namely *replicated tags*. Since replicated tags carry copies of compromised genuine tags' data (e.g., IDs and keys), they behave exactly the same as genuine tags and therefore threaten RFID applications that use the genuineness of tags to validate the authenticity of tagged objects [Bolotnyy and Robins, 2007, Lehtonen et al., 2009b]. For example, attached with replicated tags, products in RFID-enabled supply chains cause financial losses [Koh et al., 2003], healthcare facilities in RFID-aided hospitals jeopardize personal safety [Janz et al., 2005], while RFID-incorporated passport cards may even threaten national security [Koscher et al., 2009].

Replication attacks in RFID systems are, however, hard to prevent. Existing replication attack prevention approaches [Abawajy, 2009, Bolotnyy and Robins, 2007, Devadas et al., 2008, Dimitriou, 2006] use complex cryptography and encryption techniques, require additional hardware resources and key management strategies, and therefore are not affordable to most off-the-shelf low-cost tags [Lehtonen et al., 2009b,

Sarma, 2006, Spiekermann and Evdokimov, 2009]. Moreover, no prevention approach claims to completely defeat replication attacks yet [Lehtonen et al., 2009b]. Even if a break-through prevention approach arrives in the near future, it is not practical either to replace off-the-shelf tags with new tags or to recall them for upgrade—already 1.3 billion tags were in the market in 2005, and even 33 billion were expected in 2010 [web, c].

A few countermeasures against replication attacks turn to detecting them in RFID systems, focusing on two complementary application scenarios. In the first application scenario, tagged objects are distributed across multiple RFID systems. Existing detection protocols focusing on this scenario aim to secure RFID-enabled supply chains using *tag traces* [Koh et al., 2003, Lehtonen et al., 2009a, Mirowski and Hartnett, 2007, Zanetti et al., 2010b]. A tag trace comprises tag related data (e.g., ID, ownership, and location) distributed among supply chain partners [Zanetti et al., 2010b]. Redundant data corresponding to tag IDs help reveal replication attacks: When, for example, a tag with a certain ID simultaneously shows up at different places, the ID associates with replicated tags [Koh et al., 2003]. However, the detection accuracy of tag trace based protocols is hard to guarantee due to incomplete tag traces [Lehtonen et al., 2009a]; this issue becomes more challenging when some partners refuse to share the owned data due to business concerns [Zanetti et al., 2010b].

In the second application scenario, tagged objects are confined in the same RFID system. A recent detection protocol focusing on this scenario leverages tags' rewritable memory [Lehtonen et al., 2009b]. In this protocol, the reader writes a new random number on a tag's memory each time it scans the tag. A map of tag IDs and corresponding random numbers is maintained on a backend server. The reader then

detects the replication attack whenever any pair of the ID and random number is not identical with that in the map. Unlike tag trace based detection protocols [Koh et al., 2003, Lehtonen et al., 2009a, Mirowski and Hartnett, 2007, Zanetti et al., 2010b], the protocol in [Lehtonen et al., 2009b] can guarantee the detection accuracy. However, this protocol requires tag-wise scanning, which is known to be impractical and inefficient for a large-scale RFID system accommodating tens of thousands of tags [Chen et al., 2011, Kodialam and Nandagopal, 2006, Qian et al., 2010, Qiao et al., 2011, Zheng and Li, 2011]. Moreover, this protocol requires the transmission of tag IDs and thus may induce privacy leakage for applications that cast privacy-sensitive information into tag IDs [Han et al., 2010, Kodialam et al., 2007, Zhang et al., 2011a]. Even though the next generation of tags are expected to afford complex cryptographic operations for privacy protection, we have to face a new dilemma between the privacy and the efficiency in large-scale RFID systems [Lu et al., 2010].

In this chapter, we seek to efficiently detect replication attacks with guaranteed accuracy for applications that confine tagged objects in a large-scale RFID system. Consider, for example, RFID-based entrance control systems in widespread use [Finkenzeller et al., 2010]. For a private meeting or gathering that requires the credibility of each attendee, such systems alone cannot fully guarantee that an attendee is trustable as the attendee may hold a card embedded with a replicated tag. In this case, we need protocols that can quickly and accurately detect the existence of replicated tags, yet without leaking attendees' personal information probably casted into tag IDs. Furthermore, such protocols may benefit also applications that distribute tagged objects across multiple RFID systems. Imagine an RFID-enabled supply chain as a network,

and genuine tags, replicated tags as normal traffic, attack traffic, respectively. Borrowing the idea of filtering attack traffic at the source [Mirkovic et al., 2002], if we could locate the source of replicated tags in a supply chain [Koh et al., 2003], we can detect and reject replicated tags before they flood the market.

Toward efficient, accurate, and privacy-preserving detection of replication attacks, we propose a series of protocols that can secure applications confining tagged objects in a large-scale RFID system. Without resorting to complex cryptography techniques, inefficient tag-wise scanning, and privacy-unaware transmission of tag IDs, the proposed protocols achieve the goal through benefiting from three major novelties. First, we take advantage of broadcast and collisions for replication attack detection. A collision occurs when multiple tags respond to the reader simultaneously. Intuitively, when we specify only one tag with a certain ID to respond, there must exist replicas of the tag if a collision occurs. Second, we eliminate the need for the transmission of tag IDs. This elimination not only preserves identity privacy but also increases time efficiency. Third, we further enhance time and energy efficiency by requiring only particular tags to respond in consecutive time slots. The intuition is that only when multiple responses are received while only one is expected we can detect replication attacks, whereas other scenarios (e.g., zero or multiple responses are expected) bring no benefit but waste of time and energy.

We highlight the contributions of this chapter to replication attack detection for large-scale RFID systems as follows.

- Detect replication attacks with guaranteed accuracy. We formulate the problem as detecting the replication attack with a probability greater than or equal to a confidence level when the ratio of compromised tags is larger than a tolerance

level. The confidence and tolerance level can be conveniently adjusted according to system requirements.

- Leverage broadcast and collisions for replication attack detection. We accordingly propose a sampling-based protocol called BASIC (Section 3.3) and demonstrate its performance limitations to indicate the demand of privacy-preserving yet more efficient protocols.

- Propose a privacy-preserving protocol called RADar (Section 3.4) that eliminates the transmission of tag IDs. The primary idea is that the reader and tags perform the same hash operation on tag IDs to get expected and actual distribution of the number of responses. RADar detects the replication attack through comparing the distributions.

- Propose an efficient and privacy-preserving protocol called ET-RADar (Section 3.5) that can save unnecessary execution time and tag responses, yielding much higher time efficiency and energy efficiency than do BASIC and RADar. ET-RADar enforces tags choosing only expected singleton slots to respond consecutively by introducing two lightweight operations, namely vector broadcast and slot index recalculation.

- Further introduce a protocol adaptation to improve the efficiency of RADar and ET-RADar (Section 3.7.3). Simulation results show that, without leaking tag IDs, our best protocol averagely yields 98.5% higher time efficiency and 72.8% higher energy efficiency than does the state-of-the-art tag-wise scanning based protocol.

The rest of the chapter is organized as follows. Section 3.2 defines the problem and system. Section 3.3 presents a sampling-based protocol and analyzes its limitations in efficiency and privacy. Sections 3.4 and 3.5 present a privacy-preserving protocol and its efficient version, respectively. Section 3.6 discusses potential concerns, countermeasures and summarizes the proposed protocols. Section 3.7 reports simulation results. Finally, Section 3.8 concludes the chapter and indicates future work.

3.2 Problem Statement

In this section, we first provide an overview of the replication attack detection problem. We then discuss performance metrics for evaluation.

3.2.1 Problem Formulation

Consider an RFID system that consists of a backend server, some reader(s), a large number of tags each attached to an object [Lehtonen et al., 2009b]. The genuineness of tags are used to validate the authenticity of tagged objects. Without replication attacks, each tag has a unique ID. Tag IDs are stored on the server; readers communicate with the server via a secure link and have granted access to tag IDs [Lehtonen et al., 2009b]. Without loss of generality, we regard the backend server and readers as a whole and call it the "reader". When multiple readers are synchronized, we can logically treat them as one [Li et al., 2010a, Zheng and Li, 2011]. We are concerned with the replication attack through which an attacker compromises genuine tags, obtains all valid information of them (e.g., IDs and keys), and uses their replicas to impersonate genuine tags. We denote the confidence level and the tolerance ratio of compromised tags by α and λ, respectively, where $0 \leq \alpha, \lambda \leq 1$. The problem is to

detect the replication attack with a probability of at least α when compromised tags are more than λn, where n denotes the number of genuine tags.

The insensitivity to the number of replicas per compromised tag is a primary virtue of the above problem formulation. By insensitivity we mean that detection protocols based on our problem formulation can detect replication attacks with a required probability regardless of how many replicas of a compromised tag are produced. Let us clarify why this insensitivity matters through playing an interesting game. In the game, we are required to find two people resembling each other facially, in a crowd including multiple births. It is not hard to infer that the hardest scenario is when there are only twins hiding in the crowd. Triplets, or even quadruplets certainly make the game easier and easier. Coming back to the replication attack detection problem, a sophisticated attacker must know the game better than we do. To avoid the exposure, the attacker will try to compromise as many tags as possible, but produces a limited number of replicas using each compromised tag. Fortunately, the problem formulation we adopt is sensitive to exactly the ratio of compromised tags, regardless of how many replicas per compromised tag.

For now, we assume an error-free channel to ease understanding of our protocol designs. After delivering the design details, we will discuss the impact of channel errors on detection accuracy and suggest countermeasures in Section 3.6.1.

3.2.2 Performance Metrics

We evaluate detection accuracy and efficiency of the proposed protocols. The metric for evaluating detection accuracy is the detection probability. Time efficiency and energy efficiency are another two most important criteria for evaluating RFID protocols. Typical metrics for evaluating time efficiency and energy efficiency are the

execution time and the number of tag responses, respectively. We further introduce the concepts of time utility and energy utility to quantify the protocol efficiency in Section 3.3.2.

- The execution time shows time efficiency, which is highly desired for a protocol to detect the replication attack as fast as possible and to be scalable as the system scale increases. Most RFID protocols, ranging from traditional tag-cardinality estimation [Kodialam and Nandagopal, 2006, Qian et al., 2011] to recent missing-tag identification [Li et al., 2010a, Zhang et al., 2011b], consider time efficiency as a primary concern.

- The number of tag responses indicates energy efficiency, which is important when active tags are used. Active tags can significantly impulse the growth of RFID applications because of their ability of initiating communication and hundreds-of-feet communication radius, which is much longer than that of passive tags. However, active tags depend on self-equipped batteries to enable any operation. Thus, during enjoying the improved system performance brought by active tags, the energy cost should be as low as possible by controlling the number of tag responses. A few energy-efficient proposals for some other problems in RFID systems accommodating active tags are, for example, tag polling [Qiao et al., 2011], misplaced-tag pinpointing [Bu et al., 2011, 2012a], and cardinality estimation [Li et al., 2010b]. In [Bu et al., 2011] and [Bu et al., 2012a], we trade time efficiency for energy efficiency by introducing a probabilistic response technique. In this chapter, however, we will propose techniques that can simultaneously gain energy efficiency and time efficiency (Section 3.5 and Section 3.7.3).

- Time utility and energy utility measure the ratio of useful cost to total cost (Section 3.3.2). Low cost together with high utility is our goal of an efficient protocol. If most of the cost cannot benefit replication attack detection, even if the cost is low, there is still a need for efficiency improvement.

3.3 BASIC: Sampling-based Replication Attack Detection Protocol

In this section, we present a sampling-BASed replICation attack detection protocol (*BASIC*) and analyze its performance and limitations. We consider BASIC as a baseline for evaluating the proposed protocols (Section 3.7).

3.3.1 BASIC Design

The primary idea of BASIC is that replicated tags must exist if the reader receives multiple responses for confirming the presence of a certain tag. Generally, multiple responses cause a collision and the reader cannot correctly receive the responses. BASIC, however, can leverage a collision, which is exactly the evidence of multiple responses.

Leveraging the broadcast nature and collisions, we design BASIC as follows. BASIC samples tag IDs one after another without replacement. After sampling an ID, the reader broadcasts a query message containing the sampled ID and waits for tag responses. Upon receiving the query message, tags with the contained ID respond to the reader. In the Philips I-CODE system [web, b], a 10-bit string with error-detection (e.g., CRC) embedded is enough to verify a collision [Kodialam and Nandagopal, 2006]. If a collision occurs, the reader successfully detects the replication attack.

Otherwise, the reader ensures that the sampled ID associates with no replicated tag and continues by sampling another ID. Next we will analyze the maximum number of tag IDs to sample for satisfying the confidence level α.

3.3.2 Analysis

Let $P_B(n, x, k)$ denote the probability of BASIC detecting the replication attack by sampling no more than k IDs, when the ratio of compromised tags is x. Because the execution time is proportional to k, we formulate the problem as

$$\begin{aligned} &\text{minimize } k \\ &\text{subject to } \forall x > \lambda,\ P_B(n, x, k) \geq \alpha. \end{aligned} \tag{3.1}$$

Theorem 3.1. *Given n, x, and k,*

$$P_B(n, x, k) = \begin{cases} 1 - \prod_{i=1}^{k}(1 - p_i), & \text{if } 1 \leq k \leq (1-x)n, \\ 1, & \text{if } (1-x)n < k \leq n, \end{cases}$$

where $p_i = \frac{xn}{n-i+1}$.

Proof: We first derive the probability of BASIC failing to detect the replication attack, denoted as $P'_B(n, x, k)$. BASIC fails to detect replicas if there is no replicated ID contained in the k samples. Given xn replicated IDs and $(1-x)n$ nonreplicated IDs, there must be at least one replicated ID among more than $(1-x)n$ samples. Therefore,

$$P'_B(n, x, k) = 0,\ ((1-x)n < k \leq n).$$

For $1 \leq k \leq (1-x)n$, BASIC fails to detect the replication attack if all k samples

are nonreplicated IDs. Thus,

$$P'_B(n,x,k) = \prod_{i=1}^{k} \frac{(1-x)n-i+1}{n-i+1}$$
$$= \prod_{i=1}^{k}(1-\frac{xn}{n-i+1}),\ (1 \le k \le (1-x)n).$$

By the basic probability knowledge, we then have

$$P_B(n,x,k) = 1 - P'_B(n,x,k)$$
$$= \begin{cases} 1-\prod_{i=1}^{k}(1-p_i), & \text{if } 1 \le k \le (1-x)n, \\ 1, & \text{if } (1-x)n < k \le n, \end{cases}$$

where $p_i = \frac{xn}{n-i+1}$. ■

Theorem 3.2. *If we set*

$$P_B(n, \lambda + \frac{1}{n}, k) \ge \alpha,$$

BASIC can satisfy the accuracy constraint in Formula 3.1.

Proof: By Theorem 3.1, $P_B(n,x,k)$ is a monotonically increasing function of x. (The intuition behind this monotonicity is that more compromised tags yield higher probability of being detected.) The first possible value for $x > \lambda$ is $x = \lambda + \frac{1}{n}$, because $\lambda n + 1$ compromised tags correspond to the minimum number beyond tolerance (i.e. λn). If we set $P_B(n, \lambda + \frac{1}{n}, k) \ge \alpha$, we can therefore guarantee $P_B(n,x,k) \ge \alpha$ for $x > \lambda$ according to its monotonicity. ■

By Theorem 3.2, the maximum number k_{max} of tag IDs to sample toward the objective of minimizing the execution time, is as the following:

$$k_{max} = \min\{k \mid P_B(n, \lambda + \frac{1}{n}, k) \ge \alpha\}. \tag{3.2}$$

Given k_{max} by Equation 3.2, we derive the time and energy cost of BASIC as follows. Let t_{id} and t_c denote the transmission time of a tag ID and a 10-bit string

for verifying a collision, respectively. BASIC takes $t_{id} + t_c$ execution time and one response from a genuine tag per sampling operation. We do not consider energy cost by responses from replicas. Therefore, the execution time and the number of responses of BASIC, denoted as T_B and N_B, respectively, are

$$T_B \leq k_{max}(t_{id} + t_c),\ N_B \leq k_{max}.$$

To further quantify time efficiency and energy efficiency, we define *time utility* and *energy utility* as follows.

Definition 3.1. *Time (or energy) utility, denoted as U_t (or U_e), is the ratio of time (or responses) that is (or are) useful for replication attack detection to total time (or responses).*

Remark 3.1. *BASIC can achieve $U_t = 1$ and $U_e = 1$.*

Remark 3.1 is relatively straightforward, because each time BASIC samples an ID it can use the response(s) for replicas detection. BASIC detects replicas with the sampled ID if a collision occurs. Thus, all k' ($\leq k_{max}$) samples can benefit BASIC, yielding $U_t = \frac{k'(t_{id}+t_c)}{k'(t_{id}+t_c)} = 1$. Likewise, we can derive $U_e = \frac{k'}{k'} = 1$ because there is only one response from a genuine tag per sampling operation.

