ON-DEMAND RELAY COMMUNICATION INFRASTRUCTURE FOR BASE STATION CONNECTIVITY

BY

YING HUANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Professor Klara Nahrstedt, Chair
Professor Tarek F. Abdelzaher
Professor Nitin Vaidya
Professor Songwu Lu, University of California, Los Angeles

Abstract

In public safety and homeland security, reliable communication among the command center (CC), first responders (FRs) and surveillance sensors is critical to real-time monitoring and control applications. Commanders want to check measurements from all surveillance sensors in the field and respond to critical incidents in real-time. In addition, commanders desire to monitor every FR's location, health and device status to safeguard their lives. Frequently, they need to exchange dispatch commands and incident status with FRs in the field via voice. Occasionally, commanders inform FRs by transferring various data, such as building maps and fire hydrant locations, or pull back video feeds and text reports. We refer to the multi-functional mobile devices being monitored and controlled by CC as terminals. In order to support all these life-critical and mission-critical applications, a communication infrastructure offering reliable communication paths between CC and terminals is needed.

However, such an infrastructure meeting various performance requirements from emergency response operations often does not exist. It is either impaired or nonfunctional due to incompatibility of radio frequency and communication protocols. Hence, we are interested in establishing a communication infrastructure on demand and cost-effectively, so that FRs can be continuously monitored, informed, managed and protected, while mobilized around a large incident area. In order to provision such a communication infrastructure, two means are taken simultaneously: (a) installing multiple base stations (BSs) to increase the coverage of the command center; (b) dropping relays to further extend the connectivity to BSs, especially when terminals are far away from BSs. In this thesis, we study various algorithms to determine the optimal locations of relays and the installation sequence so that the total number of relays is minimized, hence the infrastructure cost. We call such problems relay management problems.

There are typically two categories for relay management problems: relay placement problems for static networks and relay deployment problems for mobile networks. Many papers have studied the relay placement and deployment problems. However, FR systems do exhibit special properties of disconnected evolving networks, which are either overlooked or treated primitively by the prior work. To the best of our knowledge, (a) none of the relay placement algorithms consider polymorphous networks with multiple topologies, due to

terminal movement, unsynchronized wakeup schedule and packet forwarding policy; (b) all the prior work on relay deployment problems drop more relays than what is necessary, because they follow the "breadcrumb" approach, by which a FR will drop a relay whenever the connectivity to BSs is about to break.

Despite having evolving network topologies, FR systems are not completely unpredictable. A large amount of operation planning and scheduling knowledge can be exploited in relay management. For polymorphous networks, we have rudimentary knowledge about potential network configurations, e.g., where terminals will be placed or moved occasionally. For mobile networks, we know the coarse-level mobility patterns of FRs, e.g., the set of locations FRs will visit, from FRs dispatch and task assignment. In this thesis, we broadly exploit the predictability to deal with evolving topologies for FR systems, while meeting unique performance requirements, such as cost-effectiveness, reliability, load balance, etc. Weigh-and-place algorithm (WPA) is proposed to optimize relay placement across topologies with balanced load, reliably if required. Energy-aware relay placement is studied so that the communication infrastructure lasts for a desired network lifetime under different transmission control schemes. M-Breadcrumb, a mission-aware constrained relay deployment algorithm, is invented to minimize the total number of relays and navigate FRs to relay deploying locations with reduced traversal distance.

To my beloved parents and my dear husband Tao.

Acknowledgments

This work would not have been completed without support and encouragement from my professors, family, friends and colleagues.

First, I want to thank Professor Klara Nahrstedt, who is an excellent advisor, outstanding researcher and sincere friend. She always welcomes me to discuss problems and issues, provides comprehensive feedback and guides me towards the right direction. She is confident in my ability, encouraging me to pursue all the areas and projects I am interested in. She motivates me to be the teaching assistant for her distributed system class. Being quite busy for a whole semester, I had the most helpful and unforgettable teaching experience and learned the firsthand secret of being a great teacher. Klara is also a role model for me. Her passion, diligence and dedication to her career inspires me. It is a great honor for me to work with her.

Second, I want to thank Professor Tarek Abdelzaher, Professor Nitin Vaidya and Professor Songwu Lu, who provided valuable feedback and constructive suggestions to my thesis. My special thanks goes to Professor Tarek Abdelzaher. He is very sharp and helped identify the potential applications and interesting problems to explore. My sincere gratitude goes to Professor Nitin Vaidya , who thoroughly examined my proof and statements and helped me perfect the thesis work. I am grateful to have Professor Songwu Lu on board. Though in a different time zone, Professor Songwu Lu is always open to discussion and spotted several errors in my presentation.

Third, I want to thank my family. My dear husband, Tao, always urges me to work hard and helps me make most of the important decisions. I would not have completed my Ph.D. journey successfully and started my new career at Google smoothly without his support and encouragement. My parents, far away in China, give me enormous support and eternal care I need, both financially and emotionally under no conditions. I dedicate this dissertation to my family.

My special thanks goes to Yan Gao, who assisted me on the problem formulation and clarification in this dissertation. I also want to thank all the friends, Yong Yang, Zeng Zheng, Debessay Fesehaye, Wenbo He, Jing Liang, etc., I have made in University of Illinois at Urbana-Champaign. Without you, my life in Urbana-Champaign could not have been so wonderful and joyful. I also want to thank all the members of the

MONET (Multimedia Operating System and Networking) group. Every research discussion with you was so intense and inspiring. You are never hesitant to give me great tips and generous help when I encounter difficulty.

Finally, I want to acknowledge Lynette Lubben and Mary Beth, who ensured that all administrative procedures for the thesis deposit went through smoothly, while I have already left the campus.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

FR	First Responder
EMS	Emergency Medical Services
BS	Base Station
CC	Command Center
RTMC	Real-Time Monitoring and Control
FRCD	First Responder Communication Device
PAN	Personal Area Network
IAN	Incident Area Network
JAN	Jurisdiction Area Network
EAN	Extended Area Network
DEN	Disconnected Evolving Network
DPN	Disconnected Polymorphous Network
DMN	Disconnected Monomorphous Network
DDN	Disconnected Dynamic Network
WPA	Weight-and-Place Algorithm
TSA	Topology Stitch Algorithm
TIA	Topology Iterative Algorithm
TC	Terminal Cluster
TTCI-TC	Two-Tier Communication Infrastructure with Terminal Clusters
ERP	Energy-aware Relay Placement Algorithm
CRP	Connectivity-only Relay Placement Algorithm
TTCI	Two-Tier Communication Infrastructure
LDP	List of Relay Deploying Points
BC	Breadcrumb Approach
MBC-Flow	M-Breadcrumb with Flow Approach
MBC-Graph	M-Breadcrumb with Graph Approach

List of Symbols

$\mathcal{TS}$	$\mathcal{TS} = \{\mathcal{T}^1 \cdots \mathcal{T}^T\}$. A set of terminal topologies.
$\mathcal{B}$	$\mathcal{B} = \{B_1 \cdots B_S\}$. A set of base stations.
$\mathcal{L}$	A set of candidate locations for relays.
C	Monitoring or command center. (C is used in formula and CC is used in description.)
$\mathcal{R}$	Relay placement or deployment solution, $\mathcal{R} \subseteq \mathcal{L}$.
d_i^t	Supply/demand of node i in $\mathcal{T}^t$.
y_l	Binary decision variable to indicate whether a relay is placed in location l. 1 means placement.
x_{ij}^t	Flow of type-t commodities on directed edge $\widehat{ij}$ of topology $\mathcal{T}^t$.
ab	Undirected edge ab.
$\widehat{ab}$	Directed edge.
$\|ij\|$	Distance between node i and j.
$\|S\|$	Size of set S.
$I\{A\}$	Indicator function. The value is 1 if event A is true; 0, otherwise.
f_t	Fraction of time when terminal topology $\mathcal{T}^t$ is present.
L	Desired network lifetime.
γ_u^t	Data traffic rate from terminal u to CC in topology $\mathcal{T}^t$.
E_u	Initial energy provision for a terminal u.
E	Initial energy provision for a relay.
s_{data}	Data packet size.
s_{ack}	Acknowledgement (ACK) packet size.
r_{data}	Fixed data rate under power control.
r_{basic}	Basic rate for ACK packets.
r_{ij}^t	Data rate from node i to j in topology $\mathcal{T}^t$ under data rate control.
$E_s(d)$	Energy consumption rate to transmit packets from a node to another node d meters apart with power control.
E_s	Energy consumption rate to transmit packets with data rate control.

E_r	Energy consumption rate to receive packets, with either power or data rate control.
$E^t_{out}(i)$	Mean energy consumption rate for outbound traffic at node i in Topology $\mathcal{T}^t$.
$E^t_{in}(i)$	Mean energy consumption rate for inbound traffic (intended for node i) at node i in topology $\mathcal{T}^t$.
$E^t_{eav}(i)$	Mean energy consumption rate in overhearing packets intended for other nodes at node i in topology $\mathcal{T}^t$.
$SCTC^t_i$	Carrier sense transmission set for node i in topology $\mathcal{T}^t$ with power control.
SCS^t_i	Carrier sense set for node i in topology $\mathcal{T}^t$ with data rate control.
$p^t_{ij}(r,s)$	One-way data packet delivery ratio from node i to node j in topology $\mathcal{T}^t$ with packet size s sent at rate r.
p^t_{ij}	One-way acknowledgement delivery ratio from node i to node j in topology $\mathcal{T}^t$. Acknowledgement packets of size s_{ack} are sent at basic rate r_{basic}.
$P^r_{ij}(P_t)$	Received packet signal from node i to j, when i uses transmit power P_t.
QU^t_j	Maximum number of bytes node j can forward on behalf of node i in topology $\mathcal{T}^t$.
$CSThres$	Carrier sense threshold.
$SINR$	Signal-to-noise-and-interference ratio.
$\mathbf{G}$	$\mathbf{G} = \{G^1, \cdots, G^N\}$. N mobile groups, where G^i is the i^{th} group.
$\mathcal{MT}^i$	$\mathcal{MT}^i =< B^i, DL^i >$. Mission topology for group G^i. B^i is the starting base station and DL^i is the list of destinations where G^i will visit. $DL^i = \{dest^i_1, \cdots, dest^i_{p^i}\}$.
$DL^{i'}$	$\{B^i\} \bigcup DL^i$.
$\mathcal{R}_c$	Complete placement solution.
$\mathcal{R}_{\mathcal{A}}$	Single cover.
$\mathcal{LB}$	$\mathcal{LB} = \mathcal{L} \bigcup \mathcal{B}$.
$G^{\mathcal{LB}}$	Undirected relay communication graph.
$\mathcal{F}_{\mathcal{B}}$	Steiner forest.

Chapter 1

Introduction

Society has long been grappling with large-scale natural disasters, such as earthquakes, tsunamis, hurricanes, and outbreaks of contagious diseases. Even with accurate disaster prediction technology to order massive evacuation in advance, history has given us insight into the considerable cost to human lives and property when a large-scale natural disaster hits without measurable sign of imminence. For example, in hurricane Katrina, which hit the Gulf Coast of the United States in 2005, at least 1,330 were killed, thousands were injured, and estimated 300, 000 homes were made uninhabitable. Today, society faces other kinds of large-scale disasters caused in part ironically due to advances in technology itself, and in part increasingly due to social-economical problems and terrorist organizations, such as terrorism on September 11 of 2001.

Large-scale disasters, both natural and man-made, in recent years have awakened the government regarding the increasing need for improving communication infrastructures to support homeland security and public safety for first responders who keep us safe and away from harm. The US Department of Homeland Security exerts among its top priorities on stronger information sharing and infrastructure protection with inter-operable communication and equipments [1]. In retrospect, if the cell towers in New Orleans had not been blown over, communication among New Orleans, nearby large cities and federal departments would not have been dissolved and more timely resource and human aid could have been obtained. If the agencies had been able to talk to each other easily on September 11, 121 firefighters could have been timely informed of dangerous situations and have survived the tower collapse. In this thesis, we present our effort to establish a reliable communication infrastructure on demand using droppable relays, wherein real-time monitoring and control applications can be built upon. Through the infrastructure umbrella with diverse mission-critical applications, first responders in the incident scene can be continuously monitored, informed, protected, and managed, while operating and moving around a large incident area.

First responders (FRs), who play critical roles in protection of lives and property, represent personnel from various public safety agencies providing support for law enforcement, emergency medical services (EMS), fire control, hazardous material management, public health service, and other mutual aid. An *incident first responder system* comprises an incident command system and a wireless communication system

[2]. The *incident command system* manages planning, operations, logistics, administration, and information dissemination, to support FRs dispatched in incident scenes under a broad spectrum of emergencies. The incident command system, usually installed in vehicles near the incident scene, also serves as the gateway to resources in Internet and private cooperate networks, for example remote database. The *wireless communication system* is a multi-hop mobile wireless network established on demand, with or without existing WLAN infrastructure support, formed by terminals, droppable relaying devices (relays), the command center (CC) and base stations mounted on emergency vehicles parked around the incident scene. *Terminals* are multi-functional devices in the field. Besides relaying packets, terminals need to extract and process information, and communicate with other terminals and CC. Examples are handheld devices carried by FRs, mobile robots, surveillance cameras, static environmental sensors and sensors attached to patients. Via the self-configuring wireless communication network, FRs exchange information rapidly and securely among themselves, CC, surveillance system and remote databases, hence enabling cooperative rescue mission and intelligent decisions.

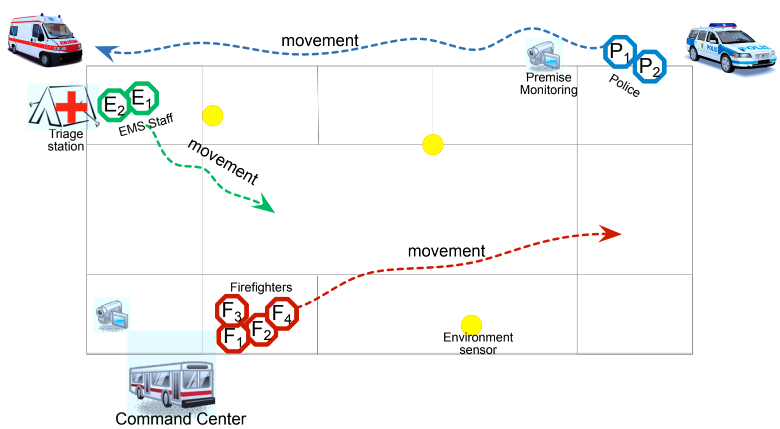


Figure 1.1: First Responder Mobile Networks

In Figure 1.1, we show a FR system, where policemen (P), firefighters (F), and EMS (E) staff move inside the incident scene in an emergency response activity. Surveillance devices, such as premise cameras and sound sensors, are arranged at special locations. A triage station is set up to prioritize victims based on their health conditions and to provide life-saving treatment and first aid. The hardcore of such a complex FR system is CC, where commanders monitor every FR's location, health and device status to safeguard their lives. Commanders closely analyze all the environment measurements from the surveillance system to watch out potential security threats and to prepare for precipitating events. They manage resource and personnel allocation globally. Frequently, commanders need to exchange dispatch commands and incident status with FRs in the field via voice communication. Occasionally, they inform FRs with various forms of data, such as building maps and fire hydrant locations and pull back feedback or reports from FRs. All of these life-critical and mission-critical applications belong to the category of distributed *real-time monitoring and control* (RTMC) applications. It is indispensable for a desired communication infrastructure to support RTMC applications. Therefore, we are interested in establishing a communication infrastructure for the incident wireless communication system to offer reliable communication paths betweens terminals and CC, so that FRs, surveillance devices and commanders are connected.

Unfortunately, such a communication infrastructure often does not exist. It may be destroyed, damaged partially, or incompatible with FR communication devices. Even if it exists, it may not be able to meet rigorous communication requirements of FR applications, such as persistent connectivity and reliability. The situation is further exaggerated by the fact that terminals with limited transmission ranges are very likely to lose connectivity from CC and each other, when scattered about a large incident scene, due to multi-path fading, interference and obstacles. Thus, we need an on-demand communication infrastructure tailed for FR applications, either established right before the incident FR system starts or gradually built during the operation. We summarize a list of requirements as follows. The communication infrastructure must be:

- *Cost-effective*: It should use as fewer relays as possible. Because of a limited supply of relays, it is better to reserve more relays for backup and recovery. Furthermore, deploying less relays means not only smaller installation overhead but also reduced retrieval overhead.

- *Adaptive*: It should provide connectivity to CC, for all the users and devices throughout their phased deployment and withdrawal. This requirement is of great value to relay placement problems since relays can only be installed before the network starts.

- *Robustness*: It should operate correctly even under all kinds of failure, such as device failures, device malfunction and broken links.

- *Sustainable*: It should operate for a desired period of time without or with minimum human intervention, considering the tedious burden and delay to locate and replace the battery-drained relays and the huge calamity brought by network disconnectivity before the manual fix.

In order to provision such a communication infrastructure for FRs, surveillance devices and CC, two means are taken simultaneously. First, multiple *base stations* (BSs) are installed at both CC and emergency vehicles, such as police vehicles and ambulances, residing in various locations around the incident scene, to increase the coverage of CC. Those BSs are interconnected with each other through satellites, cellular networks or mesh networks, which do not rely on the incident scene wireless communication system, due to reliability concern. Monitoring and control data can be transmitted between CC and terminals via any BS. Essentially, BSs are the wireless portals for CC. In presence of this tight relationship between CC and BSs, we also refer to the connectivity between terminals and CC as the one between terminals and BSs, *BS-connectivity* for short. Second, droppable relays are installed at various locations to further extend the transmission range of both BSs and terminals. Relays are communication devices, whose exclusive function is to forward packets for terminals, BSs and other relays, whenever needed.

In order to minimize the infrastructure cost, we need to strategically place BSs and relays. Unfortunately, BSs' locations cannot be controlled and finely tuned. They are mounted on top of emergency vehicles on call, which are parked at a safe distance from the incident scene, wherever available. On the contrary, relays can be strategically placed due to their portability. In this work, we only consider static relays. Unlike mobile relays equipped with motors and navigation system, static relays remain at the same place after being installed, till they are again physically accessed by human. Due to a limited supply of relay devices a FR system can afford, it is crucial to arrange relays in the optimal locations so as to eliminate unnecessary locations from coverage and to maximize the relay reuse among all the terminals. We call such problems *relay management problems*, which minimizes the the total number of relays by placing them at the optimal locations, while maintaining BS-connectivity and satisfying some other performance criteria, such as robustness and sustainability.

Prior work of art on relay management problems is mainly divided into two camps: (a) relay placement problems and (b) relay deployment problems. We summarize these two camps in Table 2.2. Relay placement problems usually lead to offline strategies, aiming for static networks. Before the network starts operation, a designated crew is dispatched into the field to install relays at the optimal locations calculated. It might happen the crew loses connectivity during movement process, since they are not terminals to be monitored and controlled. On the other hand, relay deployment problems usually lead to online strategies, mainly for mobile networks, wherein terminals of interest move around. There is no designated crew to install relays in advance. Mobile terminals are responsible to drop relays to maintain persistent connectivity whenever needed, during movement process. Majority of prior work for relay management problems fall into the

Table 1.1: Two Camps of Relay Management Problems

	Relay Placement (Offline)	Relay Deployment (Online)
Network dynamics	Static	Mobile
Who to install	Designated crew	Mobile terminals
When to install	In advance	During movement
Exemplary work	[3][4][5][6][7][8][9][10]	[11][12][13]

first category with theoretical flavor. The optimal relay placement problems can be stemmed from the theoretical problem of Steiner trees with the minimum number of Steiner points in 2D Euclidean space [3][4]. Later literatures address other interesting issues, regarding link heterogeneity [5][6], multiple BSs [7], fault tolerance [8], partial/full connectivity [6], one-tier/two-tier networks [5][9] and traffic patterns [10]. The second category, representing a small body of work, has more system and implementation flavor [11][12][13]. Heuristic and yet effective approach called *breadcrumb* is utilized to manage the relay deployment decision in real-time. In a nutshell, a relay is dropped, whenever the connectivity to BSs is going to break during movement, for breadcrumb approach.

Relay placement problems are extensively examined in a wide range of aspects. Here we summarize them according to different criteria, in terms of restriction on relay locations and network structure of the communication infrastructure. Regarding where relays can be placed, two subclasses exist: unconstrained relay placement and constrained relay placement. In *unconstrained relay placement*, relays can be placed anywhere in the network area. Simple radio propagation model of circular transmission range is assumed. In *constrained relay placement*, relays can only be placed at a subset of candidate locations [14][9]. This may be caused by the minimum distance required between adjacent relays to avoid region-correlated malfunction or crash. Or it is caused by forbidden areas where no relays can be installed, such as unreachable areas and water regions. Compared with unconstrained relay placement, constrained one deals with obstacles and radio irregularity [15] in a more elegant way, without complex manipulation. Regarding network structures, two subclasses exist: one-tier networks and two-tier networks. In *one-tier networks*, terminals and relays bear the equivalent roles in forwarding packets. However, in *two-tier networks*, relays forward all the packets, while terminals don't forward packets for other nodes [5][9]. Essentially, relays and BSs form a backbone network with terminals as the leaves in a two-tier structure.

FR systems do exhibit special properties of disconnected evolving networks which are either overlooked or treated primitively by the prior work for relay management problems. *Disconnected evolving networks* (DENs) are networks with the following properties: (a) time-evolving topologies, wherein the number of terminals and their positions change over time; and (b) connectivity to the BSs hardly exists without the help from relays. We have already made the case for disconnectivity. Let's see why FR systems have time-evolving topologies. FRs move from one location to another to execute different tasks. Temporary rescue and observation stations are set up on demand when at-the-spot medical treatment is necessary for victims and civilians. The mobile stations dissolve once the mission is accomplished. Premise cameras are present only at important locations, while the priority of a location may change over time. When these factors add up, FR networks exhibit time-evolving topologies.

When it comes to FR systems, prior work for relay management problems fall short of expectations. First, all the prior work for relay placement problems are designed for static networks, rather than evolving networks. Since relays are utilized by different topologies, they must be jointly managed across topologies. Second, relays and terminals have limited provision of battery. Therefore, relay placement should take energy consumption into account in order to meet the requirement on a desired network lifetime. Finally, breadcrumb approach for relay deployment problems is completely driven by FRs' mobility patterns. Though effective, the simple relay deploying strategy fails to maximize the relay reuse within and across multiple groups. In case of a limited supply of relays, not all the locations to be visited can be covered by relays. Also the simple relay deploying strategy may make the total number of relays far more than what is necessary.