3.3.3 Limitations

We now discuss BASIC's limitations in privacy and efficiency. First, BASIC has an inherent defect in security because of privacy leakage. BASIC needs to broadcast sampled IDs, which should be protected for applications that cast privacy-sensitive information into tag IDs [Han et al., 2010, Kodialam et al., 2007, Zhang et al., 2011a]. The attacker may overhear those broadcast IDs. The overheard IDs may also reveal some other sensitive information such as location, which can be deferred by where IDs

are overheard. As the goal of replication attack detection protocols is to secure RFID systems, it is not wise enough to induce additional security risks. Moreover, BASIC is limited in time efficiency because of the transmission of tag IDs. ID transmission is known to be time-consuming especially in large-scale RFID systems [Bu et al., 2011, 2012a, Chen et al., 2011, Kodialam and Nandagopal, 2006].

These limitations naturally raise the question of whether it is possible to design a privacy-preserving yet more efficient protocol. Another two protocols to be proposed shortly can definitely answer the question.

3.4 RADar: Privacy-Preserving Replication Attack Detection Protocol

In this section, we propose a privacy-preserving Replication Attack Detection protocol (*RADar*). We first provide an overview of RADar and then discuss the design details. Theoretical analysis and limitations will be presented as well.

3.4.1 Overview of RADar

RADar preserves identity privacy through excluding the transmission of tag IDs from the entire replication attack detection process. Tag IDs are contained neither in query messages nor tag responses. Specifically, tags are deterministically assigned to respond at a certain time according to their IDs. RADar detects the replication attack once multiple responses are received while only one is expected. Two categories of RFID protocols that can support RADar are *slotted Aloha* [Roberts, 1975] and *Tree-traversal* [Hush and Wood, 1998, Myung et al., 2007]. Because slotted Aloha is more efficient than Tree-traversal in large-scale RFID systems [Qian et al., 2010], we adopt

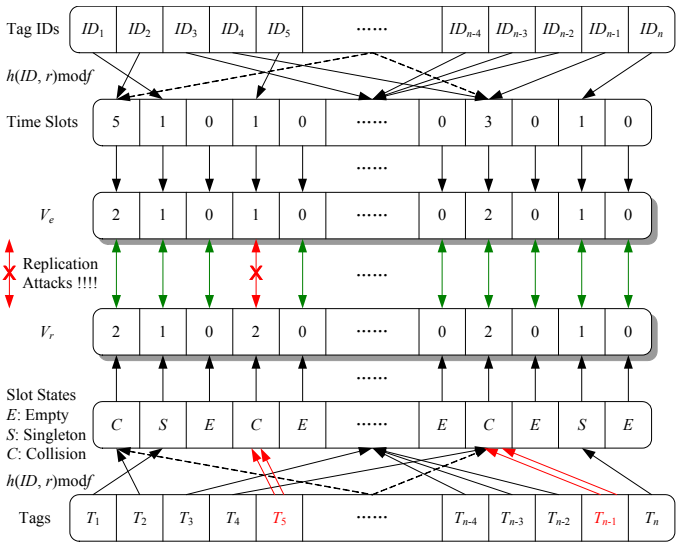


Fig. 3.1: Replication attack detection by RADar. T_i denotes a set containing all tags (a genuine tag and replicas if any) with ID_i. Dashed arrow-shaped lines indicate that one or more IDs (or Tags) are hashed to a time slot.

slotted Aloha to design RADar. In slotted Aloha, the reader sends a query message containing the number f of time slots (*frame size*) and a random seed r. Each tag picks up a time slot with index $h(ID, r) \bmod f$ to respond, where h is a hash function implemented on off-the-shelf tags. A time slot chosen by no tag, only one tag, or multiple tags is called an *empty slot*, a *singleton slot*, or a *collision slot*, respectively [Kodialam and Nandagopal, 2006]. *Empty*, *singleton*, and *collision* are called *slot states*.

Through verifying the slot states, RADar detects the replication attack if an *expected singleton slot* (i.e., a slot that is supposed to be singleton) turns out to be a collision slot. Because the transmission of tag IDs is not necessary, the time cost of RADar per time slot is only $\frac{t_c}{t_{id}+t_c} \approx 0.09$ [web, b] times that of BASIC, promising us a chance of significant time efficiency gains.

3.4.2 RADar Design

The preceding insight forms the basis for RADar. Figure 3.1 illustrates the sketch of RADar design. RADar detects replication attacks by comparing two vectors, the *expectation vector* V_e and the *response vector* V_r. Given f and r, we form V_e according to the number of tag IDs expectantly hashed to each time slot; we form V_r according to the number of tag responses actually received in each time slot. Then through comparing V_e and V_r element-wisely, RADar catches the replication attack as soon as any mismatching is found.

It is, however, challenging to determine the exact number of responses in a collision slot. Alternatively we form V_e and V_r as follows. The only useful scenario for RADar detecting the replication attack is when an expected singleton slot becomes a collision one. Two responses in a time slot are enough to cause a collision. Two or more responses make no difference if we are interested in the slot state of collision only. Therefore, given f, r, and known IDs, we define V_e and V_r to be

$$V_e[i] = \begin{cases} 0, \text{ if } |\{ID \mid h(ID,r) \bmod f = i\}| = 0, \\ 1, \text{ if } |\{ID \mid h(ID,r) \bmod f = i\}| = 1, \\ 2, \text{ if } |\{ID \mid h(ID,r) \bmod f = i\}| \geq 2. \end{cases} \tag{3.3}$$

$$V_r[i] = \begin{cases} 0, \text{ if slot } i \text{ is an empty slot}, \\ 1, \text{ if slot } i \text{ is a singleton slot}, \\ 2, \text{ if slot } i \text{ is a collision slot}. \end{cases}$$

Both V_e and V_r are f in length and $2f$ bits in size.

RADar then compares V_e and V_r element-wisely and detects the replication attack if

$$\exists i \in [0, f-1],\ V_e[i] \neq V_r[i].$$

Following the intrinsic property of probabilistic methods, RADar leads to no false positives but false negatives. We explain this by walking through scenarios when RADar can detect the replication attack and when it cannot. Taking ID_5 in Figure 3.1 for instance, it is hashed to the 4th time slot, which is chosen by no other IDs. Thus, $V_e[3] = 1$. There are replicas with ID_5 and thus the set T_5 contains more than one tags. Multiple responses from tags in T_5 will lead a collision to the 4th time slot. Therefore, $V_r[3] = 2 \neq V_e[3]$. RADar then successfully detects the replication attack that does exist, yielding no false positives. However, RADar cannot detect replicas with IDs that are originally hashed to slots with $V_e[i] = 2$. For example, ID_{n-1} in Figure 3.1 is hashed to the $(f-3)$th time slot, which is also chosen by some other IDs (e.g., ID_4). Thus, $V_e[f-4] = 2$. Although the set T_{n-1} contains one or more replicas with ID_{n-1}, RADar cannot detect them because $V_r[f-4] = 2 = V_e[f-4]$. Apparently, false negatives occur only when no replica has the tag ID with $V_e[i] = 1$. We can control the rate of false negatives through adjusting the confidence level α, that is, the higher α we set the fewer false negatives we will encounter.

3.4.3 Analysis

Let $P_R(n, x, f)$ denote the probability of RADar detecting the replication attack with frame size f, when the ratio of compromised tags is x. Because the execution time is proportional to f, we formulate the problem as

$$\begin{aligned} &\text{minimize } f \\ &\text{subject to } \forall x > \lambda,\ P_R(n, x, f) \geq \alpha. \end{aligned} \tag{3.4}$$

Theorem 3.3. *Given n, x, and f,*

$$P_R(n,x,f) \approx$$
$$1-\sum_{j=0}^{f}\binom{f}{j}(\frac{n}{f}\exp(-\frac{n}{f}))^j(1-\frac{n}{f}\exp(-\frac{n}{f}))^{f-j}(1-\frac{j}{f})^{xn}.$$

Proof: Given n tag IDs, and an f-slotted query frame, we denote by $Pr(n,f)$ the probability of a slot with $V_e[i]=1$. Because V_e corresponds to the distribution of IDs rather than tag responses, $Pr(n,f)$ is equal to the probability of any slot being singleton when there are no replicas. Therefore,

$$Pr(n,f)=\binom{n}{1}\frac{1}{f}(1-\frac{1}{f})^{n-1}\approx\frac{n}{f}\exp(-\frac{n}{f}). \tag{3.5}$$

Given f time slots, the number of slots with $V_e[i]=1$, denoted as N_1 ($\in[0,f]$), follows a binomial distribution. Thus, we define the probability of j slots with $V_e[i]=1$ to be

$$Pr(N_1=j\mid n,f)=\binom{f}{j}Pr(n,f)^j(1-Pr(n,f))^{f-j}.$$

As we discussed in Section 3.4.2, RADar can successfully detect the replication attack if at least one ID of compromised tags is hashed to a slot with $V_e[i]=1$. Therefore, we have

$$\begin{aligned}&P_R(n,x,f)\\&=1-\sum_{j=0}^{f}Pr(N_1=j\mid n,f)(1-\frac{j}{f})^{xn}\\&=1-\sum_{j=0}^{f}\binom{f}{j}Pr(n,f)^j(1-Pr(n,f))^{f-j}(1-\frac{j}{f})^{xn}.\end{aligned}$$

Substituting $Pr(n,f)$ from Equation 3.5, we derive $P_R(n,x,f)$ and therefore prove Theorem 3.3. ■

Theorem 3.4. *If we set*

$$P_R(n, \lambda + \frac{1}{n}, f) \geq \alpha,$$

RADar can satisfy the accuracy constraint in Formula 3.4. The optimal frame size f_{opt} *toward the objective of minimizing the execution time is*

$$f_{opt} = \min\{f \mid P_R(n, \lambda + \frac{1}{n}, f) \geq \alpha\}. \tag{3.6}$$

Proof: The proof of Theorem 3.4 is essentially the same as that of Theorem 3.2 (Section 3.3.2). We hereby simply skip the repetition for the sake of conciseness. ■

Given f_{opt} by Equation 3.6, we derive the execution time T_R and the number N_R of responses for RADar. Omitting the tiny transmission time of f_{opt} and r in the query message, we define T_R and N_R as

$$T_R = f_{opt} t_c, \ N_R = n.$$

Remark 3.2. *Given* n, f_{opt}, *and* N_1, *RADar can achieve* $U_t = \frac{N_1}{f_{opt}}$ *and* $U_e = \frac{N_1}{n}$.

We derive Remark 3.2 as follows. Recall that N_1 represents the number of slots with $V_e[i] = 1$. First, because only slots with $V_e[i] = 1$ are useful for replication attack detection, it is obvious that all N_1 slots out of f_{opt} slots can benefit RADar. Thus, the time utility of RADar is $U_t = \frac{N_1}{f_{opt}}$. Second, in each of these N_1 slots, there is only one response from a genuine tag. All tags, however, respond to the query message, yielding totally n responses from genuine tags. Therefore, the energy utility of RADar is $U_e = \frac{N_1}{n}$.

3.4.4 Limitations

Although RADar is privacy-preserving and yields faster detection than does BASIC, it is still limited in time efficiency and energy efficiency. By Remark 3.2, the time

utility of RADar is $U_t = \frac{N_1}{f_{opt}}$. Because N_1 is the number of slots with $V_e[i] = 1$, $\frac{N_1}{f_{opt}}$ corresponds to the ratio of singleton slots without considering replicas. The optimal ratio of singleton slots using slotted Aloha is only 36.8% [Chen et al., 2011]. Slots with $V_e[i] = 0, 2$ count most time slots of a query frame yet cannot benefit replication attack detection. Furthermore, the optimal f for satisfying the confidence level α is less than n by Theorem 3.4. Thus, the energy utility $U_e = \frac{N_1}{n} < \frac{N_1}{f_{opt}} \leq 36.8\%$; such indicates that energy inefficiency is even worse than time inefficiency. Solutions against these inefficiencies are highly desired, considering that larger and larger RFID systems have been applied [web, c].

3.5 ET-RADar: Energy- and Time-efficient RADar

In this section, we propose Energy- and Time-efficient RADar (*ET-RADar*). We first provide an overview of ET-RADar. We then expatiate on the protocol design and performance analysis.

3.5.1 Overview of ET-RADar

The primary idea of ET-RADar is to bypass all time slots that are not useful for RADar detecting the replication attack. So the execution time and tag responses induced by those bypassed slots can be saved. Because useful slots (i.e., slots with $V_e[i] = 1$) are still reserved, ET-RADar enjoys the same detection accuracy as that of RADar. In summary, ET-RADar is expected to yield faster detection and lower energy cost, without sacrificing the privacy and detection accuracy, when compared with RADar.

Current slotted Aloha, however, cannot directly support ET-RADar. As we discussed in Section 3.4.1, upon receiving a query message containing f and r, a tag determines to respond in a time slot with index $h(ID, r) \mathrm{mod} f$. There will be empty, singleton, and collision slots, even though only expected singleton slots (i.e., slots with $V_e[i] = 1$) are useful for replication attack detection. Because the reader is aware of the tag IDs mapped to expected singleton slots, it could just simply broadcast one of these IDs at a time and verify whether one or more responses will be received. Clearly, this intuitive solution degenerates to BASIC, reinducing time-consuming transmission of tag IDs as well as the privacy leakage issue.

We propose introducing two lightweight operations, *vector broadcast* and *slot index recalculation*, to adapt slotted Aloha to the ET-RADar design. Specifically, the reader broadcasts a vector to inform tags to respond if they are hashed to expected singleton slots or to keep silent otherwise. Then the tags informed to respond recalculate slot indices and respond in consecutive slots. By the first operation, ET-RADar saves energy through preventing unnecessary tag responses. By the second operation, ET-RADar reduces the execution time through leaving out empty slots. Therefore, we expect ET-RADar to outperform RADar in both energy efficiency and time efficiency. Next we will present the design details.

3.5.2 ET-RADar Design

We illustrate by Figure 3.2 the portrait of ET-RADar design. Note that the illustrated scenario is the same as that illustrated in Figure 3.1. There are replicas with ID_5 and ID_{n-1} among listed IDs. We deliberately repeat the example scenario for ease of comparison of ET-RADar and RADar.

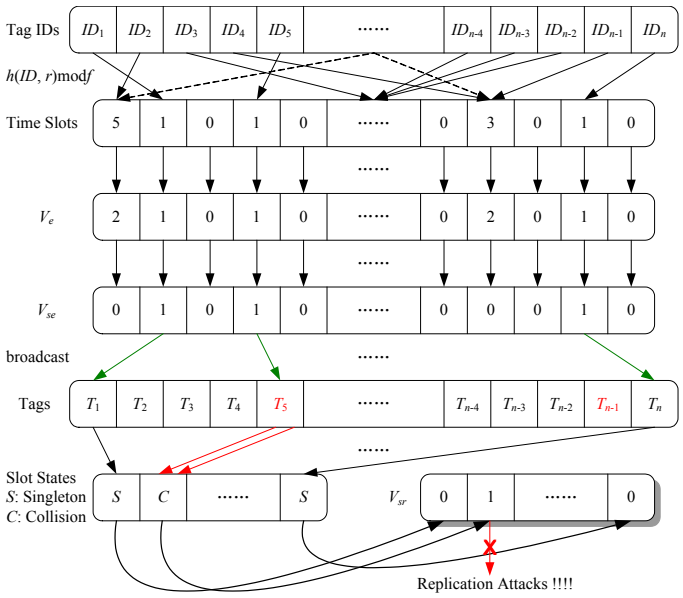


Fig. 3.2: Replication attack detection by ET-RADar. ET-RADar detects the replication attack once any $V_{sr}[i] = 1$ (e.g., $V_{sr}[1]$ as illustrated).

As shown in Figure 3.2, we first determine the optimal frame size f_{opt} by Equation 3.6, generate a random seed r, and form the expectation vector V_e by Equation 3.3. We further form a *simplified expectation vector*, denoted as V_{se}, as follows:

$$V_{se}[i] = \begin{cases} 0, \text{ if } V_e[i] \neq 1, \\ 1, \text{ if } V_e[i] = 1. \end{cases}$$

The vector V_{se} is f_{opt} in length and f_{opt} bits in size.

The reader then broadcasts a query message comprising f_{opt}, r, and V_{se}. Recall that RADar informs tags of only f_{opt} and r. Upon receiving the query message, a tag first determines the slot index by $i = h(ID, r) \text{mod} f_{opt}$. The tag, however, does not simply wait to respond in slot i. It further looks up the value of $V_{se}[i]$. If $V_{se}[i] = 1$, the tag decides to respond normally. Otherwise, the tag keeps silent. Following this process, ET-RADar avoids unnecessary tag responses and thus saves the energy.

Furthermore, because all slots with $V_{se}[i] = 0$ are ensured to be empty and are

not expected to provide any valuable information, we recalculate the slot index of a slot with $V_{se}[i] = 1$ to $i' = \sum_{j=0}^{i} V_{se}[j] - 1$. Through recalculating slot indices, all tags informed to respond will send responses in consecutive time slots. The responses are not interrupted by empty slots. Apparently, ET-RADar successfully skips all unnecessary tag responses and time slots by now. ET-RADar therefore yields higher energy efficiency and time efficiency than does RADar.