Though having time-evolving topologies, FR systems are not completely unpredictable. A large amount of operation planning and scheduling knowledge can be exploited in relay management, dealing with the unique requirements for FR communication infrastructures. Not only do we have rudimentary knowledge about potential network configurations, such as where the triage station is based and where premise cameras are placed, but also we know the coarse-level mobility patterns of FRs, e.g., the set of locations FRs will pay visit to, from FRs' dispatch and task assignment. The network topology and logistic information can be utilized by relay manage problems to pinpoint the optimal relay positions in the incident scene. We draw the highlights of the thesis work in Figure 1.2. Since relay placement problems have very distinct prior knowledge and request from relay deployment problems, we develop different models, *disconnected polymorphous networks* (DPNs) for placement problems and *disconnected dynamic networks* (DDNs) for deployment problems.

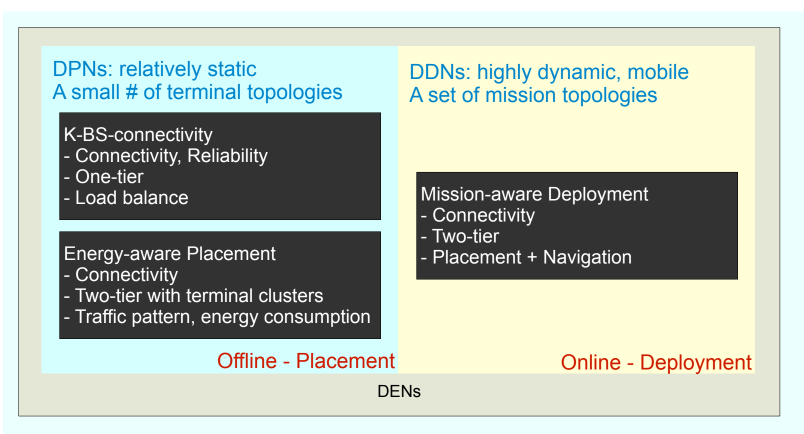


Figure 1.2: Highlights of Thesis Work

DPNs are DENs with a limited number of terminal topologies. DPNs model relatively static topologies, where terminals are mainly static or move occasionally. However, terminals may not be active all the time, due to sleep/wake schedule, or joining/leaving networks, rendering FRs' network polymorphous with multiple topologies. Two relay placement problems are studied, considering sufficiency of devices' battery and duration of FR networks. If the devices' energy provision is quite abundant, or the network runs for a short duration with light traffic, we only take BS-connectivity requirements into account when optimizing the relay placement. Our proposed offline relay placement algorithm WPA is able to handle both connectivity and reliability requirements by k-BS-connectivity, while balancing the load among all the available BSs. When $k = 1$, only connectivity is considered; otherwise, reliability is enforced for fault tolerance to deal with terminal or relay failure. If the energy provision is tight, especially when the FR networks have to survive several days' or weeks' operation with heavy traffic, it is a different story. It is vital for the FRs' communication system to continuously function for a desired network lifetime, after being deployed. Therefore, traffic patterns and energy consumption must play a role and we need to schedule the relay placement and traffic routing simultaneously in an energy-aware fashion. Realistic energy consumption models are plugged into the optimization framework, under two different transmission control mechanisms: power control and data rate control.

Unlike DPNs, DDNs are designed for mobile networks, where terminals, i.e., mobile devices carried by FRs, are mobilized around the incident scene. If we examine the whole timeline, there are an infinite number of terminal topologies, which is impossible to be captured by any algorithms. To extract our prior knowledge on mission logistics, DDNs model FR mobility patterns by mission topologies. A mission topology is the set of task destinations, where a FR mobile group sojourns a while and executes tasks. The mission topologies are then utilized to determine the optimal locations of relays, by maximizing the relay reuse within a mobile group of different tasks, and among mobile groups of different missions. Two mission-aware relay placement algorithms are proposed to take advantage of mission prior. Because we strategically place relays to maximize the reuse opportunity, FRs may have to detour from the original shortest paths towards the task destinations to those deploying places. The online navigation algorithm controls the detour with reasonable traversal overhead. It is easy to see that the relay deployment algorithm is a joint relay placement and online navigation algorithm.

In this thesis, we will focus on *constrained* relay management problems, due to the flexibility of constrained relay placement to deal with obstacles, radio irregularity and minimum distance between adjacent relays. We claim that: *On-demand relay communication infrastructure is feasible, for both relay placement and deployment problems. We propose approximation algorithms for the optimal constrained relay management problems, to maintain base station connectivity and to meet diverse performance requirements for FR systems.*

Here is the organization of this thesis work. In chapter 2, we present an overview of first responder systems, for a better understanding of the background, motivation and objective of this work. In chapter 3, we clarify various assumptions and formally define DPNs and DDNs. In chapter 4, we present our work on k-BS-connectivity for DPNs to meet both connectivity and reliability constraints. In chapter 5, energy-aware relay placement algorithms for DPNs are studied to satisfy network lifetime requirement. In chapter 6, we investigate in detail the mission-aware relay deployment algorithms for DDNs. Following a thorough examination of related work on relay management problems in chapter 7, we provide the conclusion remarks and summary of our contribution in chapter 8.

Chapter 2

Overview of First Responder Systems

First responders represent personnel from various public safety agencies providing support for law enforcement, emergency medical service, fire control, hazardous material management, public health service, and other mutual aid. The US Department of Homeland Security exerts among its top priorities on stronger information sharing and infrastructure protection with inter-operable communication and equipments for FRs in the incident scene [1]. A *first responder system* is established on demand for real-time life protection and intelligent mission execution, providing indispensable services far beyond what is offered by traditional voice communication. A FR system usually comprises an incident command system and a communication system. In this chapter, we will give an overview of first responder system, focusing on the networking aspects. Also, we will present our prior work on FR mobility patterns, whose special properties are leveraged by relay deployment algorithms in Chapter 6.

2.1 First Responder System Framework

Figure 1 shows an example of the first responder system composed of communication networks of different scales (Internet and mobile ad hoc networks) with heterogeneous devices from two agencies: firefighters and EMS staff. Since incident management comprises a wide array of operations to facilitate prevention, preparedness, response, and recovery [16], we will examine the FR system from three aspects: incident wireless communication system, network structure, and public safety application model [17].

2.1.1 Incident Wireless Communication System

To enable real-time information exchange and sharing, the incident scene is closely monitored, controlled, and managed via wireless communication. A *wireless communication system* is a multi-hop mobile wireless network established on demand, with or without existing infrastructure support. It is formed by terminals, relays, the command center (CC) and base stations (BSs) mounted on the emergency vehicles around the incident scene, as shown in Table 2.1. CC performs the essential function of monitoring and control. It

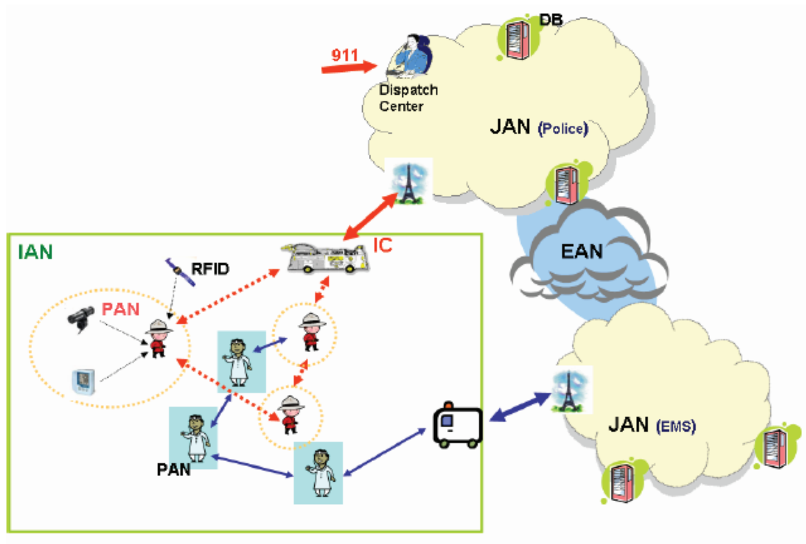


Figure 2.1: First Responder Network Composition

manages the global resource allocation, task assignment, etc. A base station, or an access point, is usually attached to CC to enable wireless communication. To increase CC's coverage, multiple BSs are installed at other emergency vehicles, besides CC, such as police vehicles and ambulances, residing in various locations around the incident scene. Those BSs are interconnected with each other via reliable control channel, which is independent of incident scene wireless communication system. *Terminals* are multi-functional wireless devices distributed in the incident scene, which perform more functions than relaying packets. Terminals need to extract and process information, and communicate with other terminals and CC frequently. Relays are communication devices, whose exclusive function is to forward packets for terminals, BSs and other relays, whenever needed. It is unnecessary to maintain BS-connectivity for relays, if they are isolated while terminals and BSs are replaced.

Table 2.1: Wireless Communication System

Name	Description
Command Center (CC)	Real-time monitoring and control center
Base Station (BS)	Wireless portal for CC, interconnected
Terminal	Multifunctional wireless device monitored and controlled by CC
Relay	Wireless device provisioning terminals with connectivity to BSs, thus CC

To support intelligent mission execution, terminals of different functions are deployed or dispatched around the incident scene. Some are mobile, such as handheld devices carried by FR, mobile robots and health sensors attached to patients. Some are static, which may occasionally be displaced if needed, premise cameras for example. According to their functions, wireless terminals are divided into two categories: interactive communication devices and surveillance sensors. Interactive communication devices are the handheld devices carried by FRs, which offer diverse two-way communication services, ranging from voice, text, image, video, to data. On the other hand, surveillance sensors perform one of most critical functions, distributed real-time monitoring and control. All the sensors should have connectivity to CC, where samples and measurements are processed, integrated and analyzed. This centralized approach is not only effective for global resource and asset coordination, management and optimization, but also critical for logging information for behind-the-scene and after-the-scene analysis and retrospect.

By the monitoring objects, surveillance sensors are further divided into three classes: environment sensors, human sensors and equipment sensors. First, environment sensors monitor the environment change and detect life signal. Live video feeds from camcorders mounted on FRs' helmet or from static premise cameras installed on demand are sent back to CC for logging and large-scale image and video analysis to detect life symptom and unauthorized trespass. Fire sensors continuous monitor the air composition and temperature to watch the progress of fire control and to alert potential threats. Second, human sensors monitor all the facets of human conditions. Heath sensors, either embedded to officers' suits or sticked to patients' body, continuously measure their heartbeat rate, breath rhythm, blood pressure, etc. Location sensors pinpoint objects' location, using technologies such as GPS, WiFi-based indoor localization, or RFID. Whenever symptom of severe health condition is observed, e.g., heart attack or suffocation, a rescue or backup team is sent to the officers' or patients' location for timely treatment. Third, equipment sensors continuously monitor the devices' status. If the resource supply of a device falls below, such as oxygen for oxygen tanks, carbon dioxide in fire extinguishers, replacement is sent or FRs are recalled to ensure operation safety. In addition, device sensors may be able to detect malfunction or misbehavior by cross-layer or cross-device analysis.

Table 2.2: Category of Terminals

Interactive Communication		
FR Communication Device (FRCD)	Voice, text, image, video, data	
Surveillance sensors		
Environment	Chemical, explosion, temperature, camcorder, sound, magnetic	
Human	Health	Heartbeat, breath rhythm, blood pressure, temperature, perspiration
	Location	GPS, WiFi indoor localization, RFID, accelerometer
Equipment	Oxygen tank, fire extinguisher	

2.1.2 Network Structure

Hundreds and thousands of devices may be integrated together in order to support the seamless end-to-end user communication in first responder systems. This complex FR system can be decomposed into four types of network: PAN, IAN, JAN and EAN according to their scale, function and characteristics [18].

- Personal Area Network (PAN): This is a small-scale wireless network used for physical communication among devices (e.g., video camera, RFID) and sensors (e.g., heart rate monitors), carried by or embedded in clothing of a person, usually a FR. PANs are widely used today for health/equipment monitoring and environmental surveillance. The aggregated statistics collected in PAN are first recorded in First Responder's Communication Devices (FRCD), which then relay event notification back to CC and other fist responders periodically or upon query. The devices are normally plug-and-play devices, supported via secure connection and high data rate over a short transmission range.
- Incident Area Network (IAN): This is a multi-hop ad hoc wireless network deployed on demand in the incident area, where communication infrastructures do not exist or have been destroyed. The devices include, but are not limited to, mobile devices (e.g., mobile robotics and BSs mounted on vehicles), portable devices (e.g., handheld devices, temporary RFIDs, and droppable relay devices), and surveillance sensors which are not associated with PANs. CC acts as a gateway to the outside of IAN.
- Jurisdiction Area Network (JAN): This is the private network of agencies, responsible for secrecy database access, certificate management, task dispatch, and resource mobilization via CC.
- Extended Area Network (EAN): This is the backbone network for interconnecting JANs and IANs, as well as provision of public Internet access (e.g., real-time weather from public sites).

Most of the wireless communication researches focus on incident scene management in IANs composed of diverse number of PANs. We are mainly interested in establishing a reliable communication infrastructure for IAN in this thesis. Here we summarize PANs' and IANs' characteristics in Table 2.3.

2.1.3 Public Safety Application Model

To support informed day-to-day routine and emergency response, diverse multimedia applications (e.g., texts, audios, graphics, videos and interactive commands) are integrated in FRCDs. The information includes three types of content: voice, data, and video. Generally the information is independent of communication devices and the underlying network protocols.

Table 2.3: Summary of PAN and IAN

	PAN	IAN
Range	1-10 meters (1-2 hops)	10-1000 meters (multi-hop)
Mobility	Fixed relative to FRCD	Mobile
Scale (# devices)	1-10	10-10000
Technology	UWB, bluetooth	802.11abgn, 3G, CDMA
Devices	Camera, heart rate sensor, FRCDs	Senors, FRCDs, droppable relays, BSs
Administration domain	Single	Mixed

1. Voice: conveys commands, reports, instructions and advices among FRs and CC. Voice, being the most popular data format, can be transmitted in a peer-to-peer, unicast or multicast mode, which has the strictest delay requirement;

2. Data: of texts, graphics and other digital formats inform first responders about both the environment and peer condition. Example data are alerts, locations of residential habitats, fire hydrant distribution, peer status, process manuals, electronic crime reports and programmable commands. Some data even come from authorized databases in JAN or EAN, such as the state motor vehicles database, and building information.

3. Video: bolsters surveillance and audit by recording the entire history and enables remote conference and assistance. With the advance of wireless and peer-to-peer technology, online video communication and offline video storage with high bandwidth, low delay and small jitter is possible.

2.2 First Responder Mobility Model

The quality of communication and resource allocation highly depend on network topologies, which are in turn determined by the FRs' movement in the incident scene. Bai et. in [19] show that different routing protocols present unique properties under different mobility models, such as throughput and control overhead. There is emerging consensus that protocol design and distributed systems should be studied under realistic and scenario-specific mobility models, considering temporary dependency [20], spatial dependency [21], geographic constraints [22] and empirical measurement of social effects [23]. This means that in simulation-based evaluation of routing and application protocols for FR incident scene wireless networks, it is crucial to use authentic mobility models.

We state that *not only do we need to study realistic mobility models for the purpose of protocol and system performance evaluation, but also we need to exploit these special characteristics of FRs' mobility patterns in* *the distributed system design.* In this thesis, those properties are widely used in relay deployment algorithms for persistent BS-connectivity.

Major mobility elements considered in mobility models are destination selection (frequency, scope and preference), trajectory calculation (criteria such as the minimal length, the shortest trip duration and safety), moving speed (temporal correlation, spatial correlation and speed regulation), pause between successive trips (frequency and duration) and coverage region. Next, we will present how those mobility factors are largely determined by transportation means, intrinsic environmental constraints and operational logistics [24][25][26] [27]. Finally, we summarize how we exploit the special properties to build a reliable communication infrastructure cost-effectively, which offers persistent BS-connectivity for FRs on the move.

2.2.1 Transportation Factor

First responders have three major modes of transportation:

- Land (e.g., foot, bicycle, horseback, various vehicles)
- Water (e.g., speed boat, water scooter)
- Air (e.g., helicopter, parachute)

Because of prevalence of transportation on land at a majority of the incident scenes, we will focus on it from now on. Means of transportation influence mobility in four aspects, namely nominal speed, coverage area, pause time and passable trajectory. Personnel on foot are usually responsible for a small region, wherein they move around at low speed of several meters per second and pause intermittently. They go around obstacles and possibly along arbitrary trajectories. Personnel riding on bicycles or horsebacks move in a region with medium size and at medium speed with small speed variation. Their trajectories are not as flexible as those on foot and somewhat constrained to existing routes, streets and paths. Personnel traveling in motor vehicles move around large coverage areas at relatively high speed with large speed variation. Their trajectories are strictly confined to roads and highways [28].

2.2.2 Environmental Factor

Environmental factors play an intrinsic role in refining mobility characteristics in a dynamic way, such as speed and trajectory. Inside physical structures, predefined hallways limit personnel's moving directions. In urban outdoor areas, paths and roads define the traversable routes for first responders on bikes and vehicles to follow. Unexpected physical obstacles alter first responders' preplanned moving routes in an ad hoc manner. Visibility conditions affect the speed at which first responders move. They may slow down due to low visibility. Hazards, such as fire, airborne pollution and explosion, could prevent first responders from entering an affected region unless they are adequately equipped and protected. For those individuals authorized to enter, they could move in the affected regions slowly with caution. An incident scene environment often changes frequently and abruptly. This dynamically triggers route recalculation process and adjusts moving trajectory.

2.2.3 Operational Factor

Contrary to environmental factors, operational factors influence mobility primarily macroscopically, such as destination selection. First responders are mission-oriented. They perform various kinds of tasks, such as inspection, video scanning, first aid, rescue, and patrol. They move around in order to complete mission-critical tasks, which often correspond to events and incidents geographically distributed in the incident scene, such as fire in the storage room, and victim trapped in the basement. Unlike other social networks studied (campus, vehicle and bicycle networks), operations of first responders are multi-organizational, collaborative and responsive.

Multi-organizational Operations

Task division among agencies from multiple organizations leads to broad mobility dissimilarity across groups. Personnel sharing the common interests with the same mission in an organization form a group or subgroup. For instance, EMS staff and firefighter are two groups and they have different interests and activity areas. EMS units track all the patients, provide first aids to victims and transport the seriously wounded to nearby hospitals. Their territory in the incident scene is usually restricted to the border. On the other hand, firefighters penetrate into the incident scenes, like a building on fire, extinguish fires and extricate the injured civilians. As we see, their mobility patterns are mainly determined by their respective interests. They select different sets of locations as destinations; they pause and move for different durations during the course of execution and resolution. Even inside one agency, officers have command structures and are assigned to different task subgroups, thus with different execution goals, destinations, coverage regions and task sequences.

Collaborative Operations

Due to precise role and task division, complexity of missions and criticality of incidents, first responders need to cooperate with one another. Formed as groups and subgroups, they typically proceed together in close proximity, safeguard one another from danger and work together on emergency situations. Therefore, mobility of a first responder affects mobility of other nearby responders in the same group or in different groups with temporal collaborative relationship. It is worthwhile to point out that the coordination relationship we consider can change frequently, depending on time-varying emergencies, their respective missions and resources availability. For example, first responder A, who is originally in Group 1 of EMS, may be reassigned to Group 2 of EMS, in a response to the lack of workforce inside Group 2. Group 3 of EMS may cooperate with Group 1 of firefighters to transfer victims from the spot and to provide first aid.

Responsive Operations

Incident scenes are usually hard to predict beforehand. Even though commanders have a rough estimation about incidents and events around the incident scene via civilian's reports, the surveillance system and initial incident scene investigation, FRs may need to respond to the incidents ad hoc, prioritizing missions according to their relative emergency, criticality, feasibility and overhead. Hence, their mobility patterns are only known at the macroscopic level, but are uncertain at the microscopic level. More specifically, the knowledge of their second-by-second location is unavailable.

2.2.4 Implication to Relay Deployment Problems

First, we will show why deploying additional relays is indispensable to maintain persistent BS-connectivity by an experiment measurement based on realistic mobility patterns designed for FRs in the incident scene. Then we will elaborate the special properties which are exploited by our relay deployment algorithms to construct a cost-effective communication infrastructure.

FRs, carrying handheld communication devices with limited transmission ranges, are likely to lose connectivity to CC, when scattered about a large incident scene, due to multi-path fading, interference and obstacles. We verify the problem of intermittent BS-connectivity under a rescue and response mission without help from droppable relays, for a FR mobile ad hoc network in Figure 2.2. The mobility trace is generated based on CORPS model [26], which is an event-driven mobility model for first responders in the incident scene. In the examined scenario, 25 first responders are dispatched into 1000m*1000m network in two phases. A base station is installed at location (1000, 500). Transmission range is set to 250m. A connection $c(i, BS)$ between terminal i and BS exists at time t if there is a communication path between i and BS at t, either direct or multi-hop through other terminals. BS-connectivity at time t is defined as $\frac{|\{i|i \in N, c(i,BS) \ exists \ at \ t\}|}{|N|}$, where N is the set of FRs and $|N|$ denotes the size of set N. We sample the BS-connectivity every 1 second during the whole simulation period. $F(c)$ is the portion of sampled time instances, at which time BS-connectivity is lower than c. As we can see, on average 3 out of 5 FRs have no connectivity to BS for more than 90 percent time. This is because FRs tend to move far away from BSs during mission. DDNs poses great challenges for the monitoring and control applications, thus the safety of first responders and their effectiveness in executing tasks and missions. Therefore, it is very important that relays are dropped to maintain persistent BS-connectivity, while FRs are moving around the incident scene.

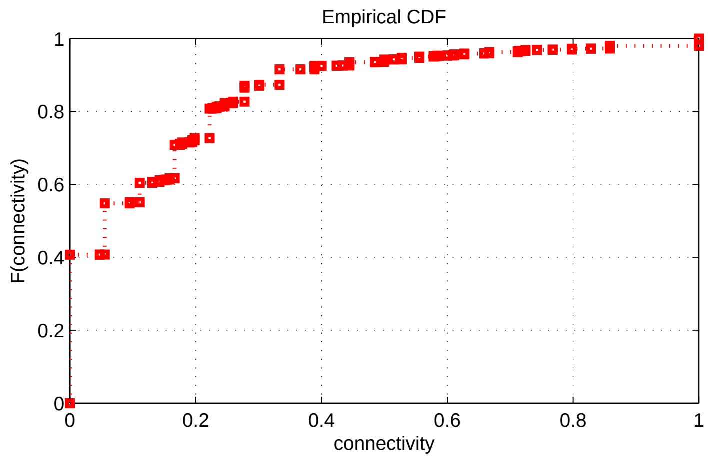


Figure 2.2: Base Station Connectivity

Even though it is hard to predict the exact time-variant moving trajectory for FRs in the field, we are able to master their macroscopical mobility patterns for each FR group. Thanks to close coordination and correlated movement, all the terminals in a mobile group can be simplified as a single virtual terminal. Also, FRs are mission-oriented and their activities are planned in a centralized way, to maximize their performance and effectiveness. Based on this predictability, there is no need to cover every location in the incident scene and the communication infrastructure only needs to cover the set of locations associated with FRs' missions. We shall exploit this knowledge of coarse-level mobility patterns in mission-aware relay deployment algorithms to maximize the relay reuse within a mobile group and among different mobile groups. Finally, uncertainty of microscopical mobility patterns requires that any relay deployment strategy should not rely on temporary network topologies formed by mobile terminals. In addition, any coupling among different mobile groups should be eliminated, since this also infers temporary terminal topologies.

Chapter 3

Assumptions and Problem Formulation

In this chapter, we clarify various assumptions and formally define our relay placement and relay deployment problems.

3.1 Assumptions

All the relays are static and they are always turned on after being installed. Because we focus on the constrained relay management problems, relays can only be placed at a subset of candidate locations. We define the set of candidate locations as a *relay placement scheme*, denoted by $\mathcal{L}$. The final set of places where relays are installed is called a *relay placement solution*, denoted by $\mathcal{R} = \{r_1, \cdots, r_R\}$. We will further refine the definitions for relay deployment problems. For constrained relay management problems, $\mathcal{R} \subseteq \mathcal{L}$. Due to the nature of relay management problems, we treat all relays equally. In another word, any two relays are interchangeable and any relay can be placed at a selected location.

Similar to relays, base stations are static and they are always turned on. BSs are connected with each other through private control channels. This connectivity ensures that the command center can communicate with a terminal via any BS the terminal is connected to.

All the relays and terminals have a limited amount of battery, while BSs have an unlimited supply of energy, because they can be powered by the attached vehicles. A terminal u is equipped with a battery with initial capacity E_u. Relays are equipped with batteries with the same initial capacity E. If a node, including all relays and some terminals, appears in several topologies, its battery provision must be shared in all the containing topologies.

A node can be a terminal, a relay or a BS. In a network topology, one-hop wireless link ab between node a and b exists if wireless communication can succeed in both directions, and we use $I(ab)$ to denote the existence of one-hop wireless link ab. All the relays have the same transmission range. Thanks to constrained relay management, terminals and BSs could have different transmission ranges, which only affects the set of defined wireless links, not our model complexity. Nevertheless, for relay deployment problems, we assume all the nodes have the same transmission range. The reason is as follows. Suppose the transmission ranges for terminals and relays is r and R, respectively and $r \neq R$. If $r < R$, there are two means to maintain persistent BS-connectivity. First, FRs could drop relays at distance r. Second, FRs could have a relay physically bundled with their communication device and any communication goes through the bundled relay proxy. Hence, they can drop relays at distance R, which is more cost-effective. If $r > R$, they still have to drop relays at distance R so that the dropped relays have wireless links to relays in previous hops. In a summary, it makes sense to assume all the nodes have the same transmission range in relay deployment problems.

3.2 Relay Placement Problems

In this section, we will formally define disconnected polymorphous networks. Following the discussion of BS-connectivity for DPNs, we formalize the relay placement problems.

3.2.1 Disconnected Polymorphous Networks

Disconnected polymorphous networks (DPNs) are networks with the following properties: (a) they are present with a small number of topologies, wherein both the number of terminals and their positions alter; and (b) connectivity to BSs hardly exists. Let us look at two examples to see why DPNs are prominent in FR systems. In Figure 3.1, we show a static surveillance sensor network composed of camcorder sensors (C), hazardous sensors (H), and BSs (B). Camcorders are installed near the incident scene premise to record all the human activity. Whenever motion is detected, full video recording capability is initiated to log all the suspicious and normal activities. Chemical and fire sensors, called hazardous sensors, are placed around unique landmarks to detect fire or pollution events, periodically. Whenever they are activated, sensors report aggregated measurements to BSs for analysis and control. First, note that sensors are generally planned to cover only regions of interest, possibly isolated. Hence, they are likely to be disconnected from BSs or from each other. Second, though all devices coexist in the same physical space, the network may form two topologies, i.e., topology 1 excluding hazardous sensors and topology 2 excluding camcorders. This can be caused, either by unsynchronized wakeup schedule or by packet forwarding policy. In the first case, camcorders are activated whenever motion is detected and hazardous sensors are activated for 1 minute every 5 minutes. For most of the time, two types of devices are not activated simultaneously. In the second case, camcorders are disallowed to forward traffic from hazardous sensors and vice visa, to remove any dependency between two sensor networks and to maximize the possibility for in-network aggregation. Therefore, the static surveillance sensor network can be polymorphous. DPNs arise more naturally in wireless networks formed by handheld

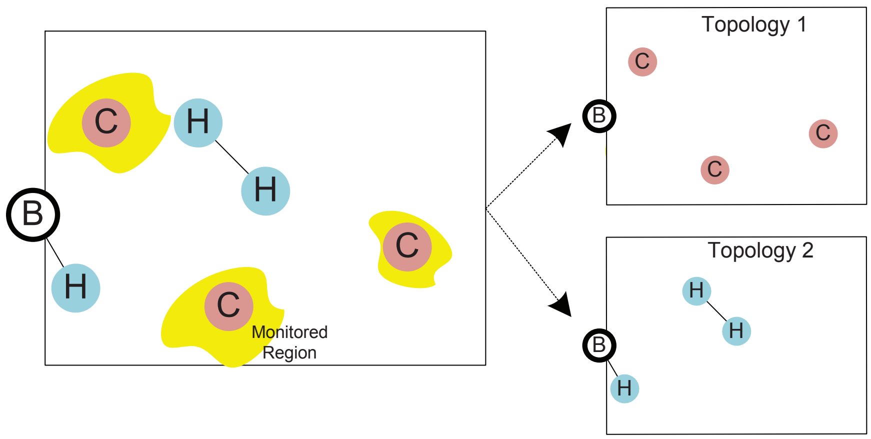


Figure 3.1: DPNs: Sensor Networks

devices carried by FRs or health sensors attached to patients, which we refer to as mission-driven networks. When missions start, the relevant devices join the network. Once the missions are completed, these devices will leave the network, and the corresponding terminal topology dissolves. In Figure 3.2, we show a first responder (FR) network with two topologies. The first topology forms during the rescue phase, composed of patients (P), EMS staff (E) and BSs (B). The second topology forms during the investigation phase, composed of hazardous control staff (H), bomb experts(B) and BSs (B). Commanders need to monitor every personnel's location, health and working environment to safeguard their lives and manage resource allocation globally. Also commanders need to send commands and documents to FRs. These monitoring and controlling data are exchanged between the commanders and terminals via any BS. Due to limited

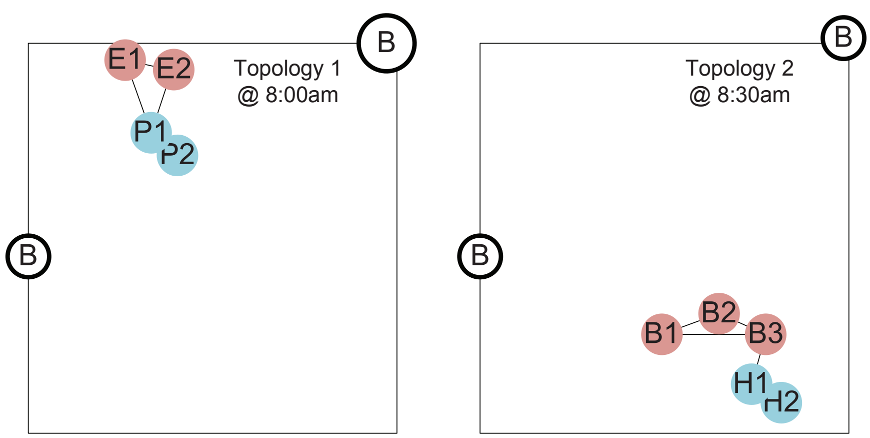


Figure 3.2: DPNs: First Responder Human Networks

transmission range of handheld devices, the large areas of the incident scene, radio interference, and group movement, terminals are likely to be disconnected from BSs and from other teams. They may join or leave the networks for different missions, or relocate from one place to another to execute different tasks. This causes both the set of terminals and their positions change. Therefore, the FR mission-driven networks are both disconnected and polymorphous.

Here, we assume that we can infer the set of network topologies for a DPN. A topology is represented by the locations of all the active terminals, which can communicate with each other if they are close. In the example of static sensor networks, we can infer the topologies by sensors' wakeup schedule and packet forwarding policy. In the example of FR mission-driven networks, we can infer the topologies from the task assignment during incident planning and preparation phase. [1]

A DPN instance is represented by the tuple $(\mathcal{TS}, \mathcal{B})$. $\mathcal{TS}$ is a set of T *terminal topologies*, $\mathcal{TS} = \{\mathcal{T}^1 \cdots \mathcal{T}^T\}$, where t_{th} terminal topology $\mathcal{T}^t$ contains n^t terminals in the Euclidean plane $\mathcal{R}^2$ (or $\mathcal{R}^3$). $\mathcal{B}$ is a set of base stations, $\mathcal{B} = \{B_1 \cdots B_S\}$.

When the communication infrastructure runs for a long period of time with high traffic load and relative tight provision of battery at each node, we are also interested in how long the infrastructure is going to last. This notion is captured by the network lifetime, which is the time a node drains its energy, thus preventing the correct network functioning. Unlike conventional sensor networks that run for years once deployed, FR networks only run for a short period of time. Hence we are interested in meeting a desired network lifetime, rather than maximizing the network lifetime. Usually, commanders have an estimate about a desired network lifetime, say L. Because the energy consumption rates for different topologies are different, network lifetime depends on the relative period for each topology in a DPN instance. We associate a terminal topology with a fractional value f_t, which means that for f_tL period of time, the network is present with topology $\mathcal{T}_t$. It's easy to see that $\sum_{t=1\cdots T} f_t = 1$.

A terminal u sends data traffic to CC at the rate γ_u^t in topology $\mathcal{T}^t$. For simplicity, we only consider upload traffic from terminals to the command center. It is not hard to incorporate download traffic of the reverse direction into our formulation.

3.2.2 BS-connectivity

A special vertex called command center C is included. One-to-one connectivity between C and any BS is established. We replace the pairwise connectivity between BSs by this one-to-one connectivity between C and every BS, as shown in Figure 3.3. If a terminal is connected to C, it must be connected to at least one

[1] If some terminals appear in several topologies, but at different locations, connectivity may be lost when they are moving from one location to another by our relay placement algorithms. This is generally acceptable if the persistent BS-connectivity for those terminals is not required.

BS. So in the future, *we use connectivity to C, whenever we mean connectivity to at least one BS*. Moreover, rather than contracting BS vertexes to C, adding C while keeping BS vertexes enables us to state the load on a BS explicitly for load balance purpose.

Given a placement solution $\mathcal{R}$, a *network topology* is the union of a terminal topology $\mathcal{T}^t$, $\mathcal{B}$, $\mathcal{R}$ and C. In total, there are T network topologies defined by T terminal topologies in a DPN instance. Given $\mathcal{R}$, we define the network topology graph for terminal topology $\mathcal{T}^t$ as $G_R^t = (V_R^t, E_R^t)$,

$$\begin{aligned} V_R^t &= \mathcal{T}^t \cup \mathcal{B} \cup \mathcal{R} \cup \{C\} \\ E_R^t &= \{ij | I(ij), i, j \in \{\mathcal{T}^t \cup \mathcal{B} \cup \mathcal{R}\}\} \cup \{CB | B \in \mathcal{B}\} \end{aligned}$$

In topology $\mathcal{T}^t$, a terminal has k-BS-connectivity if it has k (terminal, relay)-disjoint communication paths to C in G_R^t. k-BS-connectivity ensures that if at most $k-1$ relays or terminals fail, a terminal can still connect to the monitoring center C. It is acceptable to have more than 1 path out of k paths via a BS, according to this definition. The rationale behind this design is that the number of BSs is usually small and they are designed to be more reliable than terminals and relays. If a terminal can communicate with a BS directly, this connection is reliable and we assume the terminal is already k-BS-connected. With k-BS-connectivity, every terminal has reliable communication paths towards C. Reliable communication between two terminals a and b can also be achieved by two reliable communications, one between a and C and one between b and C. Compared with regular k-vertex-connectivity among any pair of terminals and BSs, k-BS-connectivity requires a smaller number of relays. In Figure 3.3, we illustrate 2-BS-connectivity for a network topology. All terminals in Figure 3.3(a) are 2-BS-connected since they all have 2 (terminal, relay)-disjoint paths to C. However, terminal 1 in Figure 3.3(b) is not 2-BS-connected.

3.2.3 Problem Formulation

Now, let's formulate our relay placement problems for DPNs.

Relay Placement to Maintain k-BS-connectivity for DPNs

The goal is to deploy the minimum number of relays at a subset of candidate locations, so that for any topology in a DPN, all the terminals have connectivity to BSs, reliable if required. Formally speaking, we want to find a placement solution $\mathcal{R} = \{r_1 \cdots r_{m^*}\}$ ($\mathcal{R} \subseteq \mathcal{L}$) with the minimum number of relays m^* so that for any topology $\mathcal{T}^t$, all n^t terminals in G_R^t are k-BS-connected. Also the load on base stations is balanced.

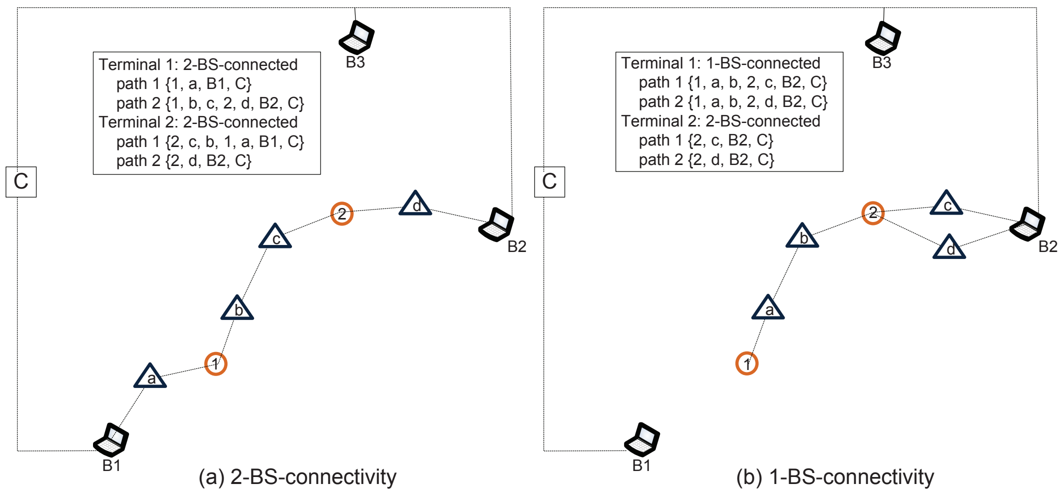


Figure 3.3: k-BS-connectivity, Terminals (1,2): circles; Relays (a,b,c,d): triangles.

Energy-aware Relay Placement to Maintain BS-connectivity for DPNs

The goal is to deploy the minimum number of relays at a subset of candidate locations, so that for any topology in a DPN, all the terminals have connectivity to BSs, with the expected network lifetime L. Formally speaking, we want to find a placement solution $\mathcal{R} = \{r_1 \cdots r_{m^*}\}$ ($\mathcal{R} \subseteq \mathcal{L}$) with the minimum number of relays m^* so that for any topology $\mathcal{T}^t$, all n^t terminals in G_R^t are BS-connected. In addition, no terminals or relays drain their battery prior to L.

3.3 Relay Deployment Problems

In this section, we will introduce mission topologies for disconnected dynamic networks, which are used to model the mission knowledge in mobile first responder networks. After presenting the design principle for on-demand communication infrastructure to maintain persistent BS-connectivity, we will formulate relay deployment problems.

3.3.1 Disconnected Dynamic Networks with Mission Topologies

Disconnected dynamic networks (DDNs) are disconnected evolving networks with the following properties: (a) there are multiple mobile terminals continuously moving around; and (b) connectivity to BSs hardly exists. If we know nothing about the mobility patterns of mobile terminals, we cannot optimize the online relay deployment decision, except using the most straightforward “breadcrumb” approach: deploying a relay whenever BS-connectivity is going to break.

Fortunately, FR’s movement is mainly driven by missions, featured with being organized, cooperative and responsive. FRs are agencies from various organizations with different missions. FRs with the same mission form a group. Different groups have diverse execution goals, targets and task sequences. Tight cooperation ensures that FRs from a group, always moving within close proximity, shield one another from danger and coordinate in emergency situations. For instance, two-in-two-out policy mandates that firefighters never go into a dangerous situation in a fire or rescue incident alone. Nevertheless, FRs may need to respond to an incident spontaneously, prioritizing tasks according to their relative emergency, criticality, feasibility and overhead. Being responsive makes deduction of instantaneous terminal topology infeasible. These properties of organization and cooperation can be naturally modeled by virtual terminals and their mission-defined macroscopic-level movement. At the same time, no microscopic-level movement is modeled, which agrees with the responsiveness property.

First, let's look at virtual terminals and understand their relationship with public safety operations.

Definition 1. *A **virtual terminal (v-terminal)** is the abstraction of a mobile group.*

There are two reasons for this abstraction, considering dynamics and coordination within FR groups. First, relative terminal locations within a mobile group in the field may vary from time to time; thus it is hard to model or even capture this dynamics rigorously. Second, abstraction to a v-terminal mirrors the real labor division within a group. Mobile terminals in a cluster usually have good connectivity among themselves for group communication, such as message multicast and distributed storage. FRs also have the ability to recover the lost connectivity among themselves by slightly moving around. As long as one terminal in a group is BS-connected, the other terminals are BS-connected automatically. In addition, FRs within a group have very detailed division of labor. Being relocated, an officer may take the special role of maintaining BS-connectivity, either strictly moving under the cover of the current communication infrastructure, or deploying additional relays. Other officers will simply match the movement of the special node by moving in its close proximity. There might be several physical terminals taking the turn for this special role. The virtual terminal can be viewed as virtualization of this role. In the following discussion, we use mobile groups and v-terminals interchangeably.

We assume that there are N mobile groups $\mathbf{G} = \{G^1, \cdots, G^N\}$, where G^i is the i^{th} group. Now, let's move on to mission topologies, which capture the mission prior.

Definition 2. *A **mission** is associated with each group, containing a list of geographically distributed **task destinations**.*

Definition 3. *A **mission topology** $\mathcal{MT}^i =< B^i, DL^i >$ for group G^i specifies G^i's starting base station B^i and a list of destinations DL^i, where G^i will visit, stay, and execute tasks. $DL^i = \{dest^i_1, \cdots, dest^i_{p^i}\}$.*

In the incident scene, often all FRs are dispatched from BSs. Different groups may initiate from different BSs. A mobile group can visit the locations in the destination list by any order, visit only a subset of locations or visit a location multiple times while they are dispatched. But all the places to be visited must be included in the list. The more accurate is the destination list, the more optimized is the performance of M-breadcrumb, since unnecessary locations aren't covered. For convenience, we use $DL^{i'}$ to denote $\{B^i\} \bigcup DL^i$.

Definition 4. *A* ***disconnected dynamic network (DDN) instance*** *is the set of mission topologies for N mobile groups, $\{\mathcal{MT}^1, \cdots, \mathcal{MT}^N\}$ with a list of M BSs $\{B_1, \cdots, B_M\}$.*

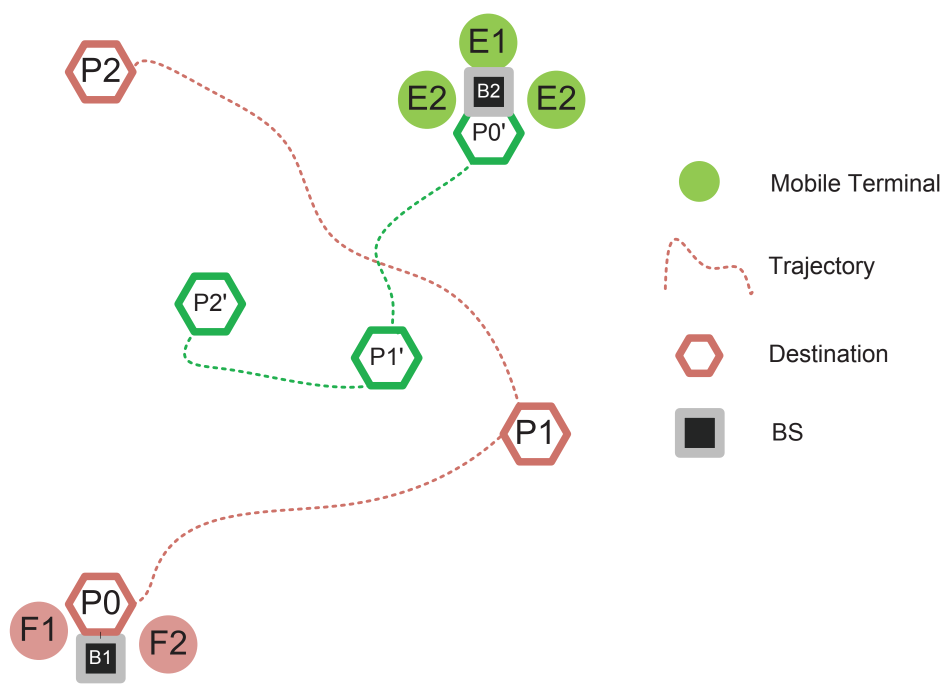


Figure 3.4: Mission Topologies. $\mathbf{G} = \{G^1, G^2\}$; $G^1 = \{F1, F2\}$; $G^2 = \{E1, E2, E3\}$; $\mathcal{MT}^1 =< B1, \{P1, P2\} >$; $\mathcal{MT}^2 =< B2, \{P1', P2'\} >$.

We show a DDN instance with 2 groups in Figure 3.4. The first group of 2 firefighters starts from $B1$ and visits locations $P1$ and $P2$. The second group, composed of 3 emergency medical staffs, starts from $B2$ and its destination list includes location $P1'$ and $P2'$.

It is important to notice that FRs commonly respond to the incidents ad hoc, such as fire and obstacles. Being responsive makes deduction of instantaneous terminal topology infeasible. Our model only captures the macroscopic-level movement. Also FRs need to prioritize tasks according to their relative emergency, criticality, feasibility and overhead. The unordered destination list enforces no ordering among the destinations to be visited. In case FRs need to visit some other locations than those specified in the list during mission at time t, we can easily incorporate those locations by rerunning the relay deployment algorithms based on adjusted destination lists after time t and deployed relays by time t.

3.3.2 Two-tier Communication Infrastructure and Persistent BS-connectivity

In this section, we will discuss the design principle for on-demand communication infrastructure and define persistent BS-connectivity in the context of two-tier communication infrastructure.