Receiving all responses, the reader forms a *simplified response vector*, denoted as V_{sr}, as follows:

$$V_{sr}[i] = \begin{cases} 0, \text{ if slot } i \text{ is a singleton slot,} \\ 1, \text{ if slot } i \text{ is a collision slot.} \end{cases}$$

The vector V_{sr} is N_1 in length and N_1 bits in size.

ET-RADar then checks the vector V_{sr} element-wisely and detects the replication attack if

$$\exists i \in [0, N_1 - 1],\ V_{sr}[i] = 1.$$

ET-RADar guarantees the same detection accuracy as that of RADar, yielding no false positives but false negatives. As shown in Figure 3.2, ET-RADar can successfully detect replicas with ID_5 but fails to detect those with ID_{n-1}. False negatives occur only when there are replicas but all the replicas are hashed to slots with $V_{se}[i] = 0$ (e.g., tags in the set T_{n-1} in Figure 3.2). Yet, we can control the rate of false negatives through adjusting the confidence level α.

3.5.3 Analysis

ET-RADar adopts Formula 3.4, Theorem 3.3, and Theorem 3.4 to formulate the problem, to guarantee the detection probability, and to determine the optimal frame size f_{opt}, respectively. We next primarily concentrate on the analysis of efficiency

gains brought by ET-RADar.

Given the optimal frame size f_{opt} by Equation 3.6, ET-RADar takes N_1 time slots and induces N_1 responses from genuine tags. The only extra time paid for upgrading RADar to ET-RADar is the transmission time of the f_{opt}-bit vector V_{se}. Let t_s represent the transmission time of a single bit. We conclude the performance of ET-RADar by Remark 3.3, Remark 3.4, and Remark 3.5. We omit the proof because they are easily digestible given the preceding interpretation and analysis.

Remark 3.3. *The execution time and the number of responses of ET-RADar, denoted as T_E and N_E, respectively, are given as follows:*

$$T_E = N_1 t_c + f_{opt} t_s, \ N_E = N_1.$$

Remark 3.4. *ET-RADar reduces the execution time of RADar by*

$$\begin{aligned} \frac{T_R - T_E}{T_R} &= \frac{f_{opt} t_c - (N_1 t_c + f_{opt} t_s)}{f_{opt} t_c} \\ &= \frac{f_{opt} t_c - (N_1 t_c + 0.1 f_{opt} t_c)}{f_{opt} t_c} \quad \textit{as } t_s = 0.1 t_c \\ &= 0.9 - \frac{N_1}{f_{opt}} \geq 53.2\%, \end{aligned}$$

and reduces the energy cost of RADar by

$$\begin{aligned} \frac{N_R - N_E}{N_R} &= \frac{n - N_1}{n} \\ &= 1 - \frac{N_1}{n} > 63.2\%. \end{aligned}$$

Remark 3.5. *ET-RADar can achieve $U_t = \frac{N_1 t_c}{N_1 t_c + f_{opt} t_s} \approx 1$ and $U_e = \frac{N_1}{N_1} = 1$.*

3.6 Discussion

In this section, we first discuss potential concerns (e.g., channel errors, manipulative attackers) and suggest countermeasures. We then summarize the proposed protocols.

3.6.1 Channel Errors

We now discuss the impact of channel errors on the proposed protocols and cope with it if any, although many RFID protocols presume an error-free channel and may not directly apply to an imperfect channel.

Channel errors can cause no false negatives but false positives. The proposed protocols detect the replication attack when the reader receives multiple responses while only one is expected. Superimposing noise on multiple responses guarantees that the reader still can verify a collision. Therefore, channel errors cannot cause any false negatives to the proposed protocols. However, a false positive occurs when there is no replication attack but noise together with a single response makes a collision. Using a certain high transmission power level to suppress the noise can avoid false positives caused by channel errors. This solution certainly puts a higher energy cost on active tags than protocols do in an error-free channel. In this case, ET-RADar is particularly more favorable than RADar, because ET-RADar avoids all unnecessary tag responses but RADar requires all tags to respond.

3.6.2 Tag Distribution

Both existing and the proposed protocols require the co-appearance of genuine and replicated tags to successfully detect the replication attack. Two complementary replication attack detection scenarios are thus with tags residing in the same RFID system or distributing across multiple RFID systems. The tag-wise scanning based protocol [Lehtonen et al., 2009b] and the proposed protocols (Sections 3.3-3.5) concentrate on the scenario where tags reside in the same RFID system. A concern naturally arises that tags may not be fully covered by the communication range of

only one reader. Multiple readers are thus necessary to guarantee full coverage for some large RFID systems. When multiple readers are well synchronized [Leong et al., 2006], they can be logically treated as one [Li et al., 2010a, Qiao et al., 2011, Zheng and Li, 2011]. In the case of tags distributing across multiple RFID systems, we turn to tag trace based protocols [Koh et al., 2003, Lehtonen et al., 2009a, Mirowski and Hartnett, 2007, Zanetti et al., 2010b] for replication attack detection. Each participating system may also deploy multiple readers to communicate with all present tags. As discussed in Section 3.1, if we could locate the source of replicated tags in, for example, a supply chain [Koh et al., 2003], we can use the proposed protocols to detect and reject replicated tags before they flood other RFID systems.

Another potential concern is that some tags may be deliberately hid at the time of detection. If this concern occurs and violates the co-appearance of genuine and replicated tags, both existing protocols and the proposed protocols fail to detect the replication attack. In this case, we have to resort to certain authority or organization for regulating an honest and full exposure of tagged objects at the time of detection. Such a regulation in, for example, a supply chain may issue proofs to tagged objects that pass the detection and thus prohibit the sale of those hid from the detection. Overall, we believe that armed with an accurate, efficient (or even privacy-preserving) replication attack detection protocol, regulations established by certain authority and organization will prevail over replication attacks in RFID applications.

3.6.3 Sophisticated Replicated Tags

Since both existing and the proposed protocols require the co-appearance of genuine and replicated tags to successfully detect the replication attack (Section 3.6.2),

sophisticated replicated tags may destroy the evidence of appearance through abnormal responses. Therefore, we should consider the cases when replicated tags emit extra responses and when they deliberately keep silent.

Sophisticated replicated tags emitting extra responses cannot decrease the accuracy of any proposed protocols and can even increase it. First, in BASIC (Section 3.3), when the reader broadcasts an ID that corresponds to some replicas, the reader will receive multiple responses and successfully detect the replication attack, whether or not any replicated tags deliberately emit additional responses. Second, in RADar (Section 3.4) and ET-RADar (Section 3.5), if a replicated tag emits extra responses in the time slot it chooses, certainly this will not affect detection accuracy. If a replicated tag deliberately emits extra responses in other time slots at least one of which is an expected singleton slot, it leads to the detection of the replication attack and thus actually helps increase detection accuracy.

Sophisticated replicated tags deliberately keeping silent, if succeeded, will avoid the exposure of replicated tags and induce false negatives. Being pessimistic, probably no detection protocol would catch replicated tags that do not send any responses. But being optimistic yet realistic, we can crack this issue based on the simplicity of low-cost tags and the very purpose of replication attacks as follows.

- The design of tag chips is not as complicated as that of sensors or of other more functional wireless devices [Finkenzeller et al., 2010]. Therefore, protocols for monitoring operations in RFID systems should also be simple enough to be affordable. The simplicity of tags and protocols leads to the similarity among various protocols. Some protocols may even be adaptable to other monitoring operations. For example, the tag-wise scanning based protocol [Lehtonen et al.,

2009b] and the proposed protocols for replication attack detection can easily identify tags or collect information by respectively querying tag IDs or stored information. This makes replicated tags hard to decide which query messages they can reply to participate monitoring operations and which ones they should not reply to avoid the exposure.

- The very purpose of replication attacks is using replicated tags to impersonate genuine tags during system monitoring operations. If replicated tags choose to keep silent, they cannot participate in monitoring operations and consequently fail the replication attack. For example, when the reader identifies tags attached to newly purchased products and registers them in the database, the reader can regard those tags that send no responses as invalid tags. The quality of products attached with invalid tags is highly suspicious and those products should not be on the shelf. In this case, replicated tags are rejected even before we run detection protocols.

3.6.4 Sophisticated Attackers

Sophisticated attackers may further challenge replication attack detection protocols as well as protocols for other system monitoring operations. First, a sophisticated attacker aware of the detection strategy may manipulate replicated tags via wireless communication and control their responses to prevent them from being detected (Section 6.3). Second, in a more general way, a sophisticated attacker may simply jam the communication between the reader and tags (with or without replicas included) to paralyze various system monitoring operations. Affected tags thus, for example, may not be identified or registered [Qian et al., 2010], may be mis-regarded as missing [Li

et al., 2010a], or may be unable to report stored information [Chen et al., 2011]. In all cases, affected tags are essentially revoked from the RFID system, although they do appear.

We have to admit that coping with such sophisticated attackers from the protocol design's point of view is very challenging. Fortunately, this challenging issue has been addressed by a recent pioneer physical layer solution called *shield* [Gollakota et al., 2011]. In short, the shield can detect and jam any signals from unauthorized readers (e.g., the aforementioned sophisticated attackers we concern) to tags without interfering the communication between authorized readers and tags. We believe that it is worth adopting the shield to combat sophisticated attackers whenever they may place serious impacts on an RFID system. Finally, we would like to refer the interested reader to [Gollakota et al., 2011] for more exciting details of the shield, yet this topic is beyond the scope of replication attack detection protocol design.

3.6.5 Distinguishing Genuine Tags from Replicas

As with any study on replication attack detection [Koh et al., 2003, Lehtonen et al., 2009a,b, Mirowski and Hartnett, 2007, Zanetti et al., 2010b], the proposed protocols cannot directly distinguish genuine tags from replicated tags. Since replicated tags copy all valid information of compromised genuine tags, they can pass any authentication as genuine tags can. Therefore, it is hard to distinguish a genuine tag from its replicated peers among multiple tags with the same ID. Against this issue, we have to resort to the physical architecture of tag hardware [Bolotnyy and Robins, 2007] or the authentication of tagged objects.

Table 3.1: Performance Comparison of BASIC, RADar, and ET-RADar

Performance	BASIC	RADar	ET-RADar
ID transmission	Yes	No	No
Privacy-preserving	No	Yes	Yes
Execution time	$\leq k_{max}(t_{id} + t_c)$;	$f_{opt}t_c$	$N_1t_c + f_{opt}t_s$
Number of responses	$\leq k_{max}$	n	N_1
Time utility	1	$\frac{N_1}{f_{opt}} \leq 36.8\%$;	≈ 1
Energy utility	1	$\frac{N_1}{n} < 36.8\%$	1

Denotations: k_{max} denotes the maximum k by Equation 3.2, and f_{opt} denotes the optimal f by Equation 3.6. N_1 represents the number of time slots with $V_e[i] = 1$ (by Equation 3.3) among f_{opt} time slots; $N_1 \leq 0.368 f_{opt}$ [Chen et al., 2011].
Notes: The tag-wise scanning based protocol in [Lehtonen et al., 2009b] requires the transmission of tag IDs and additional random numbers and thus may induce privacy leakage. The tag trace based protocols in [Koh et al., 2003, Lehtonen et al., 2009a, Mirowski and Hartnett, 2007], and [Zanetti et al., 2010b] concern another complementary scenario with tagged objects distributed across multiple places. All these protocols constitute two categories of replication attack detection for two complementary RFID application scenarios. The performance comparison of the tag-wise scanning based protocol and the proposed protocols is discussed in Section 3.7.4.

3.6.6 A First Comparison of Protocol Designs

In the preceding sections, we deliver protocol designs and performance analysis, and discuss potential concerns and countermeasures. We now summarize the proposed protocols by a simple comparison in Table 3.1, providing the big picture before we report detailed simulation results.

Table 3.1 summarizes the analysis results. If we prioritize the proposed protocols in descending order of the privacy and efficiency, the expected sequence is ET-RADar > RADar > BASIC. The baseline protocol BASIC, which needs the transmission of tag IDs, suffers from privacy leakage and time inefficiency, although its time utility and energy utility both achieve 1. RADar eliminates the transmission of tag IDs

and thus it is privacy-preserving and is more time-efficient than BASIC. Because privacy-preserving protocols are important to secure some RFID applications, we prefer RADar to BASIC, although RADar requires all tags to respond. ET-RADar significantly enhances the efficiency through avoiding all unnecessary execution time and tag responses, without sacrificing the accuracy and privacy. ET-RADar is expected to be superior to both BASIC and RADar. In the next section, we will evaluate the performance of the proposed protocols through extensive simulations.

3.7 Performance Evaluation

In this section, we evaluate the accuracy and efficiency of the proposed protocols by simulations. We further propose enhanced versions of RADar and ET-RADar, and evaluate efficiency gains brought by them. We compare our protocols against the state-of-the-art tag-wise scanning based protocol in [Lehtonen et al., 2009b]. The comparison results (Section 3.7.4) show that our best protocol averagely yields 98.5% higher time efficiency and 72.8 higher energy efficiency than does the tag-wise scanning based protocol. We, however, do not compare our protocols with tag trace based detection protocols [Koh et al., 2003, Lehtonen et al., 2009a, Mirowski and Hartnett, 2007, Zanetti et al., 2010b] because they cannot apply to scenarios confining tagged objects in the same RFID system as concerned in [Lehtonen et al., 2009b] and this chapter.

3.7.1 Environment Configuration

We simulate the system as follows. The reader communicates with tags with a power level high enough to drown the background noise. The number of genuine tags

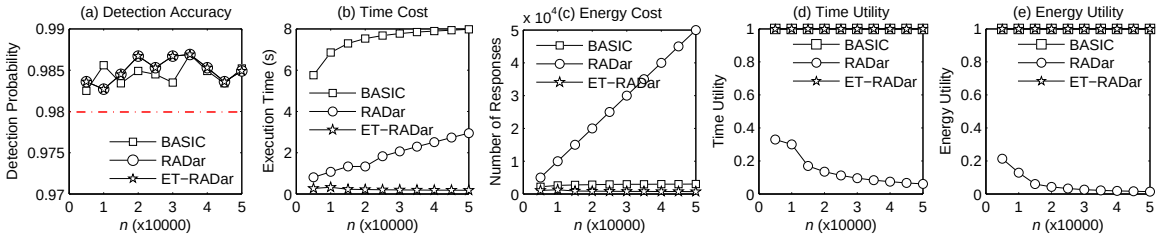


Fig. 3.3: Performance comparison of BASIC, RADar, and ET-RADar with $\alpha = 0.98$, $\lambda = 0.001$, and n varying from 5,000 to 50,000.

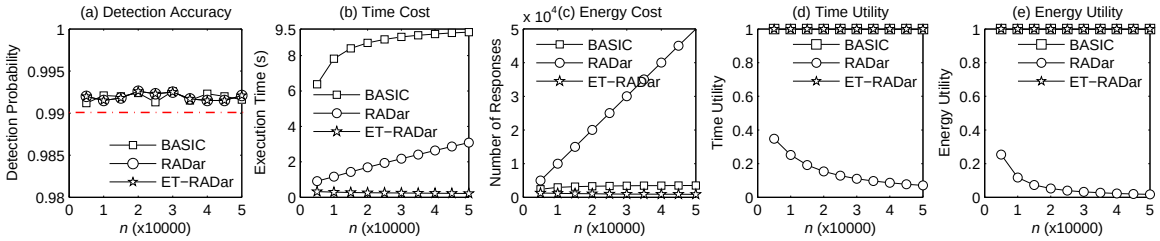


Fig. 3.4: Performance comparison of BASIC, RADar, and ET-RADar with $\alpha = 0.99$, $\lambda = 0.001$, and n varying from 5,000 to 50,000.

n varies from 5,000 to 50,000. The tag ID is 96 bits long [web, a]. Given the confidence level α and the tolerance ratio of compromised tags λ, we randomly choose $\lambda n + 1$ tags for the attacker producing replicas, which behave exactly the same as genuine tags. We run BASIC, RADar, and ET-RADar to detect the replication attack. A 10-bit response is used for verifying a collision [web, b]. We set the transmission time of a single bit to $t_s = 25$ µs [web, b]. This value may vary for different hardware, but does not affect the performance evaluation because we are interested more in the percentage of reduction in the execution time than we are interested in the execution time itself. As we discussed in Section 3.2.2, the performance metrics include the detection probability, the execution time, the number of responses from genuine tags, the time utility, and the energy utility.