Terminals within a mobile group can utilize each other as intermediate hops to reach BSs. We have already applied this idea in virtual terminals. However, it is dangerous to plan relays based on the idea that one group can use links formed by other groups. Being responsive, different groups may encounter different incidents and respond to different environment factors, thus speeding up, slowing down or even diverting from their scheduled trajectory. Thus it is extremely difficult to obtain microscopic-level mobility patterns for mobile networks of multiple groups. Unless we have precise future terminal moving trajectory annotated with speed, relay deployment algorithms cannot rely on any instantaneous terminal topology. Hence, we propose two-tier communication infrastructure to decouple mobile groups.

Definition 5. *A **two-tier communication infrastructure (TTCI)** is a connected communication network composed of relays and BSs.*

Under TTCI structure, a virtual terminal only relies on relays to reach BSs and the commanders. Hence, each mobile group is treated as an independent unit. No location correlation information of different groups is needed. In addition, TTCI is self BS-connected. If a mobile group can reach it, the group is BS-connected. We redefine persistent BS-connectivity in the context of TTCI.

Definition 6. *A v-terminal has **persistent BS-connectivity** under TTCI, if there is a communication path of deployed relays towards any BS all the time, during its movement.*

3.3.3 Problem Formulation

Now, let's formulate our constrained relay deployment problems for DDNs. Our objective is to determine the sequence to drop relays at a subset of candidate locations $\mathcal{R} = \{r_1, \cdots, r_p\}$ ($\mathcal{R} \subseteq \mathcal{L}$), while any group G^i has persistent BS-connectivity under TTCI. We call $\mathcal{R}$ a relay deployment solution. If every group G^i visits all the locations in its destination list DL^i, $\mathcal{R}$ is written as $\mathcal{R}_c$, which is called a complete placement solution. We want to find $\mathcal{R}_c^* = \{r_1 \cdots r_{p^*}\}$, to minimize the number of relays p.

Since we deal with relay deployment problems, the communication infrastructure is growing as terminals move around and drop relays here and there. Persistent BS-connectivity only utilizes the deployed relays.

However, a relay deployment solution is the final set of all the deployed relays when all missions are completed. Furthermore, the infrastructure is shared among all groups. A group of FRs can utilize the relays dropped by themselves at an earlier time or by another group.

Chapter 4

Relay Placement to Maintain k-BS-Connectivity in DPNs

DPNs bring great challenges to relay placement problems, mainly due to the fact that relays are utilized by different topologies and thus they must be jointly planned across topologies. On one hand, an optimal relay placement for one topology may not be an optimal placement for another. On the other hand, a global optimal relay placement is usually a suboptimal placement for a topology. How to design good placement algorithms for polymorphous networks is unknown.

Our goal is to deploy the minimum number of relays at a subset of candidate locations, so that for any topology in a DPN, all the terminals have connectivity to BSs, reliable if required. To the best of our knowledge, our work is the first one to study the optimal relay placement problem for DPNs. Also, we identify the load balance problem in presence of multiple BSs. The objective to minimize the total number of relays likely causes extremely unbalanced load on BSs, which contradicts our original intention to place multiple BSs for availability and performance. Often, a majority of terminals connect to one BS, causing serious congestion around this BS while wasting connectivity provided by other BSs. Thus, relay placement needs to carefully balance load and maximize resource utilization among available BSs.

In this chapter, we propose three constrained relay placement algorithms for DPNs. The first two heuristic algorithms are direct applications of state-of-the-art constrained relay placement algorithms for monomorphous networks. *Topology stitch algorithm* solves the placement problem for each topology separately and combines the solutions by pruning the redundant relays. *Topology iterative algorithm* iteratively places relays from one topology to another. Relay placement for later topologies takes into account the relays placed from previous topologies. These two algorithms are nonholistic, in the sense that they all place relays per topology, thus resulting in suboptimal solutions. In order to jointly place relays across topologies, we propose a holistic algorithm, "weigh-and-place" algorithm (WPA), which is developed based on (reliable) multi-commodity flow formulation and linear relaxation with iterative rounding. Relay placement is done per candidate location. In WPA, we place a relay at the most useful place in each iteration, considering the location's contribution to overall connectivity in all topologies.

4.1 Heuristic Algorithms

The first two heuristic algorithms are direct applications of state-of-the-art constrained relay placement algorithms for monomorphous networks. *Topology stitch algorithm* solves the placement problem for each topology separately and combines the solutions by pruning the redundant relays. *Topology iterative algorithm* iteratively places relays from one topology to another. Relay placement for later topologies takes into account the relays placed from previous topologies.

Relay placement problems have been extensively studied for monomorphous networks, which is a special case of DPNs with a single topology. It is attractive and intuitive to utilize well-known algorithms from DMNs in DPNs. In this section, we propose two heuristic algorithms built on top of state-of-the-art constrained relay placement algorithms for DMNs. RPA_m (Relay Placement Algorithm for D**M**Ns) denotes any such algorithm. Both of our proposed algorithms consider relay placement problems per topology, with some variations.

The first heuristic algorithm is the Topology Stitch Algorithm (TSA), sketched in Algorithm 1. First, relays are deployed for each topology individually to assure connectivity requirement. Then we stitch (union) all relays deployed for each topology together and remove redundant ones. Pruning process tries to remove relays one by one. A relay is removed if its removal does not violate the connectivity requirement for every terminal topology. TSA relies on the pruning process to reduce the total number of relays.

Algorithm 1: Topology Stitch Algorithm (TSA)

Input : $\mathcal{TS}, \mathcal{B}, \mathcal{L}$
Output: Placement solution $\mathcal{R}$
$\mathcal{R} = \varnothing$
foreach *Topology* $\mathcal{T}^t$ **do**
    $\mathcal{R} = \mathcal{R} \cup RPA_m(\mathcal{T}^t, \mathcal{B}, \mathcal{L})$;
end
Prune Redundant Relays;

The second heuristic algorithm is the Topology Iterative Algorithm (TIA), sketched in Algorithm 2. Relays are deployed for each topology incrementally. Extra parameter $\mathcal{R}$ is given to RPA_m to take into account the relays placed from topologies considered earlier. At the end, prune procedure is applied to remove any redundant relays, as in Algorithm 1. Note that RPA_m needs modifying so that there is no cost to place relays at locations specified in the input parameter $\mathcal{R}$.

Both TSA and TIA are greedy algorithms. They place relays to maximize the connectivity of the topology being processed. TSA relies on the pruning process to remove redundancy generated by each topology. TIA is somehow smarter than TSA, since it considers the relationship between the current topology and topologies

Algorithm 2: Topology Iterative Algorithm (TIA)

$\mathcal{R} = \varnothing$
foreach *Topology* $\mathcal{T}^t$ **do**
  $\mathcal{R} = \mathcal{R} \cup RPA_m{}^*(\mathcal{T}^t, \mathcal{B}, \mathcal{L}, \mathcal{R})$;
end
Prune Redundant Relays;

processed earlier, albeit only the end results of processed topologies. Connectivity provided by relays already placed in $\mathcal{R}$ is maximally used by the current topology. Nevertheless, TIA's performance highly depends on the processing order of topologies.

4.2 Weigh-and-Place Algorithm

For a single topology, constrained relay placement problems are equivalent to the minimum cost vertex-connectivity problems with unit cost for relay vertexes in $\mathcal{L}$ and zero cost for other vertexes [29]. The placement problem becomes complicated for multiple topologies, since relays and BSs are shared among topologies, but a terminal only belongs to a particular topology. In other topologies, the terminal might move away or be turned off. This situation can be naturally modeled by multi-commodity flow with edge capacity.

We model the relay placement problem for DPNs as a graph $G = (V, E)$, wherein $V = \cup_t V^t$ and $E = \cup_t E^t$. Vertex and edge sets for a terminal topology are defined upon $\mathcal{L}$, as if we were placing a relay at every candidate location in $\mathcal{L}$. $V^t = V^t_{\mathcal{L}}$ and $E^t = E^t_{\mathcal{L}}$. Note that we treat terminals in different topologies as different vertexes. The reason is that though two terminals in two topologies may be the same physical device, they may have diverse neighboring edges. BS connectivity for a terminal in topology $\mathcal{T}^t$ is only provided by V^t and E^t.

We model the BS connectivity by feasibility of multi-commodity flow, i.e., connectivity between C and a terminal exists if we can route some commodities from C to this terminal through the network graph G^t successfully [30]. Recall that $G^t = (V^t, E^t)$. Commodities for terminals in topology $\mathcal{T}^t$ can only be routed in G^t. Essentially, the capacity of edges not belonging to E^t is 0 for commodities originated from G^t. As shown in Figure 4.1, we have two network graphs, dashed blue lines for topology $\mathcal{T}^1$ and bold orange lines for topology $\mathcal{T}^2$. Edges outside E^t cannot route commodities originated from G^t ($t = \{1, 2\}$). We mark the capability of edges along edges. $\{t\}$ means this edge can only route commodities from G^t. Edges between relays and BSs, marked with $\{1, 2\}$, can route any commodities since they belong to both topologies.

There are two benefits from our flow-based formulation. First, we can represent all the terminal topologies

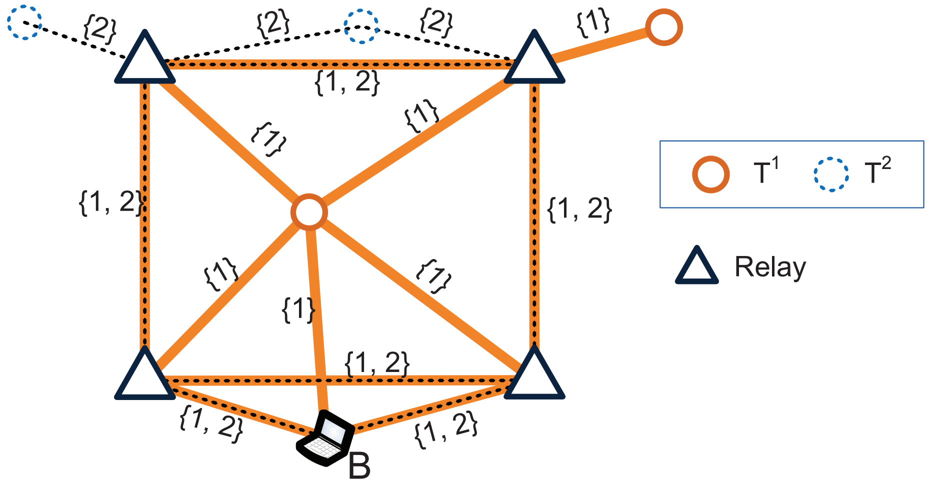


Figure 4.1: Multi-commodity Flow Problem

in one graph, yet without confusing the connectivity requirement per terminal topology. Second, flows are related to loads on vertexes, which can be used for load balancing purpose.

Since we are considering flows, we use directed edges in the following discussion. Two directed edges $\widehat{ij}$ and $\widehat{ji}$ substitute for undirected edges ij in E^t . A variable written in bold is a vector. First, let us look at 1-BS-connectivity for DPNs. Then we will move onto k-BS-connectivity.

4.2.1 Mixed Integer Programming for 1-BS-connectivity

There is a type of commodities associated with each topology, wherein C has a unit of type-t commodity for every terminal in $\mathcal{T}^t$. We need to place relays so that all commodities can flow successfully from C to their destination terminals. In total, C sends $|\mathcal{T}^t|$ units of type-t commodities and every terminal receives one unit for topology $\mathcal{T}^t$. Let d_i^t denote the supply/demand of node i for $\mathcal{T}^t$. If $d_i^t > 0$, node i is a supply node for $\mathcal{T}^t$; if $d_i^t < 0$, node i is a demand node for $\mathcal{T}^t$; if $d_i^t = 0$, node i is a forwarding node for $\mathcal{T}^t$. We have,

$$d_i^t = \begin{cases} |\mathcal{T}^t|, & i = C \\ -1, & i \in \mathcal{T}^t \\ 0, & i \in \{\mathcal{L}, \mathcal{B}\} \end{cases} \tag{4.1}$$

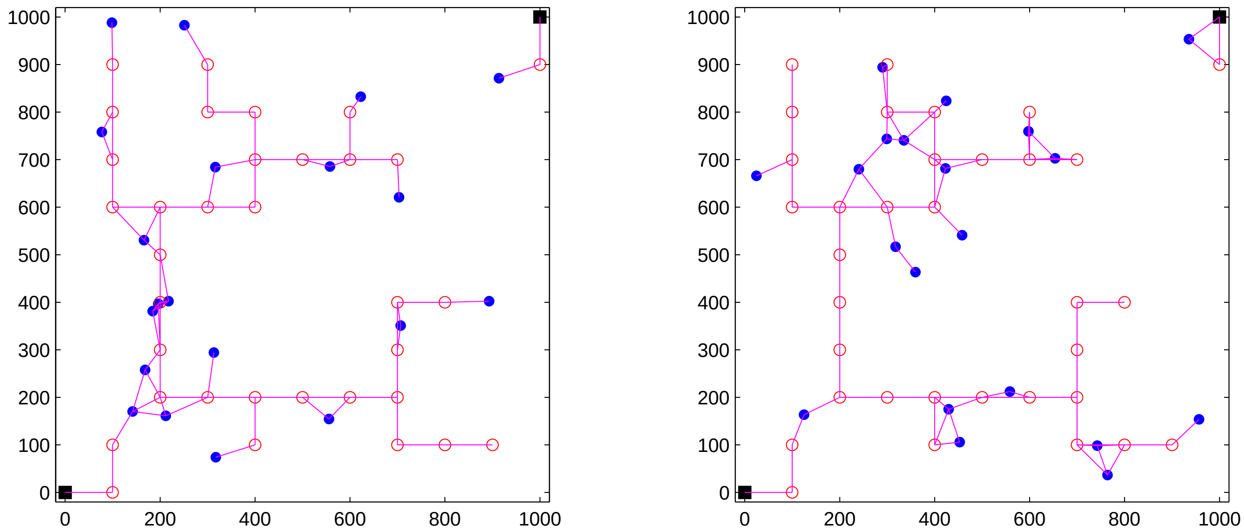
The relay placement problem to maintain 1-BS-connectivity is formulated as follows:

$$
\begin{aligned}
\min \quad & \sum_{l \in \mathcal{L}} y_l && (4.2)\\
\text{s.t.} \quad & \sum_{j:\widehat{ij} \in E^t} x_{ij}^t - \sum_{j:\widehat{ji} \in E^t} x_{ji}^t = d_i^t, \forall i \in V^t && (4.3)\\
& \sum_t \sum_{j:\widehat{lj} \in E^t} x_{lj}^t \leq y_l \sum_t |\mathcal{T}^t|, \forall l \in \mathcal{L} && (4.4)\\
& y_l \in \{0, 1\} && (4.5)\\
& x_{ij}^t \in [0, |\mathcal{T}^t|] && (4.6)
\end{aligned}
$$

The decision variables are binary variables $\mathbf{y}$, as shown in constraint (4.5). If $y_l = 1$, we place a relay in location l; otherwise, no relay is placed in location l. A placement solution is $\mathcal{R} = \{l | y_l = 1, l \in \mathcal{L}\}$. The cost of placing a relay in location l is one. Hence, the objective function in (4.2) corresponds to the total number of relays. When the objective function is minimized, a placement solution with the minimum number of relays is located. Let x_{ij}^t denote the flow of type-t commodities for topology $\mathcal{T}^t$ on directed edge $\widehat{ij}$, and type-t commodities can route through edge $\widehat{ij}$ only if $\widehat{ij} \in E^t$. We refer to the constraints in (4.3) as *mass balance constraints*. The first term for a node i represents the total outflow of the node and the second term represents the total inflow of the node for topology $\mathcal{T}^t$. The mass balance constraints state that the outflow minus inflow must equal the supply/demand of the node for $\mathcal{T}^t$. In constraints (4.4), a relay location l can route any commodity for any topology, only if a relay is placed at l ($y_l = 1$), which we refer to as *relay switch constraints*. The constraints (4.3)-(4.6) together indicate that a placement solution must have a path composed of relays, terminals and BSs between C and every terminal in the network graph G^t for any terminal topology $\mathcal{T}^t$. We denote this optimization by OPT.

A problem with this formulation is that all the terminals are likely to be connected to one BS. This is because reusing placed relays is the most cost-efficient way to minimize the number of relays. This undesirable outcome will overload and congest both the BS and nearby nodes, wasting the connectivity provided by other BSs. As shown in Figure 4.2, the majority of terminals are connected to the BS on the lower left corner, despite two BSs in a DPN of 2 topologies. Only one terminal is connected to the BS on the upper right corner in both topologies. We address this unbalanced load issue by adding the following *stress constraints*.

$$
\sum_{j:\widehat{ij} \in E^t} x_{ij}^t \leq l_i |\mathcal{T}^t|, \forall i \in \mathcal{B} \qquad (4.7)
$$



(a) Topology 1

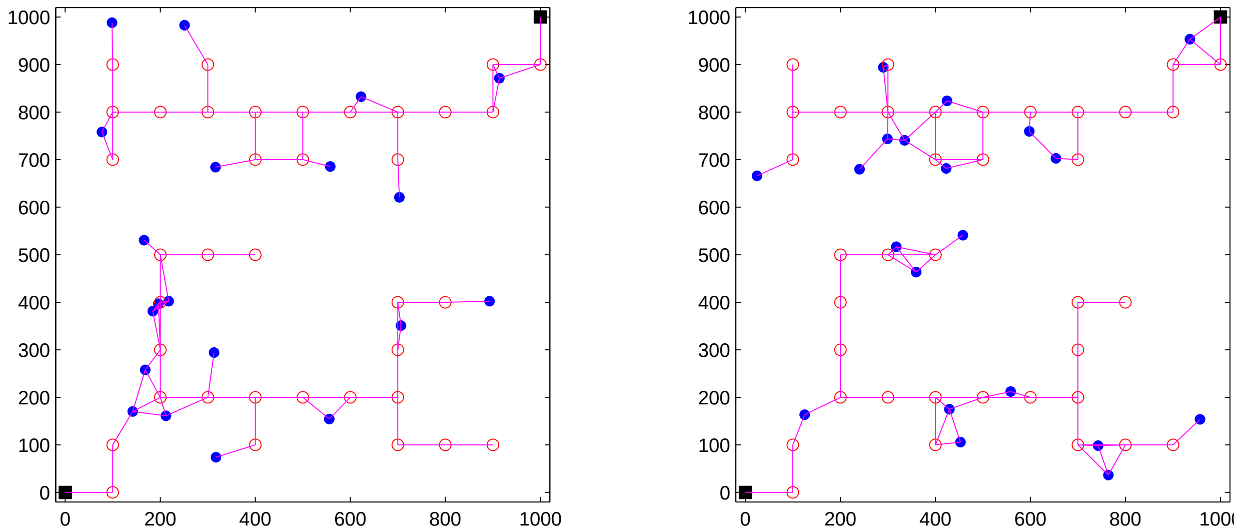
(b) Topology 2

Figure 4.2: Unbalanced Load: network area 1000*1000m; transmission range 100m; 20 terminals in 2 topologies; relay placement scheme, 100m grid; terminal, blue dot; relay, red circle; BS, solid square; one-hop link, solid line.

For $\mathcal{T}^t$, we define the load at a node i as its outflow $(\sum_{j:\widehat{ij}\in E^t} x_{ij}^t)$. Load on node i indicates how many terminals use i as an intermediate hop to reach C in $\mathcal{T}^t$. Low load corresponds to low utilization on the connectivity provided by this node, thus lighter load. We balance load by setting an upper bound $l_i|\mathcal{T}^t|$ for node i's load in $\mathcal{T}^t$, wherein $l_i(0 < l_i \leq 1)$ is called stress. In this way, node i routes traffic for at most $l_i|\mathcal{T}^t|$ terminals to C. In the stress constraints (4.7), we only restrict the load among BSs and their $\mathbf{l}$ values should be at least $1/|\mathcal{B}|$. If loads need to be restricted at other devices, we can include them as well. $\mathbf{l}$ should be carefully configured to avoid infeasible relay placement solution. Shown in Figure 4.3(a) and Figure 4.3(b), the resulting relay placement with stress constraints enables almost an even number of terminals connected to both BSs, while only incurring 3 additional relays ($l_i = 0.6$). Often, we expect the number of relays with load balance should be higher than the number without load balance.

4.2.2 Weigh-and-Place Algorithm for 1-BS-connectivity

The integral requirement on $\mathbf{y}$ prevents us from using standard linear programming (LP) techniques. With the help from branch-and-bound approach [31], we can find an exact or almost optimal solution; however, it takes an exponential number of iterations. This becomes a serious problem when we have a large number of candidate locations and terminals with many topologies. Linear programming relaxation with iterative



(a) Topology 1 (b) Topology 2

Figure 4.3: Balanced Load with Stress Constraints

rounding is often used to tackle the integral requirement [32]. We will adopt this method for optimization problem OPT.

The general idea of linear relaxation with iterative rounding (LR-IR) for binary variables is solving the LP relaxation of an integer programming, fixing a variable, which is closest to 1, to 1 in each iteration and resolving the LP until all the variables are either 1 or 0, thus forming a feasible solution. LR-IR method, when applied to relay placement problems, works as follows. Solve the LP relaxation ($y_l \geq 0$ instead of $y_l \in \{0, 1\}$ in OPT) to find an optimal solution $\mathbf{y}^*$. Pick a location l^* with highest weight y_l^* ($l^* \notin \mathcal{R}$) and add it to the placement solution $\mathcal{R}$ (initially, $\mathcal{R}$ is empty). This step essentially rounds up the variable y_{l^*}, which is closest to 1, to 1. Then resolve the LP, assuming that relays are placed in $\mathcal{R}$ (i.e., $y_l = 1, \forall l \in \mathcal{R}$). Iteratively solve the problem until all the locations outside $\mathcal{R}$ have zero weight ($y_l = 0, \forall l \in \mathcal{L} \backslash \mathcal{R}$), and then we obtain a feasible solution to the integer optimization problem. Based on this idea, we propose *Weigh-and-Place* algorithm (WPA), which adds relays iteratively, considering their weight obtained from linear programming relaxation. We illustrate the main steps of WPA in Figure 4.4.

From relay switch constraints (4.4), we know that

$$\sum_{l \in \mathcal{L}} y_l \geq \sum_{l \in \mathcal{L}} \frac{1}{\sum_t |\mathcal{T}^t|} \sum_t \sum_{j:\widehat{lj} \in E^t} x_{lj}^t \tag{4.8}$$

$$= \frac{1}{\sum_t |\mathcal{T}^t|} \sum_t \sum_{l \in \mathcal{L}} \sum_{j:\widehat{lj} \in E^t} x_{lj}^t \tag{4.9}$$

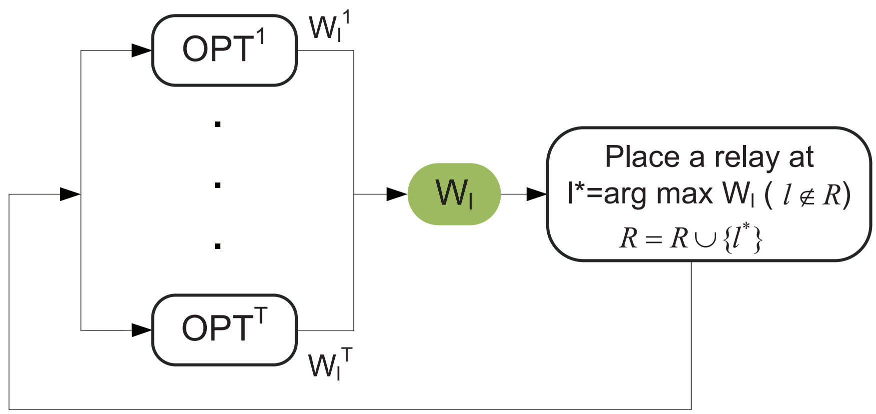


Figure 4.4: Overview of Weigh-and-Place Algorithm

Relaxing the integral requirement on $\mathbf{y}$, the objective function now becomes

$$\begin{aligned} \min \sum_{l \in \mathcal{L}} y_l \quad &\Longleftrightarrow \quad \min \frac{\sum_t \sum_{l \in \mathcal{L}} \sum_{j:\widehat{lj} \in E^t} x_{lj}^t}{\sum_t |\mathcal{T}^t|} \\ &\Longleftrightarrow \quad \frac{1}{\sum_t |\mathcal{T}^t|} \sum_t \min \sum_{l \in \mathcal{L}} \sum_{j:\widehat{lj} \in E^t} x_{lj}^t \end{aligned} \tag{4.10}$$

The global optimization across multiple topologies decomposes into T separate suboptimization problems: min cost flow problem in variables x_{ij}^t with balanced load for each individual topology $\mathcal{T}^t$, which we refer to as OPT^t.

$$\min \sum_{l \in \mathcal{L}} \sum_{j:\widehat{lj} \in E^t} x_{lj}^t \tag{4.11}$$

$$\text{s.t.} \quad \sum_{j:\widehat{ij} \in E^t} x_{ij}^t - \sum_{j:\widehat{ji} \in E^t} x_{ji}^t = d_i^t, \forall i \in V^t \tag{4.12}$$

$$\sum_{j:\widehat{ij} \in E} x_{ij}^t \leq l_i |\mathcal{T}^t|, \forall i \in \mathcal{B} \tag{4.13}$$

$$x_{ij}^t \in [0, |\mathcal{T}^t|] \tag{4.14}$$

By this decomposition, we can solve OPT^t for each topology independently. The weight w_l of candidate location l is calculated as follows:

$$w_l^t = \sum_{j:\widehat{lj} \in E^t} x_{lj}^t \tag{4.15}$$

$$w_l = \sum_t w_l^t \tag{4.16}$$

We sketch WPA in Algorithm 3 and 4. Algorithm 3 is the iterative relay placement (IRP) algorithm based on linear relaxation with iterative rounding. $\mathcal{R}$ is the relay placement solution calculated iteratively. Initially, $\mathcal{R}$ is empty. We solve min cost flow problem OPT^t for each topology, considering the cost only on the out-edges of locations outside $\mathcal{R}$. Thus we adjust the objective function as $\sum_{l\in\mathcal{L}\setminus\mathcal{R}}\sum_{j:\widehat{lj}\in E^t} x_{lj}^t$. Let f^t be the optimal objective value obtained from OPT^t. If all commodities can flow along placed relays $\mathcal{R}$, i.e., $\sum_t f^t = 0$, we return $\mathcal{R}$ as the solution. Otherwise, we pick the location l^* with highest weight and add it to the placement solution $\mathcal{R}$. IRP is guaranteed to terminate since a location is added upon each call and there are $|\mathcal{L}|$ candidate locations to consider.

Algorithm 3: Iterative Relay Placement Algorithm (IRP)

Input : Relay placement solution $\mathcal{R}$
Output: Relay placement solution $\mathcal{R}$
foreach *Topology* $\mathcal{T}^t$ **do**
  Solve OPT^t with objective function$\sum_{l\in\mathcal{L}\setminus\mathcal{R}}\sum_{j:\widehat{lj}\in E^t} x_{lj}^t$. Let the optimal solution be $(\mathbf{x}^*)$ and optimal objective value f^t.
  $w_l^t = \sum_{j:\widehat{lj}\in E^t} x_{lj}^{t*}$
end
if $\sum_t f^t = 0$ **then**
  return $\mathcal{R}$;
end
$w_l = \sum_t w_l^t$; $l^* = arg_{l\in\mathcal{L}\setminus\mathcal{R}} \max w_l$.
return $\mathcal{R}$=IRP($\mathcal{R} \cup \{l^*\}$)

Being greedy at each iteration, IRP may place redundant relays, loops for example. WPA in Algorithm 4 applies pruning process after IRP algorithm, which tries to remove placed relays one by one without violating connectivity and load balance requirements for every terminal topology.

Algorithm 4: Weigh-and-Place Algorithm (WPA)

Input : $\mathcal{TS}, \mathcal{B}, \mathcal{L}$
Output: Relay placement solution $\mathcal{R}$
$\mathcal{R}$ = IRP($\emptyset$);
// Prune Redundant Relays
foreach $r \in \mathcal{R}$ **do**
  $\mathcal{R}' = \mathcal{R} - r$;
  if *For any* $\mathcal{T}^t$*, all terminals are 1-BS-connected and BSs' loads are balanced in* $\mathcal{T}^t \cup \mathcal{B} \cup \mathcal{R}' \cup \{C\}$ **then**
    $\mathcal{R} = \mathcal{R}'$;
  end
end

4.2.3 Fault Tolerance with k-BS-connectivity

With 1-BS-connectivity, every terminal has a communication path to the monitoring center C. It is desirable to have k-BS-connectivity to retain connectivity with C if at most $k-1$ terminals or relays fail. A terminal is k-BS-connected if it has k (terminal, relay)-disjoint communication paths to C.

Again, we transform the vertex connectivity problem to multi-commodity network flow problem. However, we associate a type of commodity with each terminal in every topology. C desires to route k units of type-(ts) commodities to terminal s in terminal topology $\mathcal{T}^t$, wherein k is the connectivity requirement. The supply/demand d_i^{ts} at a node i for type-(ts) commodities is:

$$d_i^{ts} = \begin{cases} k, & i = C \\ -k, & i = s \in \mathcal{T}^t \\ 0, & i \in \{\mathcal{L}, \mathcal{B}\} \end{cases} \tag{4.17}$$

x_{ij}^{ts} denotes the flow of type-(ts) commodities on directed edge $\widehat{ij}$. Mass balance constraints in (4.19) are specified per type of commodities. Relay switch constraints in (4.20) state that a candidate location l can route any commodities for any terminal in any topology, only if a relay is placed there. Capacity constraints in (4.21) guarantee that a relay or terminal can only send out one unit of type-(ts) commodity. In this way, we enforce that k units of type-(ts) commodities for terminal s are routed through k (terminal, relay)-disjoint paths in G^t. Like OPT, stress constraints in (4.22) are enforced at all BSs. We refer to this optimization for k-BS-connectivity as OPT_k.

$$\min \quad \sum_{l \in \mathcal{L}} y_l \tag{4.18}$$

$$\text{s.t.} \quad \sum_{j:\widehat{ij} \in E^t} x_{ij}^{ts} - \sum_{j:\widehat{ji} \in E^t} x_{ji}^{ts} = d_i^{ts}, \forall i \in V^t \tag{4.19}$$

$$\sum_{t, s \in \mathcal{T}^t} \sum_{j:\widehat{lj} \in E^t} x_{lj}^{ts} \leq y_l k \sum_t |\mathcal{T}^t|, \forall l \in \mathcal{L} \tag{4.20}$$

$$\sum_{j:\widehat{ij} \in E^t} x_{ij}^{ts} \leq 1, \forall i \in \{\mathcal{T}^t, \mathcal{L}\} \tag{4.21}$$

$$\sum_{s, j:\widehat{ij} \in E^t} x_{ij}^{ts} \leq l_i |\mathcal{T}^t| k, \forall i \in \mathcal{B} \tag{4.22}$$

$$y_l \in \{0, 1\} \tag{4.23}$$

$$x_{ij}^{ts} \in [0, 1] \tag{4.24}$$

By linear relaxation on integral requirement on $\mathbf{y}$, OPT_k decomposes into T suboptimization problems OPT_k^t, which is a min cost flow problem with load balance and vertex connectivity for topology $\mathcal{T}^t$. We still use WPA to solve k-BS-connectivity problem. In IRP, OPT^t is replaced by OPT_k^t with adjusted objective function $\sum_{s\in\mathcal{T}^t}\sum_{l\in\mathcal{L}\setminus\mathcal{R}}\sum_{j:\widehat{lj}\in E^t} x_{lj}^{ts}$. Correspondingly, $w_l^t = \sum_{s\in\mathcal{T}^t}\sum_{j:\widehat{lj}\in E^t} x_{lj}^{ts}$. In the pruning step, a relay r is removed if its removal does not violate k-BS-connectivity and load balance requirements for any terminal topology. Suboptimization problem OPT_k^t for topology $\mathcal{T}^t$ is

$$\min \quad \sum_{s\in\mathcal{T}^t}\sum_{l\in\mathcal{L}}\sum_{j:\widehat{lj}\in E^t} x_{lj}^{ts} \tag{4.25}$$

$$\text{s.t.} \quad \sum_{j:\widehat{ij}\in E^t} x_{ij}^{ts} - \sum_{j:\widehat{ji}\in E^t} x_{ji}^{ts} = d_i^{ts}, \forall i \in V^t \tag{4.26}$$

$$\sum_{j:\widehat{ij}\in E^t} x_{ij}^{ts} \le 1, \forall i \in \{\mathcal{T}^t, \mathcal{L}\} \tag{4.27}$$

$$\sum_{s,j:\widehat{ij}\in E^t} x_{ij}^{ts} \le l_i|\mathcal{T}^t|k, \forall i \in \mathcal{B} \tag{4.28}$$

$$x_{ij}^{ts} \in [0,1] \tag{4.29}$$

4.3 Evaluation

In this section, we evaluate the performance of three proposed algorithms by measuring the number of relays deployed, load on base stations under various parameters, such as transmission range, the number of topologies, network area, terminal density and reliability requirement. The performance is averaged over 10 DPN instances randomly generated from a DPN configuration. A *DPN configuration* is defined by the network area, the number of terminals, the number of terminal topologies and the locations for base stations (we assume that each topology contains the same number of terminals). We generate a *DPN instance* by placing the desired number of terminals randomly in the network area for each topology. For all DPN configurations considered, two BSs are placed, one at the top-right corner and one at the bottom-left corner. For simplicity, transmission range of terminals is 100m. We vary the transmission range for BSs and relays, which is 100m by default. The relay placement scheme is a regular grid with 100m distance between adjacent locations.

In presence of multiple BSs, it is best to evenly utilize the connectivity provided by all BSs, so that no base stations are much more congested than others. Here, we use "skewness" to measure how well loads on various BSs are balanced. The *load* for base station B in a terminal topology is the number of terminals connected to B. If a terminal t has connectivity to b BSs, each of the b BSs is assigned $\frac{1}{b}$ terminal from t. *Skewness for a terminal topology* is the ratio of the maximum load and the minimum load among all BSs. *Skewness for a DPN instance* is averaged over all topologies. In case there is no terminal connected to a particular BS, the minimum load for this topology is 0. To avoid infinite skewness value, we set the number of terminals as the upper bound for the skewness value of a topology. A large skewness value corresponds to uneven loads on various BSs.

Approximation algorithm for RNPc in [14] is used as the RPA_m implementation in TSA and TIA for 1-BS-connectivity. With respect to k-BS-connectivity, we cannot directly apply approximation algorithm for RNPs in [14], since it only support 2-BS-connectivity. Based on the ideas from [14], we develop our own RPA_m implementation. First we use edge weight to approximate node weight for placing a relay. Then we apply single-source vertex-connectivity algorithm with edge cost in [33] to find the locations to place relays.

4.3.1 1-BS-connectivity

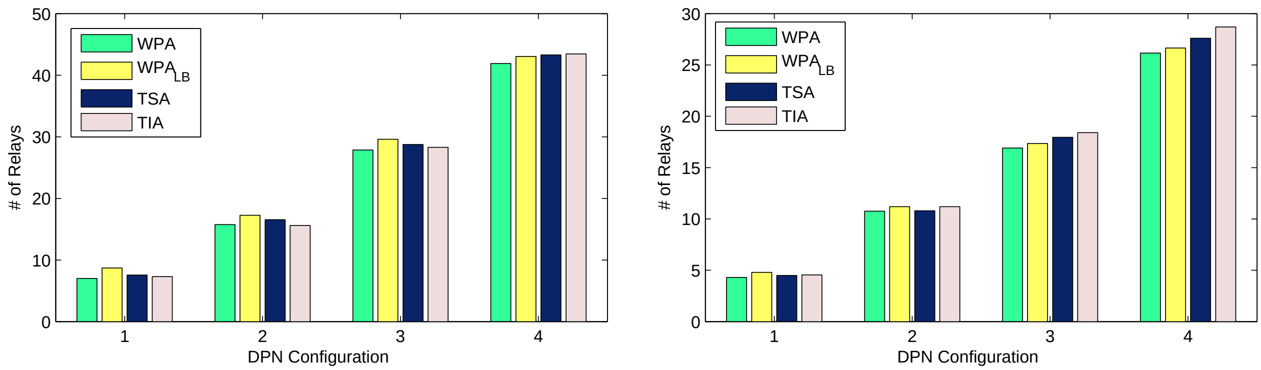
First, let us look at the performance of TSA, TIA and WPA to maintain 1-BS-connectivity. In order to show the extra overhead introduced by stress constraints, we plot the performance of WPA in two cases, one with load balance (WPA_{LB}) and one without load balance (WPA).

The first set of experiments is to examine the number of relays deployed and skewness for different transmission ranges and network sizes, as shown from Figure 4.5(a) to Figure 4.6(b). Two transmission ranges for relays and BSs are evaluated, 100m and 200m. There are four DPN configurations with the same terminal density shown in Table 4.1. All DPN configurations contain 3 topologies. For load balance, l is set to be 0.6 for all BSs to avoid excessive overhead while still maintaining good level of load balancing. In Figure 4.5(a) and Figure 4.5(b), WPA_{LB} places an extra number of relays, yet small, compared with WPA without load balance. For all the network sizes and transmission ranges examined, WPA is constantly better than TSA and TIA, in average sense. The improvement of WPA is more visible when network area is large

Table 4.1: DPN Configuration

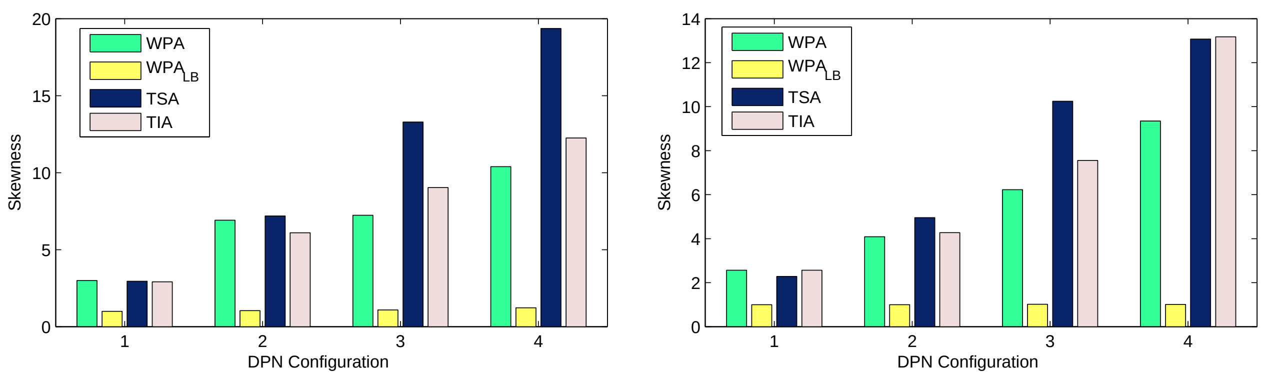
Index	Network Area	Number of Terminals Per Topology
1	400*400m	4
2	600m*600m	9
3	800m*800m	16
4	1000m*1000m	25

and the transmission range is long, DPN configuration 4 with transmission range 200m for example. One interesting thing to notice is that TSA and TIA have reversed trends in two setups. With 100m transmission range, TIA is generally better than TSA; on the contrary, with 200m transmission range, TSA is better than TIA. Possible explanation is that with 200m transmission range, candidate relay locations are dense, and the best relay placement solutions for each topology overlap together. Thus, TSA is able to combine those best placements and prune the redundant relays. For load balance shown in Figure 4.6(a) and Figure 4.6(b), WPA_{LB} always has skewness value around 1, which means two base stations are attached with almost an equal number of terminals. Skewness for WPA without load balance is large; yet it is often smaller than skewness in TSA and TIA, or at least comparable.



(a) txRange=100m (b) txRange=200m

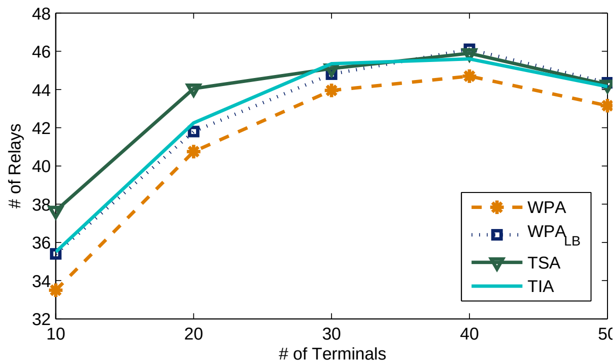
Figure 4.5: Number of Relays



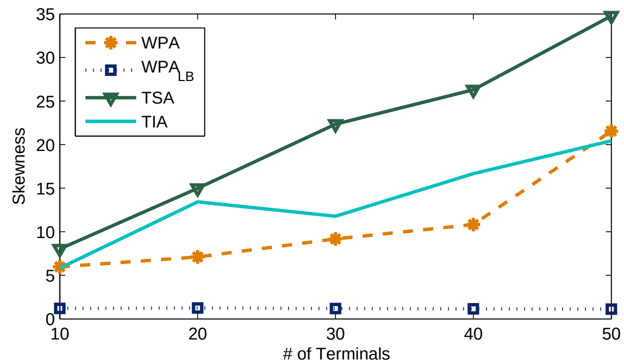
(a) txRange=100m (b) txRange=200m

Figure 4.6: Skewness

In the second set of experiments, we examine the number of relays and skewness under different network densities. The plots are shown in Figure 4.7(a) and Figure 4.7(b). We fix the network area at size 1000m*1000m. Transmission range for relays and BSs is 100m. When the number of terminals increases, the number of placed relays first increases and then decreases. The decreasing trend is caused by big connected components formed by terminals themselves when there are many terminals. WPA again has the best performance because of joint relay placement across multiple topologies. WPA$_{LB}$ incurs an additional number of relays placed, however the skewness is maintained at 1 as shown in Figure 4.7(b).



(a) Number of Relays



(b) Skewness

Figure 4.7: Varying Number of Terminals

Finally, we study the number of relays with a varying number of topologies from 1 to 5 in Figure 4.8. 40 terminals per topology are deployed in a 1000m*1000m network. For a single topology, the performance of flow-based formulation with linear relaxation is close to the performance of constrained relay placement algorithms in [14] for DMNs. However, our formulation is able to deal with load balancing among BSs. When the number of topologies increases, the advantage of WPA over TSA and TIA becomes clearer because of its optimization across topologies.

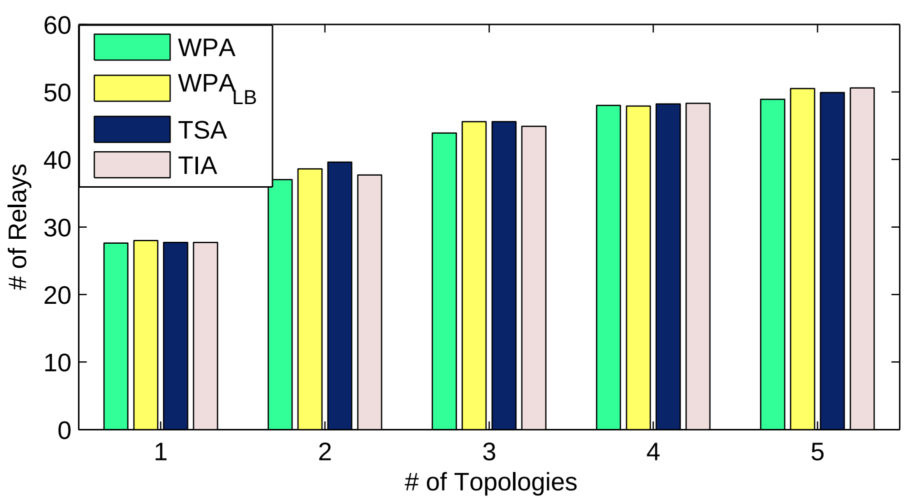
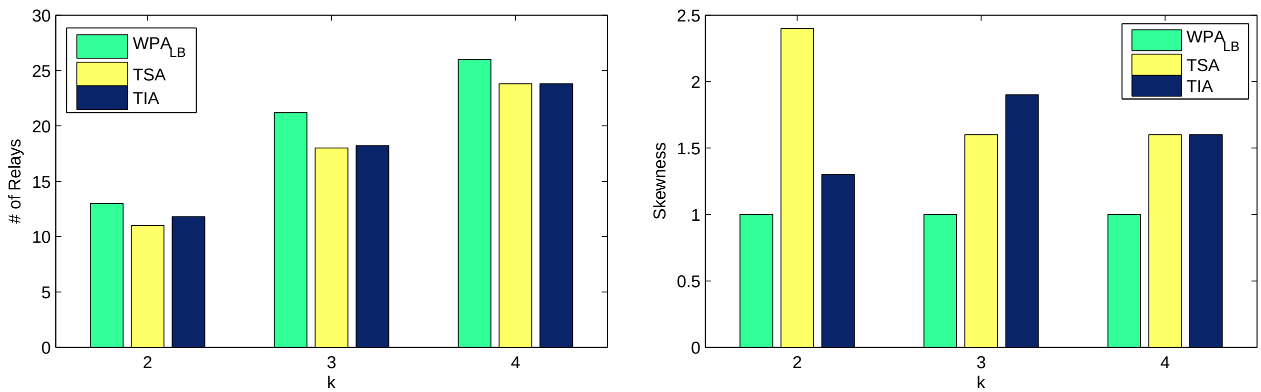


Figure 4.8: Number of Relays: Varying Number of Topologies

4.3.2 k-BS-connectivity

In this section, we study the performance for k-BS-connectivity in 400m*400m networks with 4 terminals per topology. Each DPN contains two topologies. In order to satisfy the requirement for k (terminal, relay)-disjoint paths between BSs and terminals, we increase the grid density by shrinking the distance between adjacent locations from 100m to 50m. As plotted in Figure 4.9(a), high reliability with large k demands more relays. WPA_{LB} places a few more relays than TSA and TIA; however it is able to evenly balance load among 2 BSs, shown in Figure 4.9(b).