3.7.2 Accuracy and Efficiency

Figure 3.3 and Figure 3.4 report the results under two scenarios with $\alpha = 0.98, 0.99$ for either and $\lambda = 0.001$ for both. Figure 3.3(a) shows that all protocols can satisfy the confidence level α with average detection probability 0.985. As we have expected, Figure 3.3(b)-(e) demonstrate that ET-RADar is more efficient than both BASIC and RADar. In Figure 3.3(b), ET-RADar decreases the execution time of BASIC and of RADar by 97.0% and 85.5% on average, respectively. When $n = 50{,}000$, for example, ET-RADar averagely takes 0.18 seconds, while BASIC and RADar averagely take 7.97 seconds and 2.94 seconds, respectively. In Figure 3.3(c), RADar requires all tags to respond, whereas BASIC and ET-RADar require a portion of them to respond. Specifically, ET-RADar induces 68.3% and 94.2% fewer responses than do BASIC and RADar, respectively. In Figure 3.3(d)-(e), both of BASIC and ET-RADar approach 100% cost utility, while RADar achieves the time utility and the energy utility of only 14.3% and 5.8% on average, respectively.

When $\alpha = 0.99$, as shown in Figure 3.4, all the proposed protocols require higher cost to satisfy the detection accuracy than they do when $\alpha = 0.98$. In summary, the proposed protocols detect replication attacks with average probability 0.992 (Figure 3.4(a)); ET-RADar outperforms the baseline protocol BASIC in time efficiency and energy efficiency by 96.9% (Figure 3.4(b)) and 68.0% (Figure 3.4(c)) on average, respectively. We conduct more simulations and the results are consistent.

A surprisingly interesting property of ET-RADar revealed by the results is that, given α and λ, ET-RADar takes less time and energy as n grows (Figure 3.3(b)-(c) and Figure 3.4(b)-(c)). By Remark 3.3, the cost of ET-RADar is proportional to N_1, which represents the number of expected singleton slots. ET-RADar can detect

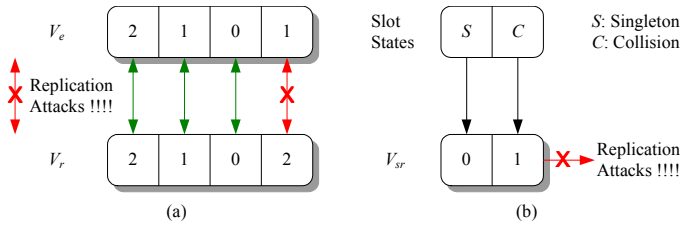


Fig. 3.5: Replication attack detection by (a) RADar-II and (b) ET-RADar-II, which are enhanced versions of RADar (Figure 3.1) and ET-RADar (Figure 3.2), respectively. RADar-II and ET-RADar-II detect replicas right after each time slot, and terminate when they detect the first replicated ID.

the replication attack if at least one replica responds in an expected singleton slot. Intuitively, the number of replicated IDs $\lambda n + 1$ increases as n increases. Therefore, a smaller N_1 could keep the probability of at least one of the N_1 slots being chosen by a replica. A more comprehensive explanation is as follows. Given f_{opt} and N_1, the probability of ET-RADar detecting the replication attack is $1 - (1 - \frac{N_1}{f_{opt}})^{\lambda n+1}$, where $0 < \frac{N_1}{f_{opt}} < 1$. When $\lambda n + 1$ increases, $1 - \frac{N_1}{f_{opt}}$ should also increase to keep $(1 - \frac{N_1}{f_{opt}})^{\lambda n+1}$ constant for satisfying the confidence level α. Therefore, $\frac{N_1}{f_{opt}}$ decreases with decreasing N_1 and increasing f_{opt} as n grows in the simulations.

3.7.3 Enhanced Versions of RADar and ET-RADar

The results reported in Section 3.7.2 demonstrate remarkable efficiency enhancement of ET-RADar over that of BASIC and RADar. Better news here to report is that a simple protocol adaptation can push its efficiency one more step toward the limit. Figure 3.5 illustrates the protocol adaptation with applying it to RADar and ET-RADar, under the scenarios shown in Figure 3.1 and Figure 3.2, respectively. The adapted protocols, namely *RADar-II* and *ET-RADar-II*, detect replication attacks

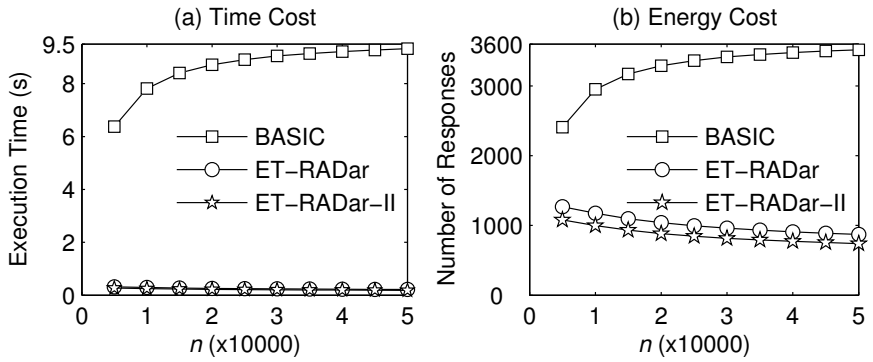


Fig. 3.6: Performance comparison of BASIC, ET-RADar, and ET-RADar-II with $\alpha = 0.99$, $\lambda = 0.001$, and n varying from 5,000 to 50,000.

per time slot and terminate after detecting the first replicated ID, rather than always wait for the entire frame. Therefore, they can further enhance time and energy efficiency without sacrificing detection accuracy.

For the ease of comparison, we report by Figure 3.6 the performance statistics of only BASIC, ET-RADar, and ET-RADar-II with $\alpha = 0.99$ and $\lambda = 0.001$. Omitting RADar and RADar-II, which are inferior to ET-RADar and ET-RADar-II, respectively, we are interested in the efficiency gap between our best protocol ET-RADar-II and the baseline protocol BASIC. ET-RADar-II averagely increases the efficiency of ET-RADar by 15.1%, yielding 97.4% higher time efficiency (Figure 3.6(a)) and 72.8% higher energy efficiency (Figure 3.6(b)) on average than does BASIC.

3.7.4 Comparison with Tag-wise Scanning

Finally, we compare the proposed protocols against the state-of-the-art tag-wise scanning based protocol (TWS) in [Lehtonen et al., 2009b]. We have reviewed the basic of TWS in (the fourth paragraph of) Section 3.1; we thus omit the repetition here. When implementing TWS, we use 32-bit random numbers as the authors do in experiments [Lehtonen et al., 2009b]. Other parameters follow the configuration in

Section 3.7.1. The performance metrics for comparison are the execution time and the number of tag responses, indicating time efficiency and energy efficiency, respectively.

Since ET-RADar-II is with the best performance among the proposed protocols (Section 3.7.3), we report by Figure 3.7 the performance statistics of only TWS and ET-RADar-II with $\alpha = 0.99$ and $\lambda = 0.001$. As we expect, ET-RADar-II outperforms TWS in both time efficiency (Figure 3.7(a)) and energy efficiency (Figure 3.7(b)). First, the superiority of ET-RADar-II over TWS in time efficiency is primarily attributed to the elimination of ID transmission. Recall that TWS transmits not only tag IDs but also additional random numbers (Section 3.1). Moreover, the elimination of ID transmission can preserve privacy for RFID applications that cast privacy-sensitive information into tag IDs [Han et al., 2010, Kodialam et al., 2007, Zhang et al., 2011a]. Second, the superiority of ET-RADar-II over TWS in energy efficiency can be explained by the similarity between TWS and BASIC. Both of TWS and BASIC are essentially based on sampling—TWS randomly chooses tags with certain IDs to scan while BASIC randomly chooses tag IDs to query tags with the chosen IDs. Therefore, ET-RADar-II outperforms TWS in energy efficiency in a comparable extent as it outperforms BASIC (Section 3.7.3).

In summary, ET-RADar-II averagely yields 98.5% higher time efficiency (Figure 3.7(a)) and 72.8% higher energy efficiency (Figure 3.7(b)) than does TWS.

3.8 Summary

The replication attack poses a significant threat to RFID applications but is hard to prevent. Motivated by limitations of existing replication attack detection protocols

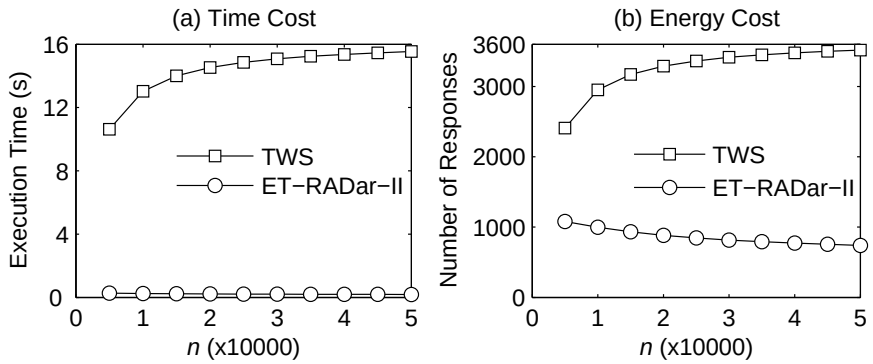


Fig. 3.7: Performance comparison of the state-of-the-art TWS and the proposed ET-RADar-II with $\alpha = 0.99$, $\lambda = 0.001$, and n varying from 5,000 to 50,000.

in accuracy, efficiency, or even privacy, this chapter studies efficient and privacy-preserving replication attack detection with guaranteed accuracy for large-scale RFID systems. To address time inefficiency and privacy leakage, the twin hobgoblins of the transmission of tag IDs, we leverage broadcast and collisions to detect the replication attack. We propose introducing lightweight operations to save unnecessary execution time and tag responses, and therefore harvest promising gains in both time efficiency and energy efficiency. Both theoretical analysis and simulation results demonstrate the accuracy and efficiency of the proposed protocols. For future work, how to identify all the replicated tags is worthy of consideration. Energy-efficient protocols in favor of the readers (e.g., the proposal in [Xu et al., 2010]) are also of interest.

Chapter 4

Efficient Detection of Replication Attacks in Large Anonymous RFID Systems

Replication attacks threaten Radio-Frequency Identification (RFID) applications but are hard to prevent. Existing replication attack detection methods are enslaved to the knowledge of tag identifiers (IDs). Tag IDs, however, should be protected to enable and secure privacy-sensitive applications in anonymous RFID systems. In a first step, this chapter tackles replication attack detection in anonymous RFID systems without requiring tag IDs as a priori. To this end, we leverage unreconciled collisions to uncover replication attacks. An unreconciled collision is probably due to responses from multiple tags with the same ID, exactly the evidence of replication attacks. This insight inspires GREAT, our pioneer protocol for replication attack detection in anonymous RFID systems. We evaluate the performance of GREAT through theoretical analysis and extensive simulations. The results show that GREAT can detect replication attacks in anonymous RFID systems fairly fast with required accuracy.

4.1 Overview

Replication attacks threaten Radio-Frequency Identification (RFID) applications but existing replication attack detection methods are enslaved to the knowledge of tag identifiers (IDs). In a replication attack, an attacker compromises genuine tags and produces their replicas (*replicated tags*) [Juels, 2006]. Holding replicated information of compromised tags, replicated tags behave exactly the same as genuine tags [Juels, 2006]. Replication attacks thus threaten many RFID applications that use the genuineness of tags to validate the quality or authenticity of tagged objects. For example, carrying replicated tags, products in an RFID-enabled supply chain lead to financial losses [Delen et al., 2007], healthcare facilities in RFID-aided hospitals jeopardize personal safety [Janz et al., 2005], while RFID-incorporated passport cards even threaten national security [Koscher et al., 2009]. Existing replication attack detection methods leverage data redundancy corresponding to tag IDs. Since normally a tag has a unique ID [web, a,b], if an ID associates simultaneously with different values of a certain attribute (e.g., tag location [Koh et al., 2003, Lehtonen et al., 2009a, Mirowski and Hartnett, 2007, Zanetti et al., 2010b] or synchronized secret [Lehtonen et al., 2009b]), the ID relates to multiple tags and reveals a replication attack.

In a first step, this chapter tackles replication attack detection in anonymous RFID systems without requiring tag IDs as a priori. More specifically, the anonymity requires that readers cannot query tag IDs from tags or backend servers. Anonymous RFID systems enable privacy-sensitive applications [Kodialam et al., 2007, Vahedi et al., 2011]. In such applications, communicating tag IDs either between backend servers and readers or between readers and tags risks leakage of tag IDs, which are private information or can be easily used to infer other private information [Kodialam

et al., 2007]. Private information of concern is, for example, trade secrets in RFID-enabled supply chains [Zanetti et al., 2010b], personal privacy in RFID-incorporated passports or driver licenses [Koscher et al., 2009], and military strength in RFID-enabled weapon tracking systems [Dean, 2006, Harris, 2008]. Requiring the awareness of tag IDs, existing replication attack detection methods are therefore not applicable in anonymous RFID systems.

Forget about seemingly better intuitions. Before we introduce our method to replication attack detection in anonymous RFID systems, let us first walk through some intuitive approaches and shake off the reverie in which they seem to be better.

Prevention? Only if we could. Of course, if we could prevent tags from being replicated, we would not bother to detect replication attacks at all. A disappointing fact is, however, that no prevention scheme claims to completely defeat replication attacks yet [Lehtonen et al., 2009b]. Most existing prevention protocols use cryptography and encryption to make tags hard to replicate [Abawajy, 2009, Devadas et al., 2008, Dimitriou, 2006]. Apart from possible failures [Lehtonen et al., 2009b], they require additional hardware resources and key management strategies [Spiekermann and Evdokimov, 2009], which are hardly affordable to low-cost tags that cannot support any operation beyond hashes [Sarma, 2006]. A more promising prevention scheme resorts to unclonable physical architecture of tags [Bolotnyy and Robins, 2007]. However, even if tags armed with cloning-resistant architectures arrive in the near future, it is still not practical either to replace off-the-shelf tags with cloning-resistant tags or to recall them for upgrade—already 1.3 billion tags were in the market in 2005, and even 33 billion were expected in 2010 [web, c]. All the preceding concerns raised by replication attack prevention necessitate replication attack detection.

Authentication? No. Since replicated tags are not genuine tags after all, some may resort to tag authentication. Authentication is a sharp weapon against counterfeit tags that carry valid IDs but forged keys [Juels, 2006, Lakafosis et al., 2011, Tan et al., 2008a]. Different from counterfeit tags, replicated tags hold not only valid IDs but also valid keys. Replicated tags, therefore, can pass authentication as can genuine tags.

Tag cardinality estimation? No. Since replication attacks make the number of tags (*tag cardinality*) exceed the number of IDs (*ID cardinality*), some may suggest first estimating tag cardinality and then leveraging the difference between those two cardinalities. If the difference exceeds a certain threshold, chances are that replicated tags exist. But adopting the suggestion faces two major hindrances, the privacy of ID cardinality and the accuracy of tag cardinality estimation. First, ID cardinality is probably as privacy-sensitive as tag IDs in anonymous RFID systems. Consider, for example, a military anonymous RFID system that tracks weapons such as firearms and shells [Dean, 2006, Harris, 2008]. In such a system, tag IDs may reveal categories and models of tagged weapons, and ID cardinality indicates exactly how many weapons therein. To avoid exposing military strength through tag IDs and ID cardinality, both of them should be protected in the considered system.

Second, even if ID cardinality is known, we still cannot simply rely on the difference between it and tag cardinality estimation. Considering inaccuracy of tag cardinality estimation protocols, we sometimes cannot determine that the difference is due to replication attacks or tag cardinality estimation error. Even worse, when replicated tags exist, tag cardinality estimation protocols may encounter large estimation errors [Kodialam and Nandagopal, 2006, Qian et al., 2008, Shah-Mansouri and Wong, 2011,

Zheng et al., 2011]. They estimate tag cardinality using the distribution of the number of tag responses in a frame of time slots. But this distribution is likely to be disturbed by responses from replicated tags and thus to induce a large estimation error.

Our approach and contributions. We propose leveraging *unreconciled collisions* for replication attack detection in anonymous RFID systems. An unreconciled collision cannot be reconciled through arbitrating channel access among tags whose responses cause the collision. The motivation for leveraging unreconciled collisions lies in how RFID tags compete for channel access. In an RFID system, tags decide when to respond according to the value of their IDs [web, a,b]. In other words, multiple tags with the same ID simultaneously respond to a query message and thus induce an unreconciled collision. Since multiple tags having the same ID is exactly the evidence of replication attacks, we can leverage unreconciled collisions to uncover replication attacks yet not require the knowledge of tag IDs.

Taking the first step toward replication attack detection in anonymous RFID systems, the chapter makes the following contributions:

- Leverage unreconciled collisions to uncover replication attacks without requiring tag IDs as a priori. This countermeasure against replication attacks can enable and secure privacy-sensitive applications in anonymous RFID systems.

- Propose GREAT, a pioneer protocol leveraging unreconciled collisions for replication attack detection in anonymous RFID systems.

- Analyze theoretically GREAT's detection accuracy and execution time. The analysis results can guide protocol configuration for satisfying required detection accuracy.