(a) Number of Relays

(b) Skewness

Figure 4.9: Varying k

Chapter 5

Energy-aware Two-Tier Relay Placement with Terminal Clusters

In previous chapter, we discuss the relay placement problems to maintain k-BS-connectivity for DPNs. Relays are planned in strategical locations to maximally utilize the connectivity already formed among terminals. However, the resulting communication infrastructure may suffer from two problems: (1) connectivity instability caused by terminal movement and fluctuation of network connectivity at remote area; (2) short network lifetime due to unbalanced energy consumption. We address these problems by two mechanisms. First, we introduce two-tier communication infrastructure with terminal clusters to improve robustness, while preserving the connectivity present inside predefined terminal clusters. Second, we propose energy-aware relay placement algorithms to meet network lifetime constraints.

The relay placement algorithms in the previous chapter construct a *one-tier* communication infrastructure, wherein terminals and relays have an equivalent role regarding forwarding packet on behalf of other nodes. However, one-tier networks suffer from connectivity instability brought by terminal movement and fluctuation of network connectivity at remote region. Figure 5.1 depicts a DPN instance with fragile BS-connectivity. In this example, a single topology $\mathcal{T}^1$ contains 3 terminals, terminal 1 to 3. Terminal 1 relies on the links jointly provided by relay a and b and terminal 2 and 3 to reach the BS. Terminal 1 is likely to be disconnected from the BS, whenever the link between relay a and terminal 2 is broken. This link is broken if any of the following cases happens: (1) terminal 2 moves out of the transmission range of relay a; (2) terminal 2 denies to forward packets from terminal 1 because its remaining battery falls below some threshold; and (3) terminal 2 fails or malfunctions.

In order to reduce the connectivity instability, we introduce two-tier communication infrastructures with terminal clusters.

Definition 7. *A terminal cluster (TC) is a subset of terminals forming a connected component, which forward packets for each other, but not for outside nodes.*

Definition 8. *A two-tier communication infrastructure with terminal clusters (TTCI-TC) is a communication infrastructure with all the relays being connected to BSs, while terminals are organized as a set of terminal clusters.*

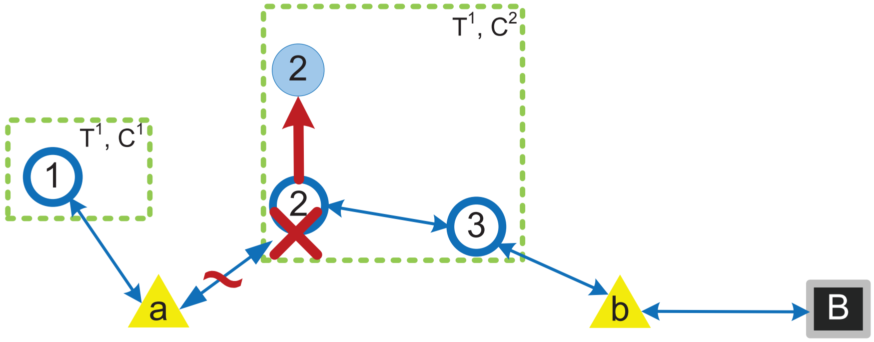


Figure 5.1: Connectivity Instability Brought by Link Failure

From the definition, relays form a backbone network, providing connectivity to BSs. Packets originated from or designated to a terminal in a cluster can only be forwarded by relays and the terminals from the same cluster in a TTCI-TC. As long as data packets originated from a terminal hit the relay backbone, they cannot be forwarded by a terminal anymore. Once data packets from the command center reach a terminal cluster, they can be only forwarded within the cluster till the destination.

We propose the concept of terminal clusters, not only for technical purpose, but also for its realistic reflection of FR operations. A terminal cluster often corresponds to a logical group of first responders or devices with the common missions or tasks. One example is a set of sensors collaboratively capturing the video feed from different angles near an animal hideout. Another example is a group of first responders from the same organization, who cooperate in the common mission and protect each other, always staying in close proximity.

TTCIs-TC reduce connectivity instability by keeping terminals of different clusters from interfering with each other. More specifically, since a cluster only relies on relays and itself to forward traffic, a broken link from other clusters will not disrupt its traffic flow. Nevertheless, it adds only a little relay placement overhead, for existing connectivity formed within terminal clusters is exploited as much as possible. TTCIs-TC may be viewed as a superset of traditionally defined two-tier networks (TTNs), where only relays forward packets for other nodes. In other words, in TTNs, all terminals must be covered by at least a relay, if they require BS-connectivity, and relays have BS-connectivity. Contrary to TTNs where all the terminals work independently from each other, TTCIs-TC only requires terminals from different clusters to work independently. We show a relay placement scheme for a TTCI-TC in Figure 5.2.

In TTCIs-TC, it is unnecessary to cover every terminal by the relay backbone. Hence the number of relays is effectively reduced. It is important to mention that terminal clusters are often capable of self-recovery. Even though some critical link inside a terminal cluster is broken, terminals are able to recover the connectivity by moving around, or via side channels such as oral communication in emergency response

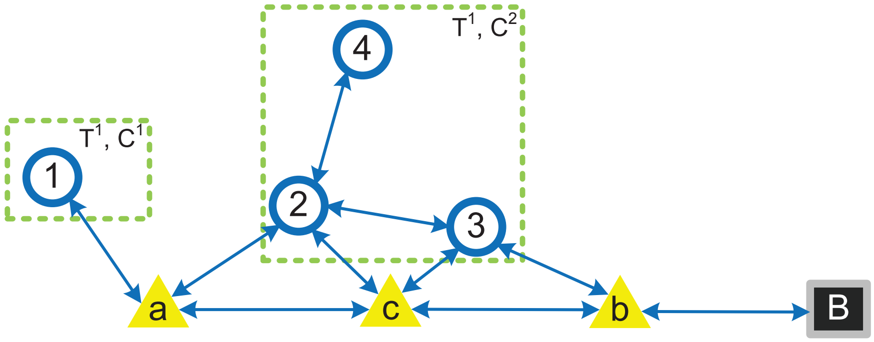


Figure 5.2: Two-Tier Communication Infrastructure with Terminal Clusters

case and other radio technology in environment sensing case. Terminal clusters are a good choice to enforce independence with self-recovery capability, while introducing a little overhead. Besides, relays are usually dedicated forwarding devices with a large transmission range and high battery capacity, designed to be durable, sustainable and robust. With a relay backbone, the whole communication network is more robust to environment incidents, such as fire and flood, and may show better performance by excluding the devices with either a small transmission range or limited battery capability in the backbone.

Besides BS-connectivity stability issue, another big concern is energy consumption. Terminals and relays are commonly equipped with limited and irreplaceable power supplies. During network operation, it is unrealistic to replace the depleted batteries, especially when those devices are scattered around, performing critical monitoring and communication functions. Among all the services, radio communication consumes most of the energy. Relay nodes near the BSs are usually the first to die out of battery because they reside in many communication paths between terminals and BSs. Once their energy is drained out, the network becomes partitioned and BS-connectivity no longer exists. This unbalanced load on various nodes can be resolved by adding the energy constraints into the optimization formulation. A common metric used to approach energy constraints is network lifetime, which is the time the network keeps functioning until one node drains its energy. Unlike traditional sensor applications, ad hoc monitoring and control applications don't need to extend network lifetime as long as possible. The network is only established for a planed time period, one day or one week for example. It dissolves after the public safety activity is finished. Hence, we are interested in network functioning for a given network lifetime, rather than the maximal network lifetime.

There are several factors affecting a node's energy consumption: routing scheme, power control and data rate control. Relay placement must be jointly scheduled with these three factors in order to minimize the total number of relays, while meeting the expected network lifetime requirement. Intelligent routing tries to balance the load among relays and terminals so that no nodes are drained much faster than other nodes. Power control dynamically adjusts the transmission power just enough to reach neighbors with acceptable received signal quality. Data rate control allows link rate adaptation for different neighbors so that packets are delivered with the highest throughput. Power control and data rate control are usually complementary to each other. Hence, we will consider either power control with fixed data rate or data rate control with fixed power, but not both simultaneously.

Formally speaking, we are interested in the following energy-aware relay placement problems for DPNs: *given a disconnected polymorphous network instance, energy provision at each node and traffic rate, how to place the minimum number of relays in a two-tier communication infrastructure with terminal clusters, so that for any topology in the DPN, all the terminals have BS-connectivity, while the expected network lifetime requirement is met?* For simplicity, we will limit our discussion to upload traffic from terminals to CC.

In order to approach the above problem, we need to answer the following problems:

- How to model two-tier communication infrastructures with terminal clusters;
- How to solve the joint scheduling problems of relay placement, routing, and power or data rate control;
- How to model power consumption in DPNs, under either power control or data rate control.

5.1 Energy Consumption Model

We assume that the wireless network is not overloaded, thus without congestion. We also ignore the impact of interference or collision on power consumption and effective link rate. Wireless communication uses standard 802.11 MAC protocol, with RTS/CTS disabled. All the data traffic are sent using the uniform packet size S_{data} and acknowledged with ACK packets of size S_{ack}. Data packets are sent at data rate r_{data} if fixed data rate is used and ACKs are always sent at basic rate r_{basic}, which is the lowest modulation rate for the adopted protocol standard.

5.1.1 Power Control

Let's first discuss the case when nodes can control their transmission power for each data packet, sent at fixed data rate r_{data}. A node can be in one of the three modes: transmission, reception and idleness. Only transmission and reception modes consume energy, which is modeled by $E_s(d)$ and E_r, representing the energy consumption rate to send traffic to a neighbor at distance d and to receive traffic respectively. Our formulation is not bound to specific energy models for wireless communication. For the discussion purpose, the following energy consumption model is used as [10].

$$E_s(d) = a * d^\alpha + b$$
$$E_r = b$$

$E_s(d)$ represents the energy consumption rate to transmit packets from a node to another node d meters apart. And E_r represents the energy consumption rate to receive packets, no matter how far the sender and the receiver are. In this model, a is determined by the transmitter amplifier's efficiency and the channel condition. b represents the energy consumption to run the transmitter or receiver circuit. α is the path loss exponent.

For the majority of off-the-shelf wireless adaptors, not only is energy consumed when the devices transmit packets and receive intended packets (either unicast or broadcast), energy is also consumed when the devices overhear packets (either correct or corrupted) designated for other nodes, if the packets' received signal is above the carrier sense threshold $CSThres$. For discussion purpose, we will use Friis propagation model shown in Equation (5.1). Formally, $P^r_{ij}(P_t)$ denotes the received packet signal from node i to j, when i uses the transmit power P_t. $freq$ is the transmission frequency. G_t and G_r are the transmit and receive antenna gains and L is the system loss. $|ij|$ is the distance between node i and node j.

$$P^r_{ij}(P_t) = \frac{P^t G_t G_r (3.0 * 10^8/freq)^2}{(4\pi|ij|)^2 L} \quad (5.1)$$

A packet is sensed, if

$$P^r \geq CSThres \quad (5.2)$$

. And, a packet is correctly received, only if

$$\frac{P^r}{noise + interference} \geq SINR \quad (5.3)$$

, where SINR is the signal-to-noise-and-interference ratio. Usually the carrier sense range defined by $CSThres$ is much larger than the transmission range defined by $SINR$. We will capture the energy consumption on carrier sensed packets designated for other nodes by eavesdropping energy cost.

5.1.2 Data Rate Control

In case of adaptive data rate control with fixed power, the sending and receiving energy consumption rates are constant. We denote them by E_s and E_r respectively. With different data rates, the receiver may experience diverse received signal qualities, thus diverse packet error rates, due to different modulation schemes and distance factors.

5.2 Energy-aware Relay Placement

We model the energy-aware relay placement problem by multi-commodity flow. One type of commodities is associated with each topology. A terminal u in topology $\mathcal{T}^t$ needs to send type-t commodities to the command center at the rate γ_u^t. Hence C receives type-t commodities at the rate $\sum_{u\in\mathcal{T}^t}\gamma_u^t$. d_i^t denotes the supply/demand of node i in $\mathcal{T}^t$. Relays neither supply nor demand anything; hence their d values are 0. We have

$$d_i^t = \begin{cases} \sum_{u\in\mathcal{T}^t}\gamma_u^t, & i = C \\ -\gamma_i^t, & i \in \mathcal{T}^t \\ 0, & i \in \mathcal{L}\cup\mathcal{B} \end{cases} \tag{5.4}$$

Next, we will discuss how to model two-tier communication infrastructures with terminal clusters via multi-commodity flow. Packets can flow freely within a terminal cluster. However, once a packet reaches the relay backbone, it cannot be routed by any terminals. This requirement can be naturally modeled by one-way directed edges from terminals to relays, as shown in Figure 5.3, which are marked with dotted red lines. All the other edges are two-way directed edges. There are no edges between terminals from different terminal clusters, even within a single topology, to enforce cluster independency. All the edge are defined based on the maximal transmission range, either with or without power control. Whether to use power control or data rate control only affects energy consumption rate.

5.2.1 General Framework - Mixed Integer Programming

The energy-aware relay placement problem is formulated as follows, which we call OPT_{relay}.

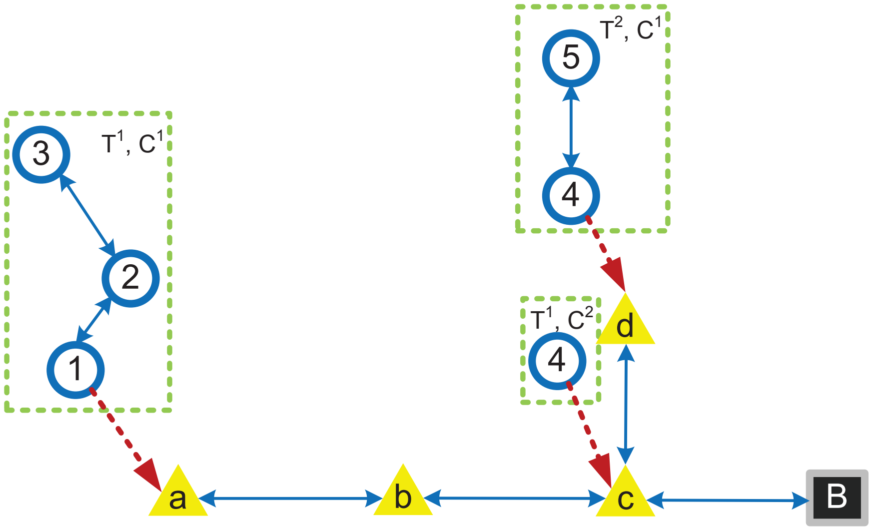


Figure 5.3: One-way Edge from Terminals to Relays

$$
\begin{aligned}
\min \quad & \sum_{l \in \mathcal{L}} y_l && (5.5)\\
\text{s.t.} \quad & \sum_{j:\widehat{ji} \in E^t} x^t_{ji} - \sum_{j:\widehat{ij} \in E^t} x^t_{ij} = d^t_i && (5.6)\\
& \sum_t \sum_{j:\widehat{lj} \in E^t} x^t_{lj} \leq y_l \sum_t \sum_{i \in \mathcal{T}^t} \gamma^t_i, \forall l \in \mathcal{L} && (5.7)\\
& \sum_{t:i \in V^t} (E^t_{in}(i) + E^t_{out}(i) + E^t_{eav}(i)) f_t L \leq E_i && (5.8)\\
& y_l \in \{0, 1\} && (5.9)\\
& x^t_{ij} > 0 && (5.10)
\end{aligned}
$$

The decision variables are binary variables y_ls and real-value variables x^t_{ij}s. A placement solution is $\mathcal{R} = \{l | y_l = 1, l \in \mathcal{L}\}$. The objective function (5.5) tries to minimize the total number of placed relays. Constraint (5.6) is the mass balance constraint, stating that the inbound flow minus outbound flow should equal to the supply/demand of the node for each topology $\mathcal{T}^t$. Constraint (5.7) is the relay switch constraint. It requires that a relay must be placed ($y_l = 1$) in order for this relay place l to route any traffic, up to the total traffic rate generated in a topology. The most interesting part of this formulation is constraint (5.8) for energy consumption. We decompose the mean energy consumption rate at node i in topology $\mathcal{T}^t$ into three parts:

$E^t_{out}(i)$ **-** mean energy consumption rate by outbound traffic;

$E^t_{in}(i)$ **-** mean energy consumption rate by inbound traffic;

$E^t_{eav}(i)$ **-** mean energy consumption rate by overheard traffic.

When multiplying the summation of all three mean energy consumption rates at node i with the active time $f_t L$ for topology $\mathcal{T}^t$, we get the total energy consumed by topology $\mathcal{T}^t$ node i belongs to. The summation of the energy consumption at all the containing topologies should not exceed the initial battery provision at node i when the network lifetime is at least L. In case node i is a relay, we have $E_i = E$. We will discuss how to calculate $E^t_{out}(i)$, $E^t_{in}(i)$ and $E^t_{eav}(i)$ in next section.

When energy provision is sufficient at both terminals and relays, the energy-aware relay placement problem degenerates to relay placement problem for 1-BS-connectivity.

5.2.2 Relay and Routing Co-scheduling Algorithm (R2CA)

Due to integral requirement on y_ls, the above formulation is a mixed integer programming, which is NP-hard. Linear relaxation with iterative rounding is again applied to handle the integral requirement. The resulting algorithm is called relay and routing co-scheduling algorithm (R2CA), shown in Algorithm 5.

The basic idea of linear relaxation with iterative rounding is to solve the optimization problem by relaxing the integral requirement, to round up the relaxed binary variable whose value is closest to 1 to 1, and then to resolve the optimization problem by fixing all the rounded variables, until all the relaxed but unfixed binary variables have integer values or all the relaxed binary variables are fixed at 1. This main process is describe in the first part of Algorithm 5.

Algorithm 5: Relay and Routing Co-scheduling Algorithm (R2CA)

Input : $\mathcal{TS}$, $\mathcal{L}$, $\mathcal{B}$, energy model, radio propagation model
Output: Relay placement solution $\mathcal{R}$, routing $\mathbf{x}^*$

// Linear relaxation with iterative rounding
$\mathcal{R} = \varnothing$
repeat
    Solve OPT_{relay} with $y_l = 1$, if $l \in \mathcal{R}$ and $y_l \in [0,1]$, if $l \notin \mathcal{R}$. Let the optimal solution be $(\mathbf{y}^*, \mathbf{x}^*)$.
    $l^* = arg_{l \in \mathcal{L} \setminus \mathcal{R}} \max y^*_l$.
    $\mathcal{R} = \mathcal{R} \cup \{l^*\}$
until $\forall l \in \mathcal{L},\ either\ y^*_l = 1\ or\ y^*_l = 0$;
// Prune redundant relays; determine optimal routing to minimize the total energy consumption
foreach $r \in \mathcal{R}$ **do**
    $\mathcal{R}' = \mathcal{R} - r$;
    if $OPT_{flow}(\mathcal{R})$ *is feasible* **then**
        $\mathcal{R} = \mathcal{R}'$;
        Let $\mathbf{x}^*$ be the optimal value for $OPT_{flow}(\mathcal{R})$
    end
end

The second part is to prune redundant relays and to determine the optimal flow routing. Being greedy at each iteration, R2CA may place redundant relays. Pruning process is applied immediately after the optimization step, which tries to remove placed relays one by one without violating either connectivity or energy constraints for every terminal topology. We combine this pruning procedure with the optimal flow routing procedure by proposing the following optimization formulation $OPT_{flow}(\mathcal{R})$, which minimizes the total energy consumption with relays placed at various locations included in $\mathcal{R}$, as shown below:

$$\min \quad \sum_{t=1}^{T} f_t L \sum_{i \in V^t} (E_{in}^t(i) + E_{out}^t(i) + E_{eav}^t(i)) \tag{5.11}$$

$$\text{s.t.} \quad \sum_{j:\widehat{ji} \in E^t} x_{ji}^t - \sum_{j:\widehat{ij} \in E^t} x_{ij}^t = d_i^t \tag{5.12}$$

$$\sum_{t:i \in V^t} (E_{in}^t(i) + E_{out}^t(i) + E_{eav}^t(i)) f_t L \leq E_i \tag{5.13}$$

$$\sum_{j:\widehat{lj} \in E^t} x_{lj}^t \leq I\{l \in \mathcal{R}\} \sum_{i \in \mathcal{T}^t} \gamma_i^t, \forall l \in \mathcal{L} \tag{5.14}$$

$$x_{ij}^t > 0 \tag{5.15}$$

For a given relay placement solution $\mathcal{R}$, there may exist various flow routing schedules if the energy consumption constraints are not tight. Among all the routing schedules, we would like to choose the one with the minimum energy consumption in objective function (5.11), which attempts to concentrate the flows along the most energy-efficient paths. A candidate relay place l can route any traffic, only when a relay is placed at location l, as indicated in constraint (5.14). $I\{A\}$ is an indicator function, whose value is 1 if event A is true; 0 otherwise. If $OPT_{flow}(\mathcal{R})$ has a feasible flow schedule, relay placement solution $\mathcal{R}$ is a valid one in terms of both connectivity and network lifetime requirements. Using $OPT_{flow}(R)$, we seamlessly combine the pruning process and the optimal routing process. We can use any standard linear programming technique to solve $OPT_{flow}(R)$ since all the variables have real values.

5.3 Calculating Mean Energy Consumption Rates

The energy-aware relay placement formulation in the previous section assumes that we know various mean energy consumption rates, $E_{in}^t(i)$, $E_{out}^t(i)$ and $E_{eav}^t(i)$, at a node i. Now, we will see how to determine those rates theoretically for a given routing schedule $\mathbf{x}$ in each topology $\mathcal{T}^t$. Besides traffic rate, energy consumption rate also depends on the transmission power and data rate. Smaller transmission power saves energy by operating the transmitter circuit at lower power level. Higher data rate saves energy by reducing the transmission time for a packet. However, both require shorter distance between the sender and the receiver to achieve an acceptable level of received signal quality. Its side effect is that we save mean energy consumption rate by deploying more relays. Therefore, we should carefully balance the tradeoff between the energy consumption and the cost of relays. Power control with fixed data rate and data rate control with fixed power is complementary to each other. In order to clearly understand their impact on the relay placement problem, we will discuss them separately. In reality, they can be combined together to further save energy.