- Validate the performance of GREAT through extensive simulations. The results show that GREAT can detect replication attacks in anonymous RFID systems fairly fast with required accuracy. When, for example, six replicated IDs hide among up to 50,000 tag IDs, GREAT can detect the replication attack in only 75.5 seconds with probability at least 0.99.

The rest of this chapter is organized as follows. Section 4.2 defines the problem of replication attack detection in anonymous RFID systems. Section 4.3 provides an overview of our method. Section 4.4 presents protocol design and theoretical analysis. Section 4.5 reports simulation results. Finally, Section 4.6 concludes the chapter and indicates future work.

4.2 System and Problem

We consider an anonymous RFID system that consists of a reader and many tags. The reader can communicate with all tags. Normally a tag attached to an object has a unique ID. Tag IDs may directly reveal private information of tagged objects or indirectly link to such information stored on a backend server. To satisfy privacy-sensitive applications, the anonymous RFID system should strictly control granting the reader access to the server and transmitting tag IDs (encrypted or not) between the reader and tags. We are concerned with replication attacks in which an attacker replicates genuine tags and attaches replicated tags to objects with questionable authenticity [Juels, 2006]. Using only the genuineness of tags to validate the authenticity of tagged objects, we cannot distinguish objects attached with genuine tags from objects attached with replicated tags. The problem is therefore to detect whether replicated tags exist in an anonymous RFID system. An implicit constraint

we would like to emphasize here is that, to participate in system operations, all tags reside in the communication region of the reader. This applies to also replicated tags if any; otherwise, they may fail the replication attack.

We formulate the problem using a probabilistic model: If the number of replicated IDs exceeds a given tolerance number, detect the replication attack with a probability no less than a given detection accuracy. A *replicated ID* corresponds to a genuine tag and some replicated tag(s). Both tolerance number and detection accuracy are set according to application requirements. By the intrinsic property of probabilistic methods, a higher tolerance number and a lower detection accuracy yield faster detection with less certainty. With detection accuracy and tolerance number set to 1 and 0, respectively, the problem is specialized to deterministic detection of the replication attack.

We do not assume the knowledge of tag IDs or of their cardinality. As we discussed, both tag IDs and their cardinality may induce privacy leakage. To best support privacy-sensitive RFID applications, we do not allow our replication attack detection method to collect tag IDs or to gain access to them on the backend server. We assume that the reader and tags communicate using a power level high enough to drown background noise; error correction coding against channel errors [Tran et al., 2009] is beyond the scope of this chapter. (As we will show at the end of Section 4.4.3, channel errors may induce false positives to replication attack detection. A feasible countermeasure against false positives is also investigated therein.) We consider a general scenario where a reader can communicate with all tags in an anonymous RFID system using a single channel [Chen et al., 2011, Kodialam and Nandagopal, 2006, Lehtonen et al., 2009b]. Adaptation of our replication attack detection method

to scenarios with multiple readers, multiple channels, or multiple subsystems for accommodating all tags is left for future work.

4.3 Methodology Overview

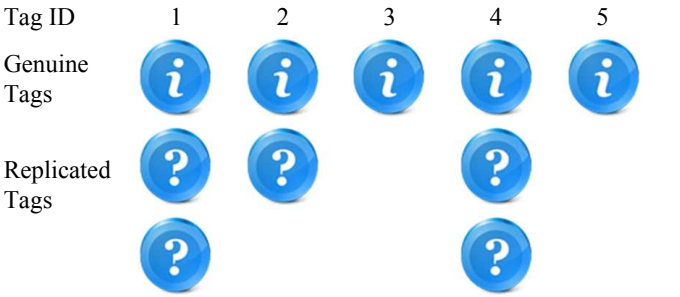
In this section, we provide an overview of replication attack detection using unreconciled collisions in anonymous RFID systems. We first discuss the motivation of unreconciled collisions by lessons from replication attack detection in identifiable RFID systems. We then discuss how to explore unreconciled collisions for uncovering replication attacks in anonymous RFID systems.

4.3.1 Lessons from Identifiable RFID Systems

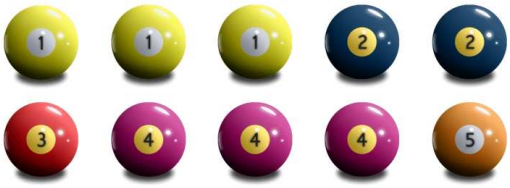
We start exploring the methodology by a warmup of replication attack detection in identifiable RFID systems. Figure 4.1(a) illustrates a replication attack instance with ten tags including five genuine tags (i.e., icons with symbol i) and five replicated tags (i.e., icons with question mark). For ease of presentation, we assign tag IDs 1 through 5. (But we do not assume that, given n genuine tags, tag IDs simply range from 1 to n.) Replicated IDs 1, 2, and 4 correspond to two, one, and two replicated tags, respectively. We will discuss two ideas of detecting the replication attack, through identification and through polling. To visualize the ideas, we transform the replication attack detection problem into the *ball drawing game* [Owen, 1999] as in Figures 4.1(b) and (c).

Detection through identification

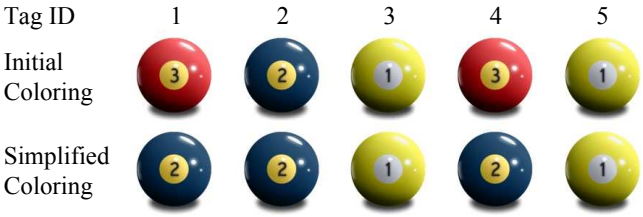
The intuition is that we can identify tags and detect the replication attack if a tag has the same ID as that of an identified tag. Figure 4.1(b) models this intuition



(a) Replication attack instance with 5 genuine tags and 5 replicated tags.



(b) Each tag is mapped to a ball. An ID is assigned with a unique color. Detecting replicated tags is transformed into the problem of drawing two balls in the same color.



(c) Each tag ID is mapped to a ball. A ball is colored according to the number of tags with the ID mapped to the ball. Detecting replicated tags is transformed into the problem of drawing a ball not in type 1 color.

Fig. 4.1: Replication attack detection in identifiable RFID systems modeled by the ball drawing game.

by the ball drawing game, in which we map each tag to a ball and assign a unique color to balls mapped from tags with the same ID. The goal is to draw two balls in the same color without replacement. Observing Figure 4.1(b), we can infer that it is more likely to achieve the goal when many balls are in the same color than to achieve it when otherwise. However, since it is not practical to identify tags in anonymous RFID systems [Kodialam et al., 2007, Vahedi et al., 2011], we in this game can hardly find any clues to detecting anonymous replicated tags.

Detection through polling

The intuition is that, if we know tag IDs in advance, we can verify whether one or more tags correspond to the same ID through polling. Figure 4.1(c) models a straightforward implementation of the intuition by the ball drawing game, in which we map each ID to a ball and assign type i color to the ball whose associated ID corresponds to i tags (we call this *initial coloring*). The goal is therefore to draw a ball not in type 1 color without replacement. Initial coloring, however, faces a dilemma: It requires in advance the number of replicated tags corresponding to each ID while those numbers are yet to obtain. Fortunately, we can escape from the dilemma by leveraging wireless broadcast. As each ID is mapped to a ball, drawing a ball is identical to a reader broadcasting a query message containing the ball's associated ID. Upon receiving the query message, a tag responds to the reader if its ID is identical to the contained one. The reader then verifies whether one or more tags respond if it receives an intact response or a collided one, respectively. The latter case reveals that multiple tags have the same ID and thus the reader detects the replication attack. We thus refine initial coloring to *simplified coloring* with only two types of colors in Figure 4.1(c)—Type 1 color for an ID corresponding to only one tag and type 2 color for an ID corresponding to multiple tags. The goal is still to draw a ball not in type 1 color without replacement; we can achieve it by leveraging wireless broadcast and response states (i.e., collision or non-collision).

So what can we learn from polling-based replication attack detection? Being optimistic, we could expect tag information (e.g., IDs and keys stored on a backend server) to be known also in an anonymous RFID system. Then we can simply apply polling-based detection. A likely modification is encrypting the broadcast IDs, which

are usually protected in anonymous systems. But being realistic, we have to prepare for no access to registered tag information. This concern is necessary because any granted access to them risks potential privacy leakage [Kodialam et al., 2007]. Such privacy leakage occurs when, for example, encrypted IDs are eavesdropped and decrypted [Juels, 2006], or the detection protocol is manipulated [Kothari et al., 2011]. The challenge is therefore to detect replication attacks among anonymous tags without knowing their IDs. Borrowing ideas from polling-based detection, if we could verify that whether a collision is caused by responses from tags with the same ID even if the ID is unknown, we can still detect replication attacks. We will shortly illustrate this idea and how we leverage it for replication attack detection in anonymous RFID systems.

4.3.2 Unreconciled Collisions in Anonymous RFID Systems

To implement the preceding idea, we expect tags to decide when to respond according to their IDs such that tags with the same ID always simultaneously respond. Tags with different IDs could, however, respond either simultaneously or asynchronously. If tags with different IDs respond simultaneously and cause a collision, we are likely to reconcile the collision by further arbitrating access to the channel among them. On the other hand, if a collision is due to responses from tags with the same ID, it is hard to reconcile. We refer to a collision that cannot be reconciled through arbitrating channel access among tags whose responses cause the collision as an *unreconciled collision*. Intuitively, an unreconciled collision is probably caused by a genuine tag and its replicated peer(s), that is, multiple tags with the same ID. Unreconciled collisions, therefore, enable us to uncover replication attacks in anonymous RFID systems.

Making tags decide when to respond according to their IDs, we do not have to

know the IDs in advance. Take, for example, a simple injection from a tag's ID to the index of the time slot in which the tag responds. Surely this straightforward injection is not desirable due to privacy leakage. Overhearing whether there is any response in each time slot, an attacker can easily infer tag IDs. Moreover, the injection method may take an unacceptable long time. Consider a general system configuration with 96-bit IDs, a 10-bit string with CRC embedded for verifying a collision, and 25 µs for transmitting a single bit [web, a,b]. Under such configuration, the injection method takes about

$$\frac{25 \times 10 \times 2^{96}}{10^6 \times 3600 \times 24 \times 365} > 0.6 \times 10^9 \times 10^9 \text{ (year)},$$

which is over 0.6 billion billion years, taking as if forever!

Collision arbitration protocols are well-investigated for arbitrating channel access among tags [Capetanakis, 1979, Roberts, 1975]. Such protocols are initially used to improve time efficiency of tag identification, which collects tag IDs without them being known in advance. Now we wonder that collision arbitration protocols may adapt to replication attack detection in anonymous RFID systems. To answer this conjecture, we will continue to review collision arbitration protocols and discuss which of them is of our interest.

4.3.3 Choice of Collision Arbitration Protocol

We briefly review two typical categories of collision arbitration protocols, *framed Aloha* [Roberts, 1975] and *tree traversal* [Capetanakis, 1979]. In framed Aloha, a reader creates a query frame with a number of time slots. The number of time slots within a query frame is usually called *frame size*. The reader then broadcasts the frame size and also a random seed. Using a hash function of the frame size, the

random seed, and its ID, a tag decides the index of the time slot in which it sends a response. A time slot chosen by no tag, only one tag, or multiple tags is known as an *empty slot*, a *singleton slot*, or a *collision slot* [Kodialam and Nandagopal, 2006]. Only in singleton slots can a reader correctly receive tag responses. In tree traversal, to collect l-bit tag IDs, a reader first creates a binary tree of height l and with each l-bit string mapped to a leaf. The reader then collects tag IDs through traversing the binary tree in a depth-first order. Specifically, the reader broadcasts the bit string corresponding to the current tree node; tags respond if their IDs are prefixed with the bit string. If no collision occurs, the reader can correctly receive the response. Otherwise, the reader continues to collect tag IDs by broadcasting the bit string of the current tree node's child. (More details about framed Aloha and tree traversal can be found in, for example, [Capetanakis, 1979, Hush and Wood, 1998, Lee et al., 2005, Roberts, 1975].)

Adapting collision arbitration protocols to replication attack detection in anonymous RFID systems, we choose framed Aloha over tree traversal. The reason for this choice is that tree traversal is susceptible to leaking a section of a tag ID or even an entire one. Overhearing the string s broadcast by a reader, an attacker can easily infer that at least one tag ID is prefixed with s after it overhears any response. Consider again the aforementioned RFID-enabled weapon tracking system [Dean, 2006, Harris, 2008]. A section of the tag ID, say s, may reveal weapon information (e.g., category and model) and thus expose military strength. When s is of length $l - 1$, the attacker can even infer that either $s0$ or $s1$ must be a tag ID if there is only one response, or both if a collision occurs. Such leakages are, of course, against the purpose of privacy-sensitive applications in anonymous RFID systems [Kodialam et al.,

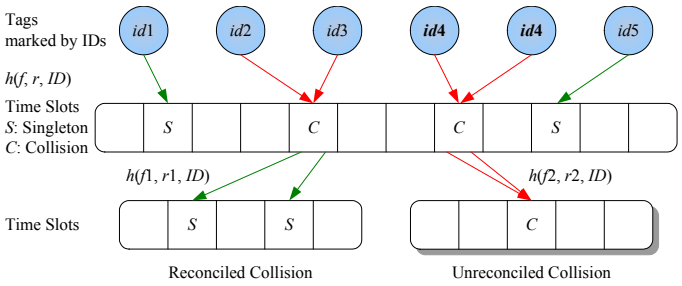


Fig. 4.2: An example of unreconciled collision caused by responses from two tags with the same ID $id4$ (i.e., a genuine tag and its replicated peer).

2007, Vahedi et al., 2011]. In framed Aloha, the attacker, however, can hardly infer a tag's ID using the hash result, that is, the index of the time slot in which the tag responds [Broder and Mitzenmacher, 2004].

4.3.4 Illustrative Example of Unreconciled Collisions

Having walked through the basics of unreconciled collisions and the choice of collision arbitration protocols for exploring unreconciled collisions, we now provide the big picture of how unreconciled collisions uncover replication attacks in anonymous RFID systems. Figure 4.2 illustrates a sample of six tags with IDs $id1$ through $id5$, among which $id4$ associates with two tags (i.e., a genuine tag and its replicated peer). For better illustration of unreconciled collisions, we deliberately make the ID of each tag explicit. In the first frame with frame size f and random seed r, a tag responds in a time slot with index decided by hash function $h(f, r, ID)$. Tags with $id1$ and $id5$ respond in two distinct singleton slots, while tags with $id2$ and $id3$ respond in the first collision slot and tags with $id4$ in the second collision slot. To reconcile the first collision, we let tags that responded in this slot (i.e., tags with $id2$ and $id3$) respond in the second frame with frame size $f1$ and random seed $r1$. We successfully reconcile the first collision because no collision occurs in the second frame. It is, however, not

hard to imagine that the second collision is unreconciled: Tags with the same ID $id4$ will still choose the same time slot to respond in the third frame with frame size $f2$ and random seed $r2$, causing a collision again.

But, of course, scenarios in anonymous RFID systems are a different story from the example in Figure 4.2—as we discussed, we are not aware of the IDs of anonymous tags in advance. Without knowing tag IDs, we can ensure only that a successfully reconciled collision is due to responses from genuine tags, whereas we cannot ensure that an unreconciled collision is due to responses from multiple tags with the same ID. So the challenge is to infer the probability of an unreconciled collision being caused by responses from multiple tags with the same ID, the very evidence of a replication attack. We next delve into leveraging unreconciled collisions to detect replication attacks with high probability in anonymous RFID systems.

4.4 GREAT: Greedy Collision-Slot–Reframing Detection Protocol

In this section, we propose the Greedy collision-slot–REfrAming deTection protocol (*GREAT*) against replication attacks in anonymous RFID systems. GREAT reframes collision slots to find unreconciled collisions and thus to detect replication attacks. We will also theoretically analyze GREAT's detection accuracy and execution time.

4.4.1 GREAT Design

GREAT detects a replication attack in an anonymous RFID system if an unreconciled collision occurs. To find an unreconciled collision, GREAT reconciles collisions

in a greedy manner: After reconciling a collision, if both some singleton slot(s) and some collision slot(s) show up, GREAT continues to reconcile the newly shown collision(s). GREAT reconciles collisions through *collision slot reframing*, an adaptation of framed Aloha. As will be detailed shortly, to reframe a collision slot, GREAT requires tags chose the slot to further respond in a new frame, as in Figure 4.2. In the new frame, if only one slot is collision and the others are empty, GREAT finds an unreconciled collision and therefore detects a replication attack.

During collision slot reframing, a challenge arises in a new frame when the first non-empty slot is collision: How can we decide whether or not to reframe the collision slot? We should reframe the collision slot if it is followed by some non-empty slot(s) (i.e., singleton or collision slot). We need not reframe the collision slot if it is followed by only some empty slot(s) or by no slot, because in both cases the collision slot exposes an unreconciled collision. To address the challenge, we quickly determine the number of non-empty slots in the new frame using 1-bit responses. If the new frame contains only one non-empty slot, it will contain only one collision slot under 10-bit responses, exactly the condition for an unreconciled collision. If the new frame contains multiple non-empty slots, we decide to reframe the collision slot under concern.