5.3.1 Power Control with Fixed Data Rate

When transmission power control is enforced, a node uses a sufficient and minimum power level to send packets to a neighbor. A power level is selected so that all the data packets and ACKs are received correctly, thus with delivery ratio almost at 1. The mean energy consumption rate for outbound traffic and inbound traffic at node i in topology $\mathcal{T}^t$ is:

$$E^t_{out}(i) = \sum_{\widehat{ij} \in E^t} E_s(|ij|)\frac{x^t_{ij}}{r_{data}} + \sum_{\widehat{ij} \in E^t} E_r \frac{x^t_{ij}}{r_{basic}} \frac{s_{ack}}{s_{data}} \tag{5.16}$$

$$E^t_{in}(i) = \sum_{\widehat{ji} \in E^t} E_r \frac{x^t_{ji}}{r_{data}} + \sum_{\widehat{ji} \in E^t} E_s(|ij|)\frac{x^t_{ji}}{r_{basic}} \frac{s_{ack}}{s_{data}} \tag{5.17}$$

For outbound traffic, node i's energy is consumed when i sends out the data packets and receives the corresponding ACKs, described by the first term and second term in Equation (5.16) respectively. $E_s(|ij|)$ is the per-second power consumption to send continuous bits from node i to node j, whose distance is denoted by $|ij|$. $\frac{x^t_{ij}}{r_{data}}$ is the fraction of time when node i is actively sending bits to node j, when data traffic are sent at rate r_{data}. A data packet triggers an acknowledgement (ACK), which is sent via basic rate r_{basic}. $x^t_{ji}\frac{s_{basic}}{s_{data}}$ is the adjusted acknowledgement rate based on data traffic rate. Dividing this number by r_{basic}, we get the fraction of time when node i is actively receiving ACK bits from node j. Similar interpretation is also applicable for Equation (5.17).

Not only is energy consumed by intended traffic, but also it is consumed by unintended traffic, whose packet signal level is above a node's carrier sense threshold, as shown in Figure 5.4 (a). Because of power control, there is no clear definition for carrier sense range. Let's define Carrier Sense Transmission Set at node i for topology $\mathcal{T}^t$ ($CSTS^t_i$) to be the set of ordered node pairs from the sender (excluding i) to the receiver (excluding i), if the sender's signal perceived at node i is above node i's carrier sense threshold $CSThres$. On the assumption of uniform radio propagation along different directions and well calibrated

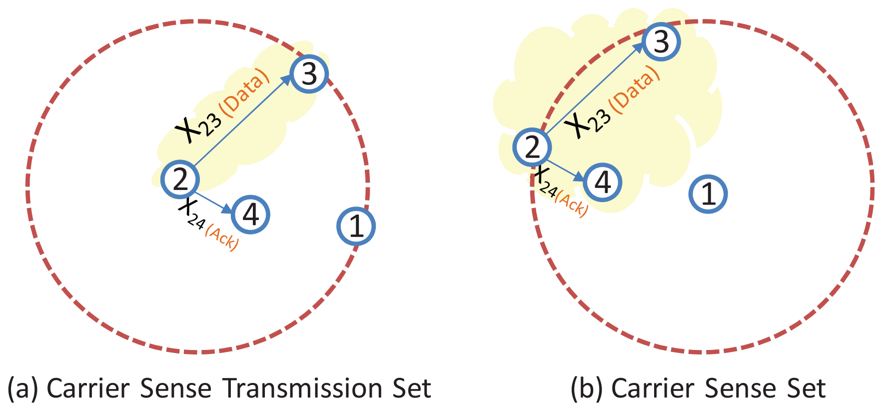


Figure 5.4: Carrier Sense Transmission Set @ node 1: {(2, 3), (3, 2), (4, 3)} and Carrier Sense Set @ Node 1 : {2, 3, 4}

wireless circuits, we have

$$CSTS_i^t = \{(j,k)|P_{ji}^r(P_{jk}^t) \geq CSThres, i\bigcap\{j,k\} = \emptyset, i,j,k \in \mathcal{T}^t\} \tag{5.18}$$

, where P_{jk}^t is the controlled transmission power from node j to k for topology $\mathcal{T}^t$. $P_{ji}^r(P_{jk}^t)$ is the received signal level at node i when node j uses transmission power P_{jk}^t. If any packets from $CSTS_i^t$ are transmitted, node i wastes energy on overhearing those packets, no matter whether packets are corrupted or not due to interference. Here, we ignore accumulated interference. Shown in Equation (5.19), the first term corresponds to data traffic and the second term corresponds to acknowledgement traffic. For the acknowledgements, we need to consider the ordered node pair from the reverse link $\widehat{kj}$ for data traffic sent along link $\widehat{jk}$.

$$E_{eav}^t(i) = \sum_{\widehat{jk} \in CSTS_i^t} E_r \frac{x_{jk}^t}{r_{data}} + \sum_{\widehat{kj} \in CSTS_i^t} I\{\widehat{jk} \in E^t\} E_r \frac{x_{jk}^t}{r_{basic}} \frac{s_{ack}}{s_{data}} \tag{5.19}$$

5.3.2 Data Rate Control with Fixed Power

With data rate control, all the nodes always transmit data packets using the same default power level. However, they can adjust the link rate so that packets are transmitted in a shorter time period, when the transmitter or receiver circuit is actively consuming energy. But the data rate cannot be arbitrarily large. On one hand, it is due to the technology limit of digital circuits. On the other hand, it is due to the noise and interference level. If packets are received with the signal level below SINR (signal to noise and interference ratio), the receiver circuit is not able to decode the packets. Due to this minimum SINR requirement, high link rate also requires good received signal level, thus smaller distance between node pairs. Relay placement algorithms should take data rate into consideration to maintain acceptable level of network lifetime when minimizing the total number of relays.

Wireless channel is a stochastic channel, rather than a binary one. Packets are often received by the intended receiver probabilistically. Whether the packet is successfully received or not depends on many factors, such as packet size, sending rate, node distance, transmission power and radio propagation model. Since we are interested in the sending rate, we will write this probability as a function of sending rate, packet size, and node pair from a topology. Let $p^t_{ij}(r, s)$ represent the one-way packet delivery ratio from node i to node j with packet size s sent at rate r in topology $\mathcal{T}^t$. We use p^t_{ij} (without parameters) to denote the one-way acknowledgement delivery ratio from node i to node j sent at basic rate r_{basic}, such as 6Mbps for 802.11g. Under data rate control, the mean energy consumption rates for outbound traffic and inbound traffic at node i are:

$$E^t_{out}(i) = \sum_{\widehat{ij} \in E^t} E_s \frac{x^t_{ij}}{p^t_{ij}(r^t_{ij}, s_{data}) p^t_{ji}} \frac{1}{r^t_{ij}} + \sum_{\widehat{ij} \in E^t} E_r \frac{x^t_{ij}}{p^t_{ji}} \frac{s_{ack}}{s_{data}} \frac{1}{r_{basic}} \tag{5.20}$$

$$E^t_{in}(i) = \sum_{\widehat{ji} \in E^t} E_r \frac{x^t_{ji}}{p^t_{ji}(r^t_{ji}, s_{data}) p^t_{ij}} \frac{1}{r^t_{ji}} + \sum_{\widehat{ji} \in E^t} E_s \frac{x^t_{ji}}{p^t_{ij}} \frac{s_{ack}}{s_{data}} \frac{1}{r_{basic}} \tag{5.21}$$

In this calculation, we assume that the sender will retransmit a data packet until it has been successfully acknowledged by the receiver. Let the date rate selected from node i to j in topology $\mathcal{T}^t$ be r^t_{ij}. The delivery ratio for a data packet from node i to node j in topology $\mathcal{T}^t$ is $p^t_{ij}(r^t_{ij}, s_{data})p^t_{ji}$ because both the data packet and its acknowledgement should be received successfully before the sender moves on to the next data packet. Hence, node i has to send data traffic at the rate $\frac{x^t_{ij}}{p^t_{ij}(r^t_{ij}, s_{data})p^t_{ji}}$ on average. Out of all the received data traffic, node j is able to decode only $\frac{x^t_{ij}}{p^t_{ji}}$, thus acknowledging the data traffic at this rate on average.

When calculating the energy consumed by unintended traffic at node i, the set of packets which are overheard by node i under data rate control is different from the set under power control, because all nodes use the same default power level for transmission, as shown in Figure 5.4 (b). We define the set of nodes whose traffic can be overheard by node i, no matter who the intended receiver is, as the carrier sense set of node i in topology $\mathcal{T}^t$, CSS^t_i for short. Suppose uniform radio propagation along all directions and well calibrated wireless circuits, we have

$$CSS^t_i = \{j | P^r_{ji}(P_t) \geq CSThres, i \neq j, i, j \in \mathcal{T}^t\} \tag{5.22}$$

, where P_t is the default transmit power and $P^r_{ji}(P_t)$ is the received signal strength at node i from node j.

It is easy to see that the mean energy consumption rate for eavesdropping traffic is

$$E^t_{eav}(i) = \sum_{j \in CSS^t_i} \sum_{\widehat{jk} \in E^t, k \neq i} E_r \frac{x^t_{jk}}{p^t_{jk}(r^t_{jk}, s_d) p^t_{kj}} \frac{1}{r^t_{jk}} + \sum_{j \in CSS^t_i} \sum_{\widehat{jk} \in E^t, k \neq i} E_r \frac{x^t_{jk}}{p^t_{kj}} \frac{s_{ack}}{s_{data}} \frac{1}{r_{basic}} \tag{5.23}$$

The first term is for data packets and the second term is for acknowledgements.

5.3.3 Principle for Rate Selection

In order to keep the network lifetime above the acceptable level, the basic way is to minimize the mean energy consumption rate at a node. Given a fixed routing schedule, even though we have no control over mean energy consumption rate for acknowledgement traffic, we can minimize the mean energy consumption rate for data traffic, by appropriate selection of data rate. From equations (5.20), (5.21), and (5.23), we find out that rate selection is actually independent of relay placement solution, despite the fact that relay placement solution depends on the selected data rate among neighboring nodes. From the three equations, the optimal data rate from node i to j should maximize $p^t_{ij}(r^t_{ij}, s_{data})r^t_{ij}$, which we call effective data throughput at data rate r^t_{ij} for node pair ij when sending data packets of size s_{data}.

We have plotted the effective data throughput for neighbors at different distances from the sender when fixing the data packet size at 1024 bytes in Table 5.1. TxR is the maximum transmission range, set to be 52 meters, which is equivalent to 52dbm for the received signal. We use packet error rate values for 802.11g PHY modes of dei80211mr implementation in NS-2 [34]. Those values have been obtained using a dedicated OFDM physical layer over different multi-path channel realizations with fixed exponential delay profile. As we can see, the highest effective data throughput for nodes $2/5TxR$ apart is 54Mbps, the maximum modulation rate of 802.11g, while it is only 10.5Mbps for nodes TxR apart, at data rate 12Mbps.

Table 5.1: Effective Data Throughput, $S_{data} = 1024$, - means 0

	2/5 TxR	3/5 TxR	4/5 TxR	TxR
6Mbps	6	6	6	5.9
9Mbps	9	9	8.5	6.3
12Mbps	12	12	11.9	10.5
18Mbps	18	17.8	12.4	0.8
24Mbps	24	23.9	4.7	-
36Mbps	36	24.8	-	-
48Mbps	48	0.11	-	-
54Mbps	54	-	-	-

5.4 Energy-aware Routing Algorithm

In order to meet the network lifetime requirement and to minimize the total energy consumption, all the upload traffic from terminals to the command center must follow the routing schedule indicated by $\mathbf{x}^*$. Since all the packets are directed to BSs, there is no need to maintain a dedicated routing table for each base station. Also the routing schedule x^* will not contain any cycle. Because if we remove the cycle, we obtain another routing schedule with smaller energy consumption. To sustain the required network lifetime, there are generally multiple paths towards a few BSs to balance load, thus multiple next-hop neighbors for outgoing traffic. We define a quota for each neighbor j in topology $\mathcal{T}^t$ to be QU_j^t, which is the maximum amount of bytes node j can forward on behalf of node i. It is easy to see, $QU_j^t = x_{ij}^t f_t L$. When nodes are activated in the field, they will forward the received packet p to the next hop j during topology $\mathcal{T}^t$, if QU_j^t is at least the size of the received packet S_p. If no such neighbor is found, the network is said to be partitioned. The time period between this moment and the moment when the network first starts is defined as the actual network lifetime. We show this simple routing algorithm in Algorithm 6.

Algorithm 6: Energy-aware Routing @ node i

Input : Routing schedule $\mathbf{x}$

// Initialization for quota
foreach *topology* $\mathcal{T}^t$*, such that* $i \in \mathcal{T}^t$ **do**
    foreach *neighbor* j*, such that* $\widehat{ij} \in E^t$ **do**
        $QU_j^t = x_{ij}^t f_t L$
    end
end
// Packet Forwarding
foreach *intended packet* p *of size* S_p *from topology* $\mathcal{T}^t$ **do**
    Randomly pick a neighbor j, such that $QU_j^t \geq S_p$;
    $QU_j^t = QU_j^t - S_p$;
    Forward p to node j
end

This algorithm lies above the MAC layer. No matter how many times a packet is retransmitted at the MAC layer, the quota is only reduced once, by the size of data packet. The random selection of the next forwarding neighbor evenly distributes the loads among all the forwarding paths towards BSs, though it is unnecessary for the purpose of preserving the desired network lifetime.

5.5 Evaluation

In this section, we evaluate the performance of our proposed energy-aware relay placement algorithm (ERP), compared with the connectivity-only relay placement algorithm (CRP) in Chapter 4. The metrics are the total number of relays, whether the communication infrastructure meets the network lifetime requirement and the total energy consumption. We use the same concept of DPN instances as in Chapter 4. We randomly assign the traffic rate in multiples of 0.1 Mbps to each terminal. The expected network lifetime is 1 hour by default. Various parameter configurations are summarized in Table 5.2 and Table 5.3. We differentiate terminal topologies by their distribution, to stress the unequal importance of flows from different topologies. The set of valid data rates is 6, 9, 12, 18, 24, 36, 48 and 54 Mbps in 802.11g.

Table 5.2: Simulation Parameters

L	1 hour
s_{data}	1024 bytes
s_{ack}	14 bytes
r_{basic}	6 Mbps
$E_u = E$	1000
E_s	0.66 /second
E_r	0.395 / second

Table 5.3: Network Lifetime Distribution

# of Topologies	Lifetime Distribution of Topologies
1	1
2	[0.6 0.4]
3	[0.5 0.3 0.2]
4	[0.4 0.3 0.2 0.1]

5.5.1 Power Control

In this section, we examine our proposed algorithm with power control. We vary the transmission range of relays and BSs from $txRange$ to $2 * txRange$, while fixing the transmission range of terminals at $txRange$. The corresponding energy consumption rate for $2*txRange$ increases to $4*0.66 = 2.64$. Each DPN instance contains 2 topologies of 4 terminals each. We vary the network area from $2 * txRange$ to $4 * txRange$. The granularity of relay placement scheme is half the transmission range.

The performance of energy-aware and connectivity-only relay placement algorithms are plotted in Figure 5.5 and Figure 5.6, when the transmission range of relays and BSs is set to be $txRange$. As we can see, ERP meets all the network lifetime requirement for 5 randomly generated instances, while CRP only meets 2 or 3 instances, when network area is relatively large. When network area is $2txRange * 2txRange$, CRP has the same performance as ERP since terminals from a topology are almost connected by themselves. Figure

5.6 claims that ERP does not incur too much overhead regarding the number of relays, while keeping the expected network lifetime above an acceptable value.

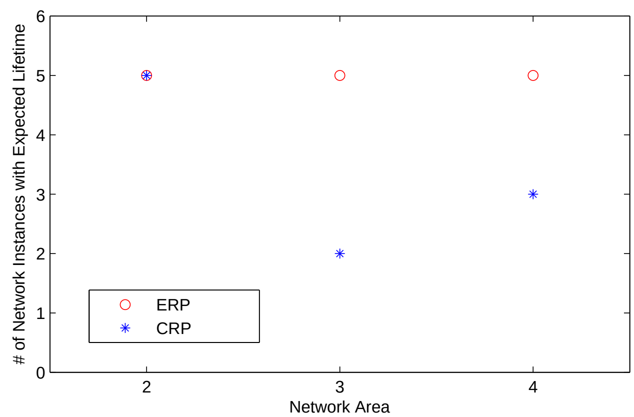


Figure 5.5: Number of Network Instances with Expected Network Lifetime, *txRange*

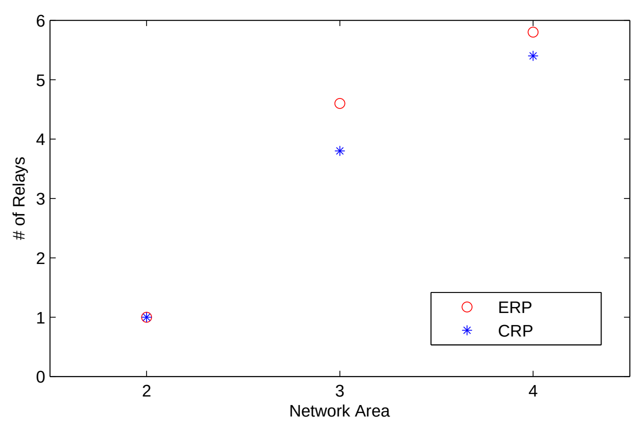


Figure 5.6: Number of Relays, *txRange*

The performance of two placement algorithms are plotted in Figure 5.7 and Figure 5.8, when the transmission range of relays and BSs has been increased to $2 * txRange$. Again, ERP beats CRP in terms of its ability to meet network lifetime requirement with less than one extra relay placed on average. Out of all the 5 DPN instances generated per DPN configuration, ERP satisfies the lifetime requirement for all instances, while CRP only satisfies 4, 1, and 1 instance.

5.5.2 Data Rate Control

In this section, we examine our proposed algorithm with data rate control. We have the same setting as the previous section, except that data rate control replaces power control. The performance of ERP and

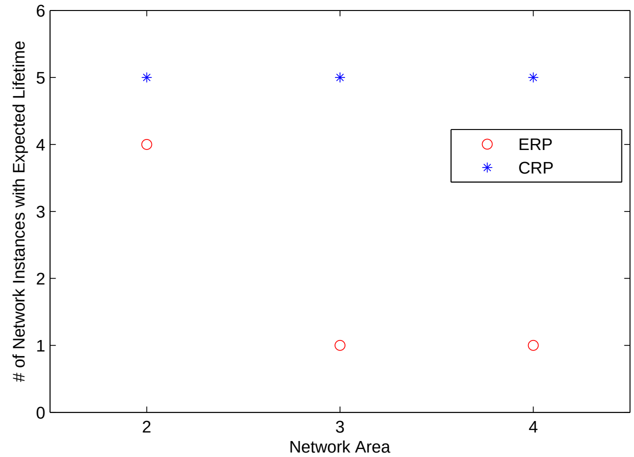


Figure 5.7: Number of Network Instances with Expected Network Lifetime, $2 * txRange$

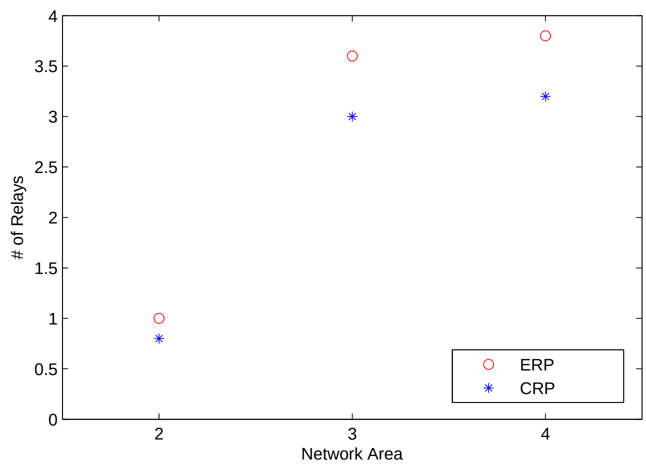


Figure 5.8: Number of Relays, $2 * txRange$

CRP are plotted in Figure 5.9 and Figure 5.10, when the transmission range of relays and BSs is set to be $txRange$. The performance of ERP and CRP are plotted in Figure 5.11 and Figure 5.12, when their transmission range increases to $2 * txRange$.

As we can see, both ERP and CRP meet all the network lifetime requirements for 5 randomly generated instances with a similar number of relays being deployed. When the transmission range for relays becomes larger, a smaller number of relays are needed.

Though CRP has a similar performance as ERP to meet network lifetime requirements with data rate control, it has different performances in terms of network-wide energy consumption. Shown in Figure 5.13, ERP is able to reduce the network-wide energy consumption when network area is large, due to its capability to include energy consumption in the optimization formulation.

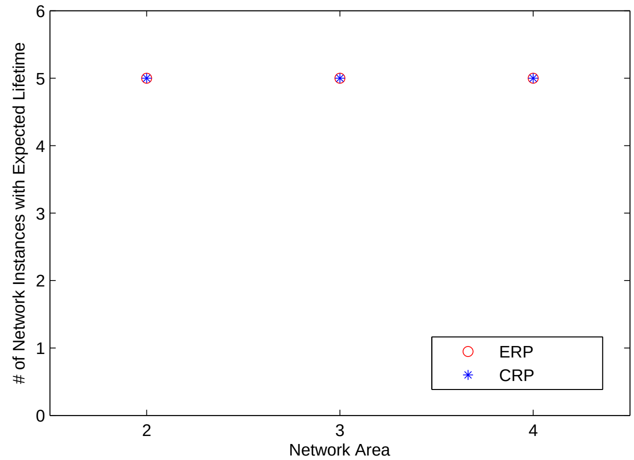


Figure 5.9: Number of Network Instance with Expected Network Lifetime, *txRange*

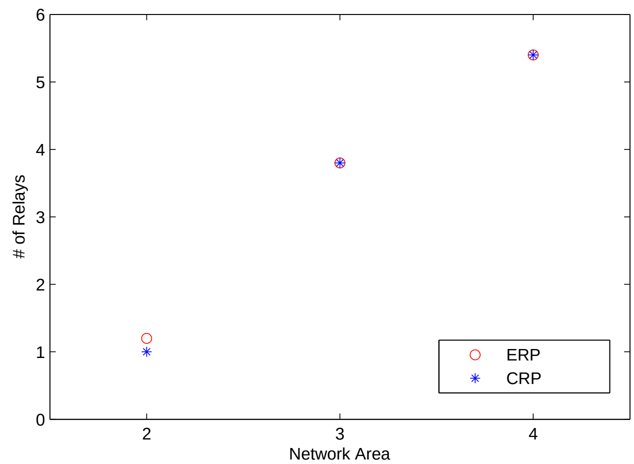


Figure 5.10: Number of Relays, *txRange*

One interesting observation from comparative results between power control and data rate control is that data rate control is more efficient than power control to reduce power consumption, thus meeting network lifetime requirement. The plausible reason is that data rate control reduces the energy consumption at sender, receiver and eavesdropper sides, while power control reduces the energy consumption only at the sender side, though it is able to reduce the total number of eavesdroppers.