We now detail the GREAT design. The reader first broadcasts a query message containing the frame size f and a random seed r. Upon receiving the query message, a tag responds in the time slot with index $h(f, r, ID)$. The hash function $h(\cdot)$ implemented on tags enables a tag to choose in which time slot to respond uniformly at random [web, a,b, Kodialam and Nandagopal, 2006]. The response is a 10-bit string with CRC embedded for the reader to verify collisions [web, b]. After verifying an

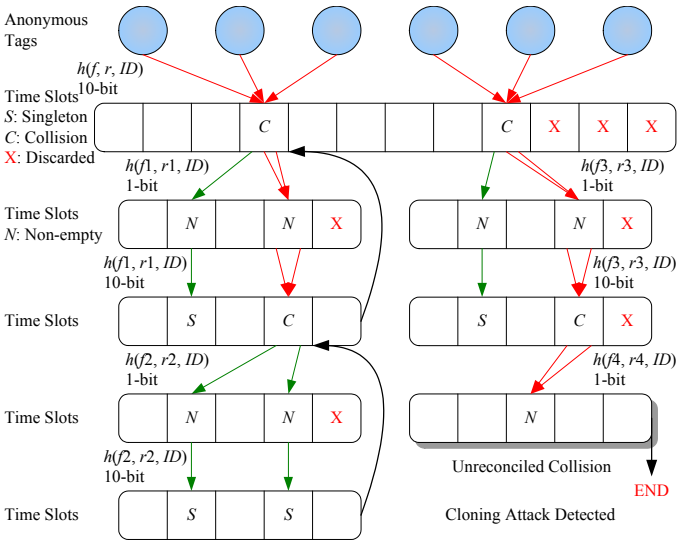


Fig. 4.3: GREAT execution instance for replication attack detection in an anonymous RFID system.

empty slot or a singleton slot, the reader issues a slot end command to trigger the next time slot. After verifying a collision slot, instead of issuing a slot end command, the reader reframes the collision slot with a new frame size f_{r} (*reframing size*) and a new random seed r_{r}, requiring 1-bit responses for quickly counting the number of non-empty slots. If the new frame contains only one non-empty slot, GREAT finds an unreconciled collision, detects a replication attack, and terminates. Otherwise, once the reader verifies the second non-empty slot, it reframes the collision slot again with f_{r} and r_{r}, requiring 10-bit responses. Then the reader greedily reframes a collision slot whenever it verifies one. After verifying all non-empty slots (and reframing collision slots if any) in an f_{r}-slotted frame, the reader traces back to the collision slot it just frames, and issues a slot end command to trigger the following time slot.

For ease of understanding the GREAT design, Figure 4.3 illustrates a GREAT execution instance. After verifying the first collision slot in the f-slotted frame, the

reader reframes it with $f_{\mathrm{r}} = f1$ and $r_{\mathrm{r}} = r1$, requiring 1-bit responses. After verifying the second non-empty slot in the $f1$-slotted frame, the reader again reframes the collision slot with $f1$ and $r1$, but requiring 10-bit responses. Using another frame with $f_{\mathrm{r}} = f2$ and $r_{\mathrm{r}} = r2$ to successfully reconcile the collision, the reader traces back to the first collision slot in the f-slotted frame, issues a slot end command, and continues to verify remaining slots. Similarly, the reader reframes the second collision slot in the f-slotted frame with $f_{\mathrm{r}} = f3$ and $r_{\mathrm{r}} = r3$ and reframes the first collision slot in the $f3$-slotted frame with $f_{\mathrm{r}} = f4$ and $r_{\mathrm{r}} = r4$. Since only one non-empty slot shows up in the $f4$-slotted frame, GREAT finds an unreconciled collision and detects the replication attack.

4.4.2 False Negative Rate

We analyze the maximum number $s_{\max}$ of slots in the f-slotted frame that GREAT needs to verify (and to reframe if any collision slot) to satisfy a false negative rate α. Note that in what follows, the analysis reckons hash values of tag IDs as following a uniform distribution, as considered in established literature (e.g., references [Chen et al., 2011, Kodialam and Nandagopal, 2006, Kodialam et al., 2007, Qian et al., 2008, Shahzad and Liu, 2012, Zheng et al., 2011], to name a few). More specifically, a tag ID has the same probability of being hashed into each time slot in a frame.

Lemma 4.1. *Given the frame size f, the tolerance number m of replicated IDs, when GREAT verifies up to s slots in the f-slotted frame, the false negative rate $P_{fn}(f, m, s)$ is upper bounded as the following:*

$$P_{fn}(f, m, s) \leq (1 - \frac{s}{f})^{m+1}. \tag{4.1}$$

Proof: Since GREAT detects replication attacks through greedy collision-slot reframing, GREAT can find an unreconciled collision and detect the replication attack

if at least one replicated tag responds in the s slots. A false negative thus occurs when all replicated tags respond in the last $f-s$ slots. Let m', where $m' > m$, denote the number of replicated IDs. The false negative rate $P_{\text{fn}}(f, m, s)$ can be defined as

$$P_{\text{fn}}(f, m, s) = (\frac{f-s}{f})^{m'} = (1 - \frac{s}{f})^{m'}. \tag{4.2}$$

Given certain f and s, $P_{\text{fn}}(f, m, s)$ in Equation 4.2 is a monotonically decreasing function of m'. Because $m' = m+1$ is the first integer that satisfies $m' > m$, we have

$$P_{\text{fn}}(f, m, s) = (1 - \frac{s}{f})^{m'} \leq (1 - \frac{s}{f})^{m+1},$$

using the monotonicity of $P_{\text{fn}}(f, m, s)$. ■

Theorem 4.1. *Given the frame size f, the tolerance number m of replicated IDs, the maximum number s_{max} of slots in the f-slotted frame GREAT verifies to satisfy a false negative α is as the following:*

$$s_{max} = \lceil (1 - \alpha^{\frac{1}{m+1}}) f \rceil.$$

Proof: By Equation 4.2, the false negative rate $P_{\text{fn}}(f, m, s)$ is a monotonically decreasing function of s. To minimize the execution time, GREAT should terminate right after it verifies the sth slot where $P_{\text{fn}}(f, m, s) \leq \alpha$, that is,

$$s_{\max} = \min\{s \mid P_{\text{fn}}(f, m, s) \leq \alpha\}. \tag{4.3}$$

By Lemma 4.1, $P_{\text{fn}}(f, m, s)$ is upper bounded. To satisfy $P_{\text{fn}}(f, m, s) \leq \alpha$, we must satisfy that the upper bound of $P_{\text{fn}}(f, m, s)$ is less than or equal to α. By plugging the upper bound in Formula 4.1 into Equation 4.3, we thus have

$$\begin{aligned} s_{\max} &= \min\{s \mid (1 - \frac{s}{f})^{m+1} \leq \alpha\} \\ &= \min\{s \mid s \geq (1 - \alpha^{\frac{1}{m+1}}) f\} \\ &= \lceil (1 - \alpha^{\frac{1}{m+1}}) f \rceil. \end{aligned}$$

■

4.4.3 False Positive Rate

We now analyze the minimum reframing size f_{rmin} that GREAT uses for collision slot reframing to satisfy a false positive rate β.

Lemma 4.2. *Given an unreconciled collision reframed by GREAT with an f_r-slotted frame, the false positive rate $P_{fp}(f_r)$ is upper bounded as the following:*

$$P_{fp}(f_r) \leq \frac{1}{f_r^2}. \tag{4.4}$$

Proof: Let n_{c} denote the number of the IDs of tags that cause the unreconciled collision. When $n_{\text{c}} = 1$, the unreconciled collision is due to responses from a genuine tag and its replicated peer(s), inducing no false positives. A false positive, however, occurs when $n_{\text{c}} \geq 2$ and all n_{c} IDs fall into the same slot in the f_{r}-slotted frame. The false positive rate $P_{\text{fp}}(f_{\text{r}})$ thus can be defined as

$$P_{\text{fp}}(f_{\text{r}}) = \sum_{i=0}^{f_{\text{r}}-1} \frac{1}{f_{\text{r}}} \frac{1}{f_{\text{r}}^{n_{\text{c}}}} = \frac{1}{f_{\text{r}}^{n_{\text{c}}}}. \tag{4.5}$$

Given a certain f_{r}, $P_{\text{fp}}(f_{\text{r}})$ in Equation 4.5 is a monotonically decreasing function of n_{c}. Because $n_{\text{c}} = 2$ is the first integer that satisfies $n_{\text{c}} \geq 2$, we have

$$P_{\text{fp}}(f_{\text{r}}) = \frac{1}{f_{\text{r}}^{n_{\text{c}}}} \leq \frac{1}{f_{\text{r}}^2},$$

using the monotonicity of $P_{\text{fp}}(f_{\text{r}})$. ■

Theorem 4.2. *Given an unreconciled collision, the minimum reframing size f_{rmin} for GREAT to satisfy a false positive rate β is as the following:*

$$f_{rmin} = \lceil \beta^{-\frac{1}{2}} \rceil.$$

Proof: By Equation 4.5, the false positive rate $P_{\text{fp}}(f_{\text{r}})$ is a monotonically decreasing function of f_{r}. To minimize the time for reframing a collision slot, GREAT should set the minimum f_{r} that satisfies $P_{\text{fp}}(f_{\text{r}}) \leq \beta$, that is,

$$f_{\text{rmin}} = \min\{f_{\text{r}} \mid P_{\text{fp}}(f_{\text{r}}) \leq \beta\}. \tag{4.6}$$

By Lemma 4.2, $P_{\text{fp}}(f_{\text{r}})$ is upper bounded. To satisfy $P_{\text{fp}}(f_{\text{r}}) \leq \beta$, we must satisfy that the upper bound of $P_{\text{fp}}(f_{\text{r}})$ is less than or equal to β. By plugging the upper bound in Formula 4.4 into Equation 4.6, we thus have

$$\begin{aligned} f_{\text{rmin}} &= \min\{f_{\text{r}} \mid \frac{1}{f_{\text{r}}^2} \leq \beta\} \\ &= \min\{f_{\text{r}} \mid f_{\text{r}} \geq \beta^{-\frac{1}{2}}\} \\ &= \lceil \beta^{-\frac{1}{2}} \rceil. \end{aligned}$$

■

Corollary 4.1. *Given an unreconciled collision reframed by GREAT with an f_r-slotted frame, $f'_{rmin} = 2$ is the minimum f_r to satisfy that the probability of a replication attack is greater than the probability of a false positive.*

Proof: Given the false positive rate $P_{\text{fp}}(f_{\text{r}})$, the probability that the unreconciled collision is due to a replication attack is $1 - P_{\text{fp}}(f_{\text{r}})$. By Lemma 4.2, $P_{\text{fp}}(f_{\text{r}})$ is upper bounded. $1 - P_{\text{fp}}(f_{\text{r}})$ is, therefore, lower bounded. To guarantee that $(1 - P_{\text{fp}}(f_{\text{r}})) > P_{\text{fp}}(f_{\text{r}})$, we derive f'_{rmin} as follows:

$$\begin{aligned} f'_{\text{rmin}} &= \min\{f_{\text{r}} \mid \min(1 - P_{\text{fp}}(f_{\text{r}})) > \max(P_{\text{fp}}(f_{\text{r}}))\} \\ &= \min\{f_{\text{r}} \mid 1 - \frac{1}{f_{\text{r}}^2} > \frac{1}{f_{\text{r}}^2}\} \\ &= \min\{f_{\text{r}} \mid f_{\text{r}} > \sqrt{2}\} \\ &= 2. \end{aligned}$$

■

Post-detection operations, such as replicated-tag identification, can eliminate false positives. Following replication attack detection, replicated-tag identification aims to identify all replicated IDs and thus to identify replicated tags with certainty. Toward certainty, replicated-tag identification must be granted an access to tag IDs and keys. Based on the accessed tag information, a straightforward replicated-tag identification is through polling, as we discussed in Section 4.3.1. To avoid privacy leakage and time inefficiency by transmitting encrypted IDs during polling, a better method is

to adapt GREAT. Given accessed IDs, GREAT can pre-hash the IDs and ensure in advance exactly which IDs are in which slots. During tags respond to the reader, if a collision occurs in a slot into which only one ID is pre-hashed, the ID must correspond to some replicated tag(s). Thus we can identify all replicated tags within a number of iterations, and in return, eliminate false positives if any. It is worth mentioning also that the above adaptation of GREAT is analogous to the information collection problem [Chen et al., 2011] that collects data from tags of which the IDs are known a priori. The data for GREAT to collect from a tag are just a random bitstring long enough for detecting a collision. For interested readers wondering whether verifying a number of tag IDs is time-consuming, the approximate execution time if leveraging the proposal in [Chen et al., 2011] is about 1.6 times the lower bound—$1.6nt_{\mathrm{c}}$, where n represents the ID cardinality and t_{c} denotes the time to detect a collision slot. Note that the above discussed replicated-tag identification can also combat false positives due to channel errors or noises. Such false positives occur when channel errors or noises turn an intact response in a singleton slot into a collided one.

4.4.4 Detection Accuracy

We now analyze the detection accuracy measured by the probability of detecting an existing replication attack.

Theorem 4.3. *Given an f-slotted frame and the tolerance number m of replicated IDs, when GREAT detects an existing replication attack by verifying the first s slots and reframing collision slots with the reframing size f_r, the detection accuracy $P_d(f, m, s, f_r)$ is lower bounded as the following:*

$$P_d(f, m, s, f_r) \geq 1 - (1 - \frac{s}{f})^{m+1} + P'_d(f, m, s, f_r), \tag{4.7}$$

where $0 \leq P'_d(f, m, s, f_r) \leq (1 - \frac{s}{f})^{m+1} \frac{1}{f_r^2}$.

Proof: GREAT can detect an existing replication attack in two cases. First, if at least one replicated ID corresponds to responses in the first s slots, GREAT can find an unreconciled collision and detect the replication attack. The first case, therefore, occurs when no false negative occurs. Second, if no replicated ID corresponds to responses in the first s slots, GREAT can also find an unreconciled collision due to a false positive and thus detect the replication attack. Combining detection probabilities in these two cases, we have

$$P_{\mathrm{d}}(f, m, s, f_{\mathrm{r}}) = (1 - P_{\mathrm{fn}}(f, m, s)) \cdot 1 + P_{\mathrm{fn}}(f, m, s) \cdot P_{\mathrm{fp}}(f_{\mathrm{r}}).$$

$P_{\mathrm{fn}}(f, m, s)$ and $P_{\mathrm{fp}}(f_{\mathrm{r}})$ as in Lemma 4.1 and Lemma 4.2, respectively, are both upper bounded. Using the upper bounds therein, we can derive that

$$1 - P_{\mathrm{fn}}(f, m, s) \geq 1 - (1 - \frac{s}{f})^{m+1},$$
$$P_{\mathrm{fn}}(f, m, s) \cdot P_{\mathrm{fp}}(f_{\mathrm{r}}) \leq (1 - \frac{s}{f})^{m+1} \frac{1}{f_{\mathrm{r}}^2}.$$

Let $P'_{\mathrm{d}}(f, m, s, f_{\mathrm{r}}) = P_{\mathrm{fn}}(f, m, s) \cdot P_{\mathrm{fp}}(f_{\mathrm{r}})$. Plugging the above two inequalities into the expression of $P_{\mathrm{d}}(f, m, s, f_{\mathrm{r}})$, we derive Formula 4.7 and prove Theorem 4.3. ■

To satisfy required false negative rate α and false positive rate β, from Theorem 4.3 follows easily Corollary 4.2.

Corollary 4.2. *Given an f-slotted frame and the tolerance number m of replicated IDs, when GREAT detects an existing replication attack by verifying the first s_{max} slots and reframing collision slots with the reframing size f_{rmin} to satisfy a false negative rate α and a false positive rate β, the detection accuracy $P_d(f, m, s_{max}, f_{rmin})$ is lower bounded as*

$$P_d(f, m, s_{max}, f_{rmin}) \geq 1 - \alpha + f(\alpha, \beta),$$

where $0 \leq f(\alpha, \beta) \leq \alpha\beta$.

4.4.5 Execution Time

As we will show, the expected execution time of GREAT is upper bounded by a function of ID cardinality n. Although GREAT does not require n to be known, system managers or whoever adopt GREAT and know the value of n can benefit from the expected execution time upper bound. A possible benefit is, for example, to facilitate scheduling multiple tag monitoring operations [Bu et al., 2011, 2012a, Chen et al., 2011, Xiao et al., 2012].

Theorem 4.4. *Given the ID cardinality n, the frame size f, the number of slots s in the f-slotted frame GREAT verifies, and the reframing size f_r, the expected execution time of GREAT $E[T(n, f, s, f_r)]$ is upper bounded as*

$$E[T(n, f, s, f_r)] \leq \frac{nsf_r}{f} t_e + (\frac{nsf_r}{f} + s)t_c,$$

where t_e denotes the time to detect an empty slot, and t_c denotes the time to detect a collision slot.

Proof: When GREAT verifies only the first s slots in the f-slotted frame, we expect $\frac{s}{f}n$ IDs in s slots. To detect replicated IDs among these $\frac{s}{f}n$ IDs, GREAT takes the maximum execution time when it verifies all $\frac{s}{f}n$ IDs, in two cases. The first case is when there is no replicated ID among the $\frac{s}{f}n$ ones. The second case is when there is only one replicated ID among the $\frac{s}{f}n$ IDs but the replicated ID is the $\frac{s}{f}n$th one for GREAT to verify.

The proof turns to finding the maximum time for GREAT to verify all $\frac{s}{f}n$ IDs. By the GREAT design (Section 4.4.1), GREAT normally ends empty and singleton slots, and further reframes collision slots. The maximum number of collision slots to reframe thus yields the maximum execution time. For $\frac{s}{f}n$ IDs to yield the maximum number of collision slots, there should be no singleton slot among the s ones. The proof turns to making IDs in each collision slot to yield the maximum number of collision slots

to reframe. Let s_c denote the number of collision slots in the s ones. Given n_j IDs in a collision slot, where $0 \leq j \leq s_c - 1$ and $\sum_{j=0}^{s_c-1} n_j = \frac{s}{f}n$, the maximum time for reconciling it occurs when there are only two non-empty slots under 1-bit responses and there are one singleton slot and one collision slot with $n_j - 1$ IDs under 10-bit responses. The same scenario applies to reframing the collision slot with $n_j - 1$ IDs, that is, there are two non-empty slots under 1-bit responses and there are one single slot and one collision slot with $(n_j - 1) - 1$ IDs under 10-bit responses. Following this recursion, we conclude that it maximally takes $n_j(t_e + t_c)$ to reconcile a collision caused by n_j IDs. Combining st_c taken by s slots, we have

$$\begin{aligned} E[T(n, f, s, f_r)] &\leq st_c + \sum_{j=0}^{s_c-1} n_j f_r(t_e + t_c) \\ &= st_c + \frac{s}{f} n f_r(t_e + t_c) \\ &= \frac{nsf_r}{f} t_e + (\frac{nsf_r}{f} + s)t_c. \end{aligned}$$

■

From Theorems 4.1, 4.2, and 4.4 follows easily the following Corollary 4.3.

Corollary 4.3. *Given the ID cardinality n, the tolerance number m of replicated IDs, and the frame size f, to satisfy a false negative rate α and a false positive rate β, the expected execution time of GREAT $E[T(n, f, m, \alpha, \beta)]$ is upper bounded as the following:*

$$E[T(n, f, m, \alpha, \beta)] \leq \frac{ns_{max}f_{rmin}}{f} t_e + (\frac{ns_{max}f_{rmin}}{f} + s_{max})t_c,$$

where $s_{max} = \lceil (1 - \alpha^{\frac{1}{m+1}})f \rceil$, $f_{rmin} = \lceil \beta^{-\frac{1}{2}} \rceil$, t_e denotes the time to detect an empty slot, and t_c denotes the time to detect a singleton or a collision slot.

4.4.6 Limitation: Generating Tag Profiles

A potential limitation of GREAT is that several runs of GREAT with a tag may generate the tags's profile. Specifically, a tag profile consists of a series of vectors comprising frame size f_i, random seed r_i, and slot index $h(f_i, r_i, ID)$ corresponding

to the ith run. Tag profiles may be exploited to track certain behaviors of tagged objects. The reader can thus leverage tag profiles to monitor tags of interest without using exact tag IDs. If, however, manipulated by malicious readers, tag profiles may indirectly reveal some sensitive information of tagged objects (e.g., locations inferred from where tag profiles are extracted). Against this challenging issue, we can resort to a pioneer proposal in [Gollakota et al., 2011]. For conciseness, we (1) emphasize here the primary finding that the proposal in [Gollakota et al., 2011] can detect and jam signals from unauthorized readers without interfering the communication between reliable readers and tags, and (2) refer interested readers to [Gollakota et al., 2011] for more advanced details.

4.5 Performance Evaluation

In this section, we evaluate the performance of GREAT through simulations. Since GREAT is, to the best of our knowledge, the first protocol for replication attack detection in anonymous RFID systems, we conduct simulations with no comparison other. Simulation results show that GREAT can detect replication attacks in an anonymous RFID system fairly fast with required accuracy. When, for example, six replicated IDs hide among 50,000 tag IDs, GREAT can detect the replication attack in only 75.5 seconds with probability at least 0.99.

4.5.1 Environment Configuration

We simulate the anonymous RFID system defined in Section 4.2: An RFID reader and many tags, including genuine tags and replicated tags, communicate via a single channel, using a power level high enough to drown background noise [Chen et al.,

2011, Kodialam and Nandagopal, 2006, Lehtonen et al., 2009b]. This scenario is not that complex but general enough for us to acquire insights for replication attack detection in anonymous RFID systems and to validate the proposed detection protocol. Scenarios of our future interest are, for example, with channel errors, multiple readers, multiple channels, or multiple subsystems for accommodating all tags [Mohsenian-Rad et al., 2010, Tran et al., 2009].

The primary performance metric is the execution time of GREAT for satisfying required detection accuracy. Slot timings are set according to the Philips I-CODE specification [web, b]: A reader requires $t_{\mathrm{e}} = 0.4$ ms to detect an empty slot or a non-empty slot, and $t_{\mathrm{c}} = 0.8$ ms to detect a singleton slot or a collision slot. Detection accuracy, by Corollary 4.2, is lower bounded by a function of false negative rate α and false positive rate β, $1 - \alpha + f(\alpha, \beta)$, where $0 \leq f(\alpha, \beta) \leq \alpha\beta$. Therefore, given required false negative rate α, GREAT is expected to detect the replication attack with accuracy no less than $1 - \alpha$. The maximum number $s_{\max}$ of slots in an f-slotted frame to verify for satisfying α and the minimum reframing size f_{rmin} to reframe collision slots for satisfying β are determined by Theorem 4.1 and Theorem 4.2, respectively. As shown in Figure 4.4(a), $s_{\max}$ is equal to f for $\alpha = 0$ and decreases with both α and tolerance number m of replicated IDs for $0 < \alpha \leq 1$. Figure 4.4(b) plots f_{rmin} with varying β; $f_{\mathrm{rmin}} = 32$ is enough for GREAT to satisfy $\beta = 0.001$.

4.5.2 Varying Frame Size f

We first investigate the impact of frame size f on the execution time of GREAT. Figure 4.5(a) shows the results under scenarios with false negative rate $\alpha = 0$, false positive rate $\beta = 0.001$, $m+1 = 1$ replicated IDs, and ID cardinality $n = 1{,}000$, 1,500, and 2,000. An interesting finding is that, given a certain n, the execution time of

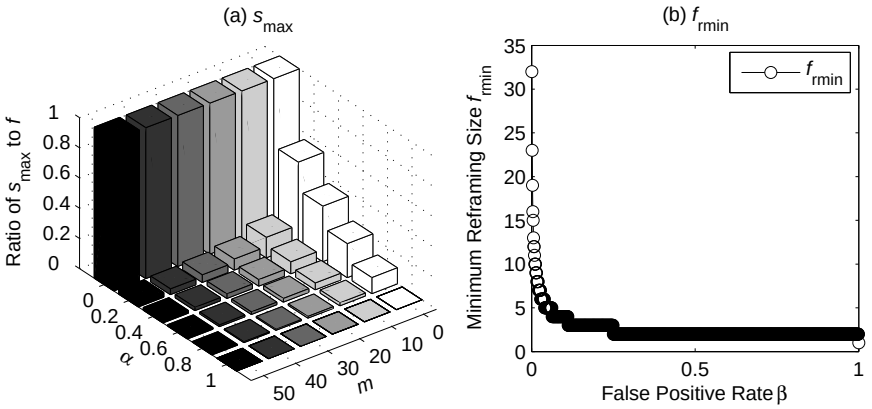


Fig. 4.4: Parameter setting. (a) The maximum number $s_{\max}$ of slots in an f-slotted frame to verify for satisfying tolerance number m of replicated IDs and false negative rate α. (b) The minimum reframing size f_{rmin} to reframe collision slots for satisfying false positive rate β.

GREAT is not a monotonically increasing function of f. As f increases, the execution time first decreases and then increases, approaching the minimum at $f \approx 6n$. Such variation of the execution time is essentially related to the variation of the numbers of empty, singleton, and collision slots. Intuitively, as f increases, the number of empty slots increases, the number of singleton slots increases up to $n - (m + 1) = n - 1$, while the number of collision slots decreases down to $m + 1 = 1$. By the GREAT design (Section 4.4.1), collision slot reframing makes a collision slot take more time than does an empty or a single slot. The execution time of GREAT thus decreases if the time reduction by collision slots exceeds the time increase by empty and singleton slots, and increases otherwise.

4.5.3 Varying Tolerance Number m of Replicated IDs

We now investigate the impact of tolerance number m of replicated IDs on the execution time of GREAT. Figure 4.5(b) shows the results under scenarios with false negative rate $\alpha = 0.001$, false positive rate $\beta = 0.001$, ID cardinality $n = 2{,}000$, and

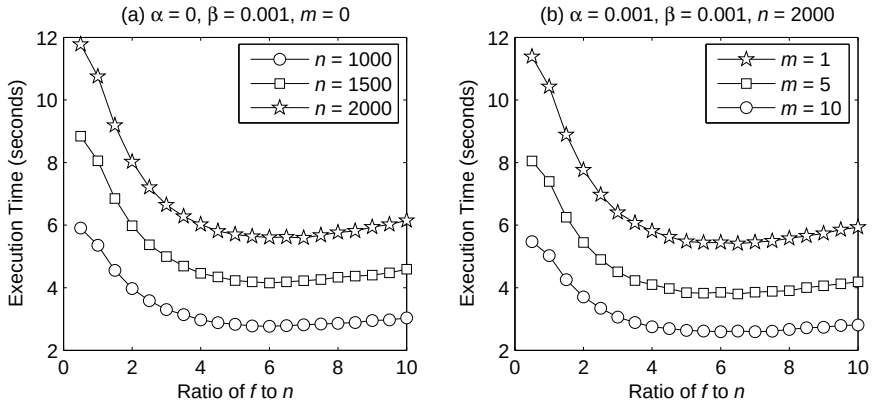


Fig. 4.5: Execution time of GREAT with varying frame size f under given ID cardinality n, tolerance number m of replicated IDs, false negative rate α, and false positive rate β.

$m + 1 = 2$, 6, and 11 replicated IDs. An obvious observation on the results is that, when $m > 0$, the execution time varies following the same trend as when $m = 0$ (Figure 4.5(a)) with varying f. As f increases, the execution time first decreases and then increases, approaching the minimum at $f \approx 6n$. Another observation on the results is that a higher m yields faster detection. The execution time of GREAT depends on how many slots among f ones to verify. By Theorem 4.1, given certain α and f, the maximum number of slots to verify is $s_{\max} = \lceil (1 - \alpha^{\frac{1}{m+1}}) f \rceil$ and decreases with m; so the execution time decreases with m.

4.5.4 Varying ID Cardinality n

We further evaluate the execution time of GREAT in larger anonymous RFID systems with ID cardinality $n = 5,000$ to $50,000$. For ease of presentation, we report the results under scenarios only when frame size $f = 6n$ in Tables 4.1 and 4.2. Table 4.1 reports the execution time with false negative rate $\alpha = 0.001$, false positive rate $\beta = 0.001$ and varying tolerance number m of replicated IDs. Given a certain m,

Table 4.1: Execution Time of GREAT with varying ID cardinality n, varying tolerance number m of replicated IDs, frame size $f = 6n$, false negative rate $\alpha = 0.001$, and false positive rate $\beta = 0.001$

n	Execution Time in Seconds				
	$m = 2$	$m = 4$	$m = 6$	$m = 8$	$m = 10$
5,000	12.6	10.5	8.8	7.5	6.5
10,000	25.3	21.0	17.6	15.1	13.1
15,000	38.0	31.6	26.5	22.6	19.7
20,000	50.7	42.2	35.3	30.2	26.2
25,000	63.3	52.7	44.1	37.7	32.8
30,000	76.0	63.3	53.0	45.3	39.4
35,000	88.7	73.8	61.8	52.8	45.9
40,000	101.4	84.3	70.6	60.3	52.5
45,000	114.0	94.9	79.5	67.9	59.1
50,000	126.8	105.4	88.3	75.4	65.6

the execution time increases with n; given a certain n, the execution time decreases with m. Table 4.2 reports the execution time with $\beta = 0.001$, $m = 5$, and varying α. Given a certain α, the execution time increases with n; given a certain n, the execution time decreases with α. In summary, (1) given certain α and m, the execution time increases with n; and (2) given a certain n, higher α and m yield faster detection.

4.6 Summary

We have studied replication attack detection in anonymous RFID systems. To enable and secure privacy-sensitive applications in anonymous RFID systems, we cannot simply turn to existing replication attack detection protocols that require the knowledge of tag IDs. We therefore tackle replication attack detection in anonymous RFID systems without requiring tag IDs as a priori and propose a pioneer protocol. The proposed protocol leverages unreconciled collisions to uncover replication attacks.

Table 4.2: Execution Time of GREAT with varying ID cardinality n, tolerance number $m = 5$ of replicated IDs, frame size $f = 6n$, varying false negative rate α, and false positive rate $\beta = 0.001$

n	Execution Time in Seconds			
	$\alpha = 0.002$	$\alpha = 0.003$	$\alpha = 0.005$	$\alpha = 0.010$
5,000	9.0	8.7	8.2	7.5
10,000	18.1	17.4	16.5	15.1
15,000	27.3	26.2	24.7	22.6
20,000	36.3	34.9	33.0	30.2
25,000	45.4	43.7	41.3	37.7
30,000	54.5	52.4	49.5	45.3
35,000	63.6	61.1	57.8	52.8
40,000	72.7	69.9	66.0	60.3
45,000	81.7	78.6	74.3	67.9
50,000	90.8	87.3	82.6	75.5

Simulation results show that the proposed protocol can detect replication attacks in anonymous RFID systems fairly fast with required accuracy. Future work lies in error correction coding against channel errors [Tran et al., 2009] and adaptation of the proposed protocol to multi-reader, multi-channel, or multi-subsystem scenarios [Mohsenian-Rad et al., 2010]. Of conceivable challenge is the adaptation to multi-reader scenarios. As when multiple readers are necessary for monitoring tags, it is possible that no reader covers some replicated tag(s) and corresponding genuine tag(s) in its communication region. If this is the case, readers may hardly find any reconciled collision and therefore the proposed protocol fails to detect replication attacks. Further efforts are thus dedicated primarily to replication attack detection in multi-reader anonymous RFID systems.

Chapter 5

Conclusion and Future Work

5.1 Conclusions

In the thesis, we have proposed a series of efficient protocols for two important monitoring operations in large-scale RFID systems, namely misplaced-tag pinpointing in Chapter 2 and replication attack detection in Chapter 3 and Chapter 4. We design the proposed protocols with making the most of bits in mind. The fewer bits an RFID protocol requires readers and tags to transmit, the more efficiency it promises to large RFID systems.

We first propose a series of protocols toward efficient misplaced-tag pinpointing in large RFID systems. The proposed protocols detect misplaced tags based on reader vectors instead of tag vectors. In favor of energy saving for active tags, the proposed protocols can detect misplaced tags by requiring only a subset of tags to respond. We also extend the proposed protocols to scenarios with distributed execution and tag mobility. Both analysis and simulation results show that the proposed protocols yield significantly increased time efficiency and energy efficiency compared with basic solutions based on tag-wise positioning.

We then study efficient and privacy-preserving replication attack detection with

guaranteed accuracy for large RFID systems. We propose detection protocols without resorting to complex cryptography techniques, inefficient tag-wise scanning, or privacy-unaware transmission of tag IDs. The proposed protocols leverage the broadcast nature and collisions, being affordable to off-the-shelf low-cost tags. We further introduce lightweight operations to save unnecessary execution time and tag responses, and therefore harvest promising gains in both time efficiency and energy efficiency.

Considering that tag IDs should be protected to enable and secure privacy-sensitive applications in anonymous RFID systems, we also tackle replication attack detection without requiring tag IDs as a priori. We leverage unreconciled collisions to uncover replication attacks. An unreconciled collision is probably due to responses from multiple tags with the same ID, the very evidence of replication attacks. We accordingly propose a pioneer protocol for replication attack detection in anonymous RFID systems. We evaluate the performance of the proposed protocol through theoretical analysis and extensive simulations. The results show that the proposed protocol can detect replication attacks in anonymous RFID systems fairly fast with guaranteed accuracy.

5.2 Future Work

Future work lies in the following four directions. First, the positioning accuracy of the scheme in [Wang et al., 2007] may not satisfy requirements of certain applications. Inspired by the proliferation of sensor network localization [Bu et al., 2012b, Liu et al., 2010, Xiao et al., 2010], we could borrow some ideas therein to improve tag positioning accuracy. Second, of conceivable challenge is adapting the proposed

replication attack detection protocol for anonymous RFID systems to multi-reader scenarios. As when multiple readers monitor tags, it is possible that no reader covers some replicated tag(s) and corresponding genuine tag(s) in its communication region. If this is the case, readers may hardly find any reconciled collision and therefore the proposed protocol fails to detect replication attacks. Further efforts are thus dedicated primarily to replication attack detection in multi-reader anonymous RFID systems. Third, although evaluating research on large-scale RFID systems depends primarily on simulation nowadays, we urge our future work to evaluate and refine the proposed protocols in real RFID systems. Fourth, energy-efficient protocols in favor of the readers (e.g., the proposal in [Xu et al., 2010]) are also of interest.

Bibliography

EPC class-1 generation-2 RFID protocol v.1.0.9, a. URL http://www.epcglobalinc.org/home.

Philips semiconductors I-CODE smart label RFID tags, b. URL http://www.semiconductors.philips.com/acrobat_download/other/identification/SL092030.pdf.

Explosive growth projected in next five years for rfid tags, c. URL http://www.instat.com/press.asp?ID=1545.

Olympics technology: RFID's the ticket for secure games, d. URL http://www.eetimes.com/electronics-news/4078120/Olympics-technology-RFID-s-the-ticket-for-secure-Games.

IDENTEC SOLUTIONS, e. URL http://www.identecsolutions.com/.

Joseph pulitzer endowment, f. URL http://www.onlineconcepts.com/pulitzer/endow.htm.

Mobile payments 2012 - my mobile, my wallet?, g. URL http://www.innopay.com/content/mobile-payments-2012.

World's largest cruise ship launches RFID-based passenger-tracking system, h. URL http://www.rfidjournal.com/article/view/7415/.

J. Abawajy. Enhancing RFID tag resistance against cloning attack. In *Proc. of IEEE NSS*, pages 18–23, 2009.

W. Bishop. Documenting the value of merchandising. Technical report, National Association for Retail Merchandising Service, 2003.

L. Bolotnyy and G. Robins. Physically unclonable function-based security and privacy in RFID systems. In *Proc. of IEEE PerCom*, pages 211–220, 2007.

A. Broder and M. Mitzenmacher. Network applications of bloom filters: A survey. *Internet Mathematics*, 1(4):485–509, 2004.

K. Bu, B. Xiao, Q. Xiao, and S. Chen. Efficient pinpointing of misplaced tags in large RFID systems. In *Proc. of IEEE SECON*, pages 287–295, 2011.

K. Bu, B. Xiao, Q. Xiao, and S. Chen. Efficient misplaced-tag pinpointing in large RFID systems. *IEEE Transactions on Parallel and Distributed Systems*, 23(11): 2094–2106, 2012a.

K. Bu, Q. Xiao, Z. Sun, and Q. Xiao. Toward collinearity-aware and conflict-friendly localization for wireless sensor networks. *Computer Communications*, 35(13):1549–1560, 2012b.

J. Capetanakis. Tree algorithms for packet broadcast channels. *IEEE Transactions on Information Theory*, 25(5):505–515, 1979.

S. Chen, M. Zhang, and B. Xiao. Efficient information collection protocols for sensor-augmented RFID networks. In *Proc. of IEEE INFOCOM*, pages 3101–3109, 2011.

I. Chlamtac, C. Petrioli, and J. Redi. Energy-conserving access protocols for identification networks. *IEEE/ACM Transactions on Networking*, 7(1):51–59, 1999.

G. Dean. RFID weapons and armoury management system. *Retrieved August*, 22, 2006.

D. Delen, B.C. Hardgrave, and R. Sharda. RFID for better supply-chain management through enhanced information visibility. *Production and Operations Management*, 16(5):613–624, 2007.

S. Devadas, E. Suh, S. Paral, R. Sowell, T. Ziola, and V. Khandelwal. Design and implementation of PUF-based unclonable RFID ICs for anti-counterfeiting and security applications. In *Proc. of IEEE RFID*, pages 58–64, 2008.

T. Dimitriou. A lightweight RFID protocol to protect against traceability and cloning attacks. In *Proc. of IEEE SecureComm*, pages 59–66, 2006.

L.W. Ferreira Chaves, E. Buchmann, and K. Böhm. Finding misplaced items in retail by clustering RFID data. In *Proc. of ACM EDBT*, pages 501–512, 2010.

K. Finkenzeller et al. *RFID Handbook: Fundamentals and applications in contactless smart cards, radio frequency identification and near-field communication.* Wiley, 2010.

S. Gollakota, H. Hassanieh, B. Ransford, D. Katabi, and K. Fu. They can hear your heartbeats: Non-invasive security for implantable medical devices. In *Proc. of ACM SIGCOMM*, pages 2–13, 2011.

H. Han, B. Sheng, C.C. Tan, Q. Li, W. Mao, and S. Lu. Counting RFID tags efficiently and anonymously. In *Proc. of IEEE INFOCOM*, pages 1–9, 2010.

R.R. Harris. Feasibility of radio frequency identification (RFID) and item unique identification (iuid) in the marine corps small arms weapons tracking system. Technical report, DTIC Document, 2008.

Bruce Hendrickson. Conditions for unique graph realizations. *SIAM Journal of Computing*, 21:65–84, 1992.

D. Huang and H. Kapoor. Towards lightweight secure communication protocols for passive RFIDs. In *Proc. of IEEE SECON*, pages 1–9, 2009.

D.R. Hush and C. Wood. Analysis of tree algorithms for RFID arbitration. In *Proc. of IEEE ISIT*, page 107, 1998.

B.D. Janz, M.G. Pitts, and R.F. Otondo. Information systems and health care-II: Back to the future with RFID: Lessons learned-some old, some new. *Communications of the Association for Information Systems*, 15(1):132–148, 2005.

D.G. Jeong and W.S. Jeon. Performance of adaptive sleep period control for wireless communications systems. *IEEE Transactions on Wireless Communications*, 5(11): 3012–3016, 2006.

A. Juels. RFID security and privacy: A research survey. *IEEE Journal on Selected Areas in Communications*, 24(2):381–394, 2006.

M. Kodialam and T. Nandagopal. Fast and reliable estimation schemes in RFID systems. In *Proc. of ACM MobiCom*, pages 322–333, 2006.

M. Kodialam, T. Nandagopal, and W.C. Lau. Anonymous tracking using RFID tags. In *Proc. of IEEE INFOCOM*, pages 1217–1225, 2007.

R. Koh, E.W. Schuster, I. Chackrabarti, and A. Bellman. Securing the pharmaceutical supply chain. *Auto-ID Center MIT, White Paper*, 2003.

K. Koscher, A. Juels, V. Brajkovic, and T. Kohno. EPC RFID tag security weaknesses and defenses: Passport cards, enhanced drivers licenses, and beyond. In *Proc. of ACM CCS*, pages 33–42, 2009.

N. Kothari, R. Mahajan, T. Millstein, R. Govindan, and M. Musuvathi. Finding protocol manipulation attacks. In *Proc. of ACM SIGCOMM*, pages 26–37, 2011.

V. Lakafosis, A. Traille, H. Lee, E. Gebara, M.M. Tentzeris, G. DeJean, and D. Kirovski. RFID-CoA: The RFID tags as certificates of authenticity. In *Proc. of IEEE RFID*, pages 207–214, 2011.

S.R. Lee, S.D. Joo, and C.W. Lee. An enhanced dynamic framed slotted ALOHA algorithm for RFID tag identification. In *Proc, of IEEE MobiQuitous*, pages 166–172, 2005.

M. Lehtonen, F. Michahelles, and E. Fleisch. How to detect cloned tags in a reliable way from incomplete RFID traces. In *Proc. of IEEE RFID*, pages 257–264, 2009a.

M. Lehtonen, D. Ostojic, A. Ilic, and F. Michahelles. Securing RFID systems by detecting tag cloning. *Pervasive Computing*, 5538:291–308, 2009b.

K.S. Leong, M.L. Ng, A.R. Grasso, and P.H. Cole. Synchronization of RFID readers for dense RFID reader environments. In *Proc. of International Symposium on Applications and the Internet Workshops (SAINTW)*, 2006.

T. Li, S. Chen, and Y. Ling. Identifying the missing tags in a large RFID system. In *Proc. of ACM MobiHoc*, pages 1–10, 2010a.

T. Li, S. Wu, S. Chen, and M. Yang. Energy efficient algorithms for the RFID estimation problem. In *Proc. of IEEE INFOCOM*, pages 1–9, 2010b.

Xuan Liu, Shigeng Zhang, Kai Bu, and Bin Xiao. Complete and fast unknown tag identification in large RFID systems. In *Proc. of IEEE MASS*, pages 47–55, 2012.

Y. Liu, Z. Yang, X. Wang, and L. Jian. Location, localization, and localizability. *Journal of Computer Science and Technology*, 25(2):274–297, 2010.

L. Lu, Y. Liu, and X.Y. Li. Refresh: Weak privacy model for RFID systems. In *Proc. of IEEE INFOCOM*, pages 1–9, 2010.

Wen Luo, Shigang Chen, Tao Li, and Yan Qiao. Probabilistic missing-tag detection and energy-time tradeoff in large-scale RFID systems. Proc. of ACM MobiHoc, pages 95–104, 2012.

A. Matic, A. Papliatseyeu, V. Osmani, and O. Mayora-Ibarra. Tuning to your position: FM radio based indoor localization with spontaneous recalibration. In *Proc. of IEEE PerCom*, pages 153–161, 2010.

J. Mirkovic, G. Prier, and P. Reiher. Attacking DDoS at the source. In *Proc. of IEEE ICNP*, pages 312–321, 2002.

L. Mirowski and J. Hartnett. Deckard: A system to detect change of RFID tag ownership. *International Journal of Computer Science and Network Security*, 7(7): 89–98, 2007.

A.-H. Mohsenian-Rad, V. Shah-Mansouri, V.W.S. Wong, and R. Schober. Distributed channel selection and randomized interrogation algorithms for large-scale and dense RFID systems. *IEEE Transactions on Wireless Communications*, 9(4):1402–1413, 2010.

J. Myung, W. Lee, J. Srivastava, and T.K. Shih. Tag-splitting: adaptive collision arbitration protocols for rfid tag identification. *IEEE Transactions on Parallel and Distributed Systems*, 18(6):763–775, 2007.

Vinod Namboodiri and Lixin Gao. Energy-aware tag anti-collision protocols for RFID systems. In *Proc. of IEEE PerCom*, pages 23–36, 2007.

G. Owen. *Discrete mathematics and game theory*. Springer, 1999.

C. Qian, H. Ngan, and Y. Liu. Cardinality estimation for large-scale RFID systems. In *Proc. of IEEE PerCom*, pages 30–39, 2008.

C. Qian, Y. Liu, H. Ngan, and L.M. Ni. ASAP: scalable identification and counting for contactless RFID systems. In *Proc. of IEEE ICDCS*, pages 52–61, 2010.

C. Qian, H. Ngan, Y. Liu, and L.M. Ni. Cardinality estimation for large-scale RFID systems. *IEEE Transactions on Parallel and Distributed Systems*, 22(9):1441–1454, 2011.

Y. Qiao, S. Chen, T. Li, and S. Chen. Energy-efficient polling protocols in RFID systems. In *Proc. of ACM MobiHoc*, pages 1–9, 2011.

A. Raman, N. DeHoratius, and T. Zeynep. Execution: The missing link in retail operations. *California Management Review*, 43(3):136–152, 2001.

Y. Rekik, E. Sahin, and Y. Dallery. Analysis of the impact of the RFID technology on reducing product misplacement errors at retail stores. *International Journal of Production Economics*, 112(1):264–278, 2008.

L.G. Roberts. ALOHA packet system with and without slots and capture. *ACM SIGCOMM Computer Communication Review*, 5(2):28–42, 1975.

S. Sarma. Introductory talk: Some issues related to RFID and security. In *Proc. of Workshop on RFID Security (RFIDSec)*, 2006.

V. Shah-Mansouri and V.W.S. Wong. Cardinality estimation in RFID systems with multiple readers. *IEEE Transactions on Wireless Communications*, 10(5):1458–1469, 2011.

M. Shahzad and A.X. Liu. Every bit counts: fast and scalable RFID estimation. In *Proc.of ACM MobiCom*, pages 365–376, 2012.

B. Sheng, C.C. Tan, Q. Li, and W. Mao. Finding popular categories for RFID tags. In *Proc. of ACM MobiHoc*, pages 159–168, 2008.

B. Sheng, Q. Li, and W. Mao. Efficient continuous scanning in RFID systems. In *Proc. of IEEE INFOCOM*, pages 1–9, 2010.

S. Spiekermann and S. Evdokimov. Privacy enhancing technologies for RFID - A critical investigation of state of the art research. In *Proc. of IEEE Privacy and Security*, 2009.

C.C. Tan, B. Sheng, and Q. Li. Secure and serverless RFID authentication and search protocols. *IEEE Transactions on Wireless Communications*, 7(4):1400–1407, 2008a.

C.C. Tan, B. Sheng, and Q. Li. How to monitor for missing RFID tags. In *Proc. of IEEE ICDCS*, pages 295–302, 2008b.

T. Tran, T. Nguyen, B. Bose, and V. Gopal. A hybrid network coding technique for single-hop wireless networks. *IEEE Journal on Selected Areas in Communications*, 27(5):685–698, 2009.

E. Vahedi, V. Shah-Mansouri, V.W.S. Wong, I.F. Blake, and R.K. Ward. Probabilistic analysis of blocking attack in RFID systems. *IEEE Transactions on Information Forensics and Security*, 6(3):803–817, 2011.

C. Wang, H. Wu, and N.F. Tzeng. RFID-based 3-D positioning schemes. In *Proc. of IEEE INFOCOM*, pages 1235–1243, 2007.

S. Weis, S. Sarma, R. Rivest, and D. Engels. Security and privacy aspects of low-cost radio frequency identification systems. *Security in Pervasive Computing, Lecture Notes in Compute Science*, 2802:201–212, 2004.

Q. Xiao, B. Xiao, J. Cao, and J. Wang. Multihop range-free localization in anisotropic wireless sensor networks: A pattern-driven scheme. *IEEE Transactions on Mobile Computing*, 9:1592–1607, 2010.

Q. Xiao, K. Bu, and B. Xiao. Efficient monitoring of dynamic tag populations in RFID systems. In *Proc. of IEEE/IFIP EUC*, pages 106–113, 2011.

Q. Xiao, K. Bu, B. Xiao, and L. Sun. Efficient Protocol Design for Dynamic Tag Population Monitoring in Large-Scale RFID Systems. *Concurrency and Computation: Practice and Experience*, 2012.

Q. Xiao, B. Xiao, and S. Chen. Differential estimation in dynamic RFID systems. In *Proc. of IEEE INFOCOM*, 2013.

X. Xu, L. Gu, J. Wang, and G. Xing. Negotiate power and performance in the reality of RFID systems. In *Proc. of IEEE PerCom*, pages 88–97, 2010.

H. Yue, C. Zhang, M. Pan, Y. Fang, and S. Chen. A time-efficient information collection protocol for large-scale RFID systems. In *Proc. of IEEE INFOCOM*, pages 2158–2166, 2012.

D. Zanetti, B. Danev, et al. Physical-layer identification of UHF RFID tags. In *Proc. of ACM MobiCom*, pages 353–364, 2010a.

D. Zanetti, L. Fellmann, and S. Capkun. Privacy-preserving clone detection for RFID-enabled supply chains. In *Proc. of IEEE RFID*, pages 37–44, 2010b.

D. Zhang, J. Ma, Q. Chen, and L.M. Ni. An RF-based system for tracking transceiver-free objects. In *Proc. of IEEE PerCom*, pages 135–144, 2007.

D. Zhang, J. Zhou, M. Guo, J. Cao, and T. Li. Tasa: Tag-free activity sensing using rfid tag arrays. *IEEE Transactions on Parallel and Distributed Systems*, 22(4): 558–570, 2010.

D. Zhang, J. Zhou, M. Guo, J. Cao, and T. Li. Tasa: Tag-free activity sensing using RFID tag arrays. *IEEE Transactions on Parallel and Distributed Systems*, 22(4): 558–570, 2011a.

R. Zhang, Y. Liu, Y. Zhang, and J. Sun. Fast identification of the missing tags in a large RFID system. In *Proc. of IEEE SECON*, pages 278–286, 2011b.

Y. Zheng and M. Li. Fast tag searching protocol for large-scale RFID systems. In *Proc. of IEEE ICNP*, pages 363–372, 2011.

Y. Zheng, M. Li, and C. Qian. PET: Probabilistic estimating tree for large-scale RFID estimation. In *Proc. of IEEE ICDCS*, pages 37–46, 2011.

Z. Zhou, H. Gupta, S.R. Das, and X. Zhu. Slotted scheduled tag access in multi-reader RFID systems. In *Proc. of IEEE ICNP*, pages 61–70, 2007.