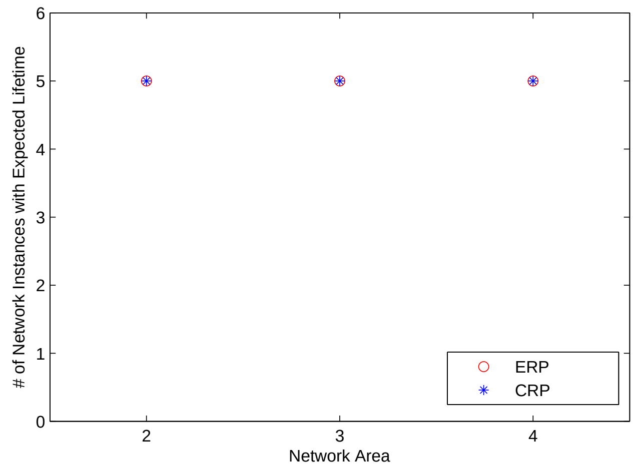


Figure 5.11: Number of Network Instances with Expected Network Lifetime, $2 * txRange$

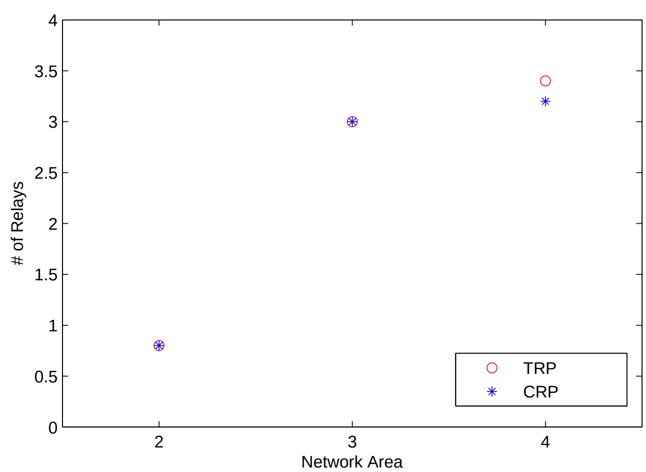


Figure 5.12: Number of Relays, $2 * txRange$

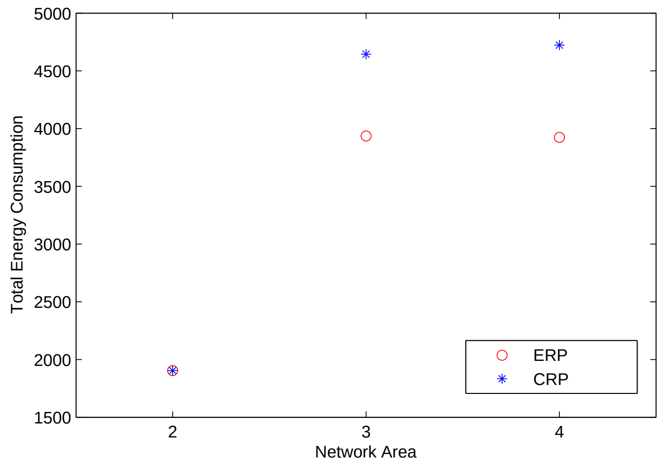


Figure 5.13: Total Energy Consumption, $2 * txRange$

Chapter 6

M-Breadcrumb: Mission-aware Relay Deployment in DDNs

In order to support all the real-time monitoring and control applications, a communication infrastructure offering reliable communication paths betweens commanders and FRs is needed. In this chapter, we are interested in establishing such an infrastructure on demand for DDNs cost-effectively, so that FRs can be continuously monitored, informed, managed and protected, while mobilized around a large incident area. Assuming locations of all BSs and mission logistics are given, we need to decide how to strategically deploy relays so that all the terminals maintain BS-connectivity at all time, or maintain *persistent BS-connectivity*. The resulting relay deployment algorithm running at terminals instructs the FR to drop relays at a series of locations, while the officer is mobilized around to execute missions with a sequence of tasks.

Without any prior knowledge, a straightforward method is to deploy a relay whenever the connectivity to BSs is about to break [11]. The deployed relays have been referred to as "breadcrumbs" in literature and this approach is also called *breadcrumb approach*. Besides maintaining persistent BS-connectivity, breadcrumb approach decently copes with environment dynamics, regarding radio propagation and obstacles. Nevertheless, where to drop relays is totally driven by FRs' mobility patterns and environment factors.

However, the breadcrumb approach does not exploit the unique properties of FRs' mobility patterns: (a) being organized and coordinated as mobile groups; (b) being mission-oriented. Thus, it may deploy more relays than necessary. First, FRs with the same mission form a group, inside which tight cooperation ensures that all FRs always move within close proximity and shield one another from danger in emergency situations. For instance, two-in-two-out policy mandates that firefighters never go into a dangerous situation alone. Hence, we can abstract a group of FRs by a virtual node. Second, FRs come from various organizations and agencies with different missions. Different groups have diverse targets and task sequences, geographically distributed. Because of careful operation planning in advance and global resource management, this mission knowledge is accessible and can be exploited to reduce the cost of relay deployment.

Relays are valuable resources with a limited supply. It is desirable for relay deployment algorithms to minimize the number of deployed relays, while preserving persistent BS-connectivity for every mobile group. This reduces not only the deployment cost, but also the retrieval cost when operations are completed. We have proposed a mission-aware constrained relay deployment algorithm, *M-Breadcrumb*, to address the relay cost. Our idea to achieve fewer relays is to reuse them among FR groups with different missions and among different tasks of a FR group. Because we strategically place relays to maximize the reuse opportunity, FRs may have to detour from the original shortest path towards the task destination to those deploying places. It is important to state that we don't change their mobility patterns on the coarse level. They still move from one destination to another, where they sojourn a while and execute tasks. What we adjust is their trajectory towards the destinations so that they are always moving within the transmission range of the available communication infrastructure, and they will bypass the relay dropping locations, if needed. Hence, M-Breadcrumb is a joint relay placement and online navigation algorithm. The relay placement algorithm maximizes relay reuse opportunity using mission information, while the online navigation algorithm controls the detour with reasonable traversal overhead. To the best of our knowledge, our work is the first one to study the optimal constrained relay deployment problems.

6.1 Main Framework

In this section, we outline the main framework of our mission-aware relay deployment algorithms. Though missions are not amendable, FRs can adjust their trajectory between the current and next destinations. FRs will be directed to relay deploying locations chosen from $\mathcal{R}_c^*$, which is globally calculated to minimize the total number of relays based on mission topologies. When FRs move among destinations and deploying locations, they are assured to be covered by TTCI, thus connected to BSs. Our M-Breadcrumb algorithm is decomposed into two steps: (a) mission-aware relay placement; (b) online navigation. The mission-aware relay placement algorithms compute a complete relay placement solution $\mathcal{R}_c^*$, which connects all the destinations in the list DL^i to the starting BS B^i for every group G^i. Because B^i is BS-connected, all the other destinations are BS-connected. The algorithms try to minimize the number of relays, by maximizing the relay reuse by (a) different destinations of a mission and (b) different mobile groups. Online navigation algorithm determines a subset of candidate locations ($\subseteq \mathcal{R}_c^*$) to drop relays, and guides FRs towards those deploying points and the next destination, when FRs relocate from one destination to another, with small traversal overhead.

In Figure 6.1, we show a complete relay placement solution and deployment sequence for the scenario in Figure 3.4. Here, group G^1 visits locations $P1$ and $P2$ in order, while G^2 only visits $P1'$. The set of triangles is a complete relay placement solution $\mathcal{R}_c^*$, while labeled ones compose a deployment solution $\mathcal{R}$. A relay labeled with numbers $i - j$ means that it is j_{th} relay being deployed by group G^i. In the next two sections, we will elaborate these two problems separately.

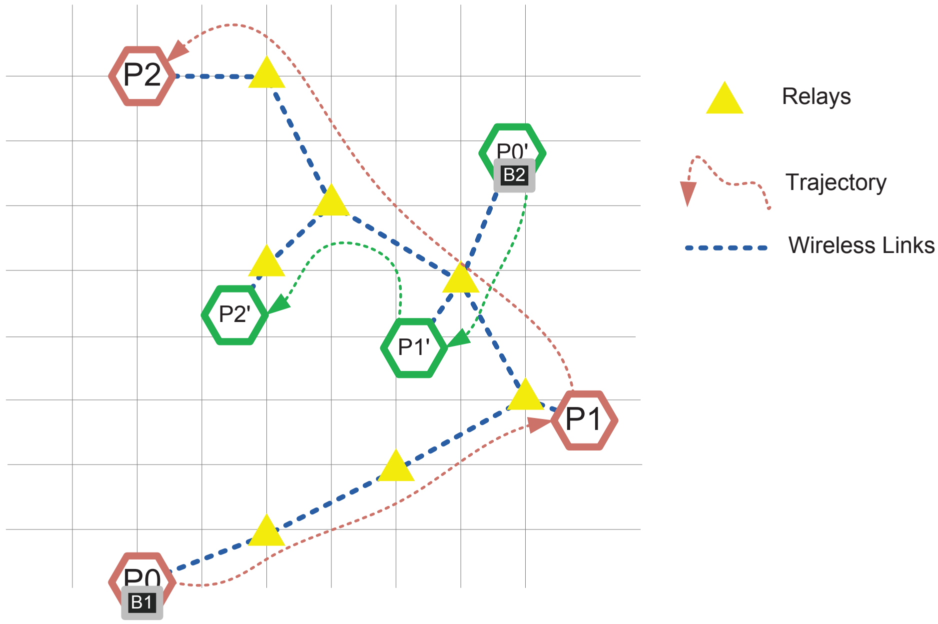


Figure 6.1: Persistent BS-Connectivity for Mobile Groups

.

6.2 Mission-aware Relay Placement

In this section, we will formalize the mission-aware relay placement problems. Two mission-aware relay placement algorithms are described. The first one is based on multi-commodity flow formulation. We formulate the problem as a mixed integer programming problem and apply the iterative rounding technique for solution. The second algorithm is based on graph theory. First, a disk cover on all the destinations is calculated. Then, a Steiner forest is constructed to connect all the covering disks for each mission topology.

6.2.1 Problem Formulation

Now, we will formally define mission-aware relay placement problems. Given a DDN instance, we infer an undirected graph $G = (V, E)$, wherein $V = \cup_i \mathcal{MT}^i \cup \mathcal{LB}$ and

$$E = \{uv||uv| \leq txR, u, v \in V, \{u, v\} \bigcap \mathcal{LB} \neq \varnothing\} \tag{6.1}$$

, where $\mathcal{LB} = \mathcal{L} \bigcup \mathcal{B}$ and $|uv|$ means the Euclidean distance between node u and v. Vertex and edge sets are defined upon $\mathcal{L}$, as if we were placing a relay at every candidate location.

Our objective is to find a complete placement solution $\mathcal{R}_c^* = \{r_1 \cdots r_{p^*}\}$ ($\mathcal{R}_c^* \subseteq \mathcal{L}$) with the minimum number of relays p^* so that for any group G^i, all vertexes in $DL^{i'}$ are connected. Destinations must be leaves. This problem is very similar to Steiner forest problem, in the way that a Steiner forest connects all the vertices for each mobile group together. However, no destinations serve as inner nodes of forests, except BSs. In Figure 6.1, we show a complete placement solution of relays depicted in triangles.

6.2.2 Flow-based Solution

The first solution for mission-aware relay placement problem is based on multi-commodity flow. We associate one type of commodities with each mission topology. In G^i, each destination $dest_k^i, (k \geq 1)$ needs to send one unit of type-i commodities to B^i. Hence, B^i receives p_i units of type-i commodities. d_u^i is used to denote the supply/demand of node u in G^i. Relays neither supply nor demand anything; hence their d values are 0. We have

$$d_u^i = \begin{cases} p^i, & i = B^i, \, p^i = |DL^i| \\ -1, & i \in DL^i \\ 0, & i \in \mathcal{LB} - \{B^i\} \end{cases} \tag{6.2}$$

. Destinations cannot forward any commodities, except generating their own. In addition, flows have directions. Hence, we refine each undirected edge with two directed edges, and remove all edges pointed from a relay to a destination. The refined directed edge set is

$$\begin{aligned} E^* \quad = \quad & \{\widehat{uv} | uv \in E, u, v \in \mathcal{LB}\} \qquad (6.3) \\ & \cup \{\widehat{uv} | uv \in E, \exists G^i, u \in DL^i, v \in \mathcal{LB}\} \end{aligned}$$

The optimization of relay placement problems is as follows:

$$\min \quad \sum_{l \in \mathcal{L}} y_l \tag{6.4}$$

$$\text{s.t.} \quad \sum_{j:\widehat{vu} \in E^*} x_{vu}^i - \sum_{j:\widehat{uv} \in E^*} x_{uv}^i = d_u^i \tag{6.5}$$

$$\sum_{G^i} \sum_{u:\widehat{lu} \in E^*} x_{lu}^i \leq y_l \sum_{G^i} |DL^i|, \forall l \in \mathcal{L} \tag{6.6}$$

$$y_l \in \{0, 1\} \tag{6.7}$$

$$x_{uv}^i > 0 \tag{6.8}$$

The decision variables are binary variables y_ls, as shown in constraint (6.7). If $y_l = 1$, we place a relay in location l; otherwise, no relay is placed. A complete placement solution is $\mathcal{R}_c = \{l|y_l = 1, l \in \mathcal{L}\}$. The cost of placing a relay in location l is one. Hence, the objective function in (6.4) corresponds to the total number of relays. When the objective function is minimized, a complete placement solution with the minimum number of relays is located. Let x^i_{uv} denote the flow of type-i commodities for mission topology $\mathcal{MT}^i$ on directed edge $\widehat{uv}$, and type-i commodities can route through edge $\widehat{uv}$ only if $\widehat{uv} \in E^*$. We refer to the constraints in (6.5) as *mass balance constraints*. The first term for a node u represents the total inflow of the node and the second term represents the total outflow of the node for topology $\mathcal{MT}^i$. The mass balance constraints state that the inflow minus outflow must equal the supply/demand of the node for $\mathcal{MT}^i$. In constraints (6.6), a relay location l can route any commodity, only if a relay is placed at l ($y_l = 1$), which we refer to as *relay switch constraints*. The constraints (6.5)-(6.8) together indicate that a placement solution must have a path composed of relay connecting every destination with the starting BS for every mission topology. We use standard technique of linear relaxation with iterative rounding to obtain an integral solution for y_ls.

6.2.3 Graph-based Solution

Adopting the methods in [9], we propose an approximation algorithm for mission-aware relay placement problems, which contains two parts. In the first part, any known algorithm $\mathcal{A}$ is used to compute a single-cover $\mathcal{R}_\mathcal{A}$ of all the destinations in $\bigcup_i \mathcal{MT}^i$. In the second part, any known algorithm $\mathcal{B}$ for Steiner forest problem is used to augment $\mathcal{R}_\mathcal{A}$ into $\mathcal{R}_c$ so that the covering set is connected for every mission topology.

Details are sketched in Algorithm 8. In step 1, a single-cover $\mathcal{R}_\mathcal{A}$ is computed. But, BSs receive special treatment, in the way that we assume a relay is always placed at a BS. In step 2, we construct an undirected relay communication graph $G^{\mathcal{LB}}$, which contains all candidate relay locations and BSs. Because node-weighted Steiner forest is hard to compute, the node weight is shifted to edge weight for calculation of an edge-weighted Steiner forest in step 3. The weight of an edge is set to the number of relay places, not included in $\mathcal{R}_\mathcal{A}$, since a place in $\mathcal{R}_\mathcal{A}$ incurs no extra cost. In step 4, the minimum covering set for every mission topology is obtained. Step 5 calculates a low-weight Steiner forest $\mathcal{F}_\mathcal{B}$. Placement solution $\mathcal{R}_c$ contains all relay vertexes in $\mathcal{F}_\mathcal{B}$.

In Figure 6.2, we illustrate the two steps for the DDN instance in Figure 3.4. In (a), a single-cover contains the set of all the relays in triangles, including two at BSs. In (b), a Steiner forest is shown to connect all the covering relays for each mission topology. More specifically, the forest connects all the relays marked by 1 (2) together, which is a minimumm cover for $\mathcal{MT}^1$ ($\mathcal{MT}^2$). Note that one triangle is marked by both 1 and 2. It is just a coincidence that the Steiner forest becomes a Steiner tree in this example.

Algorithm 7: Graph-based Relay Placement Algorithm

Input : DDN instance, $\mathcal{L}$

Output: Complete relay placement solution $\mathcal{R}_c$

Applying algorithm $\mathcal{A}$ to obtain a single-cover* $\mathcal{R}_{\mathcal{A}}$ of the DDN instance. Without loss of generality, we assume that $\mathcal{R}_{\mathcal{A}}$ is minimal, meaning that none of its proper subset is a single-cover of the DDN instance.

Construct an undirected relay communication graph $G^{\mathcal{LB}} = (\mathcal{LB}, E)$, and $E = \{uv | uv \in E, u, v \in \mathcal{LB}\}$

Set edge uv's weight: $w(uv) = |(\mathcal{LB} - \mathcal{R}_{\mathcal{A}}) \bigcap \{u, v\}|$.

From $\mathcal{R}_{\mathcal{A}}$, let the minimum set of covering disks for $\mathcal{MT}^i$ be S^i.

Applying algorithm $\mathcal{B}$ to compute a low weight Steiner forest $\mathcal{F}_{\mathcal{B}}$ of $G^{\mathcal{LB}}$, so that each pair of vertexes belonging to the same set S^i is connected.

Let $\mathcal{R}_c$ be the vertexes in $\mathcal{F}_{\mathcal{B}}$, excluding BSs.

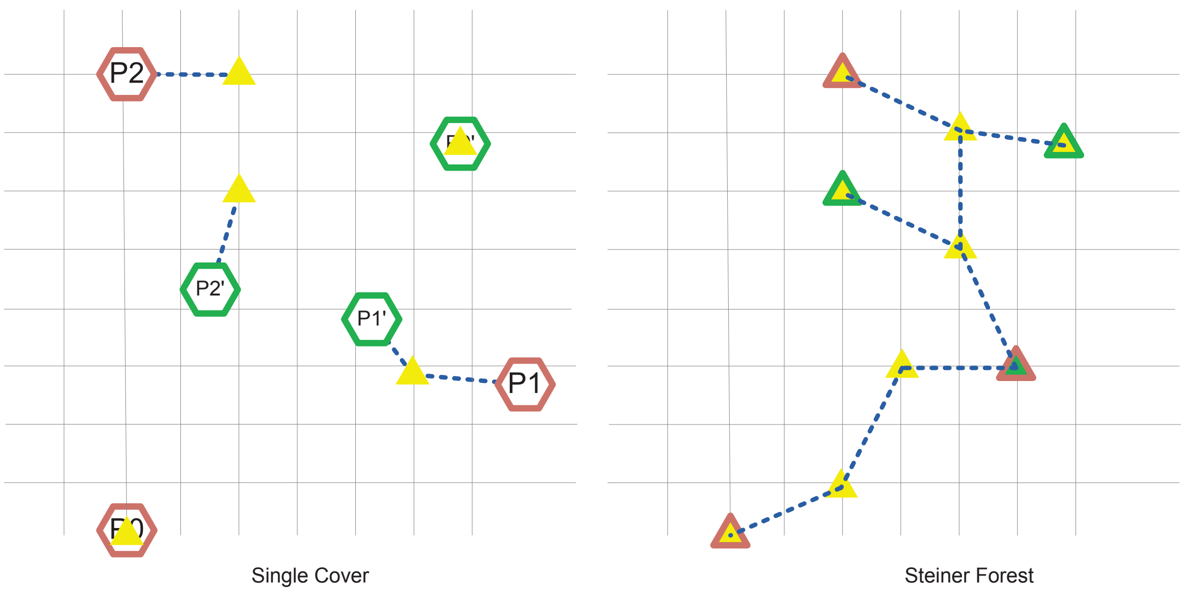


Figure 6.2: Graph-based solution

6.3 Online Navigation

Suppose a mobile group wants to move from destination $dest_i$ to destination $dest_{i+1}$, online navigation algorithm will direct a route with persistent BS-connectivity, towards $dest_{i+1}$, while passing necessary relay deploying points. Online navigation has to address two problems: (a) Assuming $dest_i$ and $dest_{i+1}$ are not reachable via the existing communication infrastructure, where shall the group deploy additional relays to minimize the traveling distance; and (b) Suppose the current and destination locations are reachable by the existing infrastructure, what the trajectory looks like to minimize the traveling distance. Two algorithms are proposed to solve these two problems. Deploying point planning algorithm calculates various locations where FRs should drop additional relays. Trajectory planning algorithm computes the shortest path between the current location and the next deploying point or task destination.

6.3.1 Deploying Point Planning

The deploying point planning algorithm is called when a group wants to move from the current task destination $dest_i$ to the next one $dest_{i+1}$, as described in Algorithm 8. Additional inputs are the complete

Algorithm 8: Deploying Point Planning Algorithm

Input : Complete placement solution $\mathcal{R}_c$; comm. infrastructure $\mathcal{R}_t$ at current time t; current and next destination $dest_i$ and $dest_{i+1}$
Output: A list of relay deploying points LDP

1 **if** $Conn(\mathcal{R}_t, dest_i, dest_{i+1})$ **then**
2 $\quad LDP = \emptyset$
3 **else**
4 $\quad$ Generate undirected subgraph $H = (V, E)$, where $V = \mathcal{R}_c \cup \{dest_i, dest_{i+1}\}$ and $E = \{uv \| uv| \leq txR, u, v \in V, \{u, v\} \bigcap \mathcal{R}_c \neq \varnothing\}$
5 $\quad$ Set the weight of edge uv: $w(uv) = |uv|$;
6 $\quad$ Find the shortest path p from $dest_i$ to $dest_{i+1}$ by Dijkstra's algorithm. Say $p = \{p_0, p_1, \cdots, p_k\}$, where $p_0 = dest_i$ and $p_k = dest_{i+1}$;
7 $\quad LDP = \{p_1, \cdots, p_{k-1}\} - \mathcal{R}_t$ (order reserved)
8 **end**

placement solution $\mathcal{R}_c$ calculated from section 6.2 and the existing communication infrastructure R_t by current time t. The output is the list of relay deploying points, denoted by LDP. If the group can move under the cover of R_t to $dest_{i+1}$ in line 1, there is no need to drop additional relays. Hence, we set LDP to be empty in line 2. Let $Conn(\mathcal{R}_t, dest_i, dest_{i+1})$ be such a procedure to determine whether a trajectory exists or not. The procedure details will be provided later. If no feasible trajectory exists ($Conn = false$), relay deploying points are calculated in *else* branch. In line 4, we generate an undirected subgraph $H = (V, E)$ on the vertex set containing placement solution $\mathcal{R}_c$ and the current and next destinations. Note that no edge exists between two destinations. In line 5, we set the weight for all edges to be their Euclidean distance between its two end points. In line 6, we calculate the shortest path from $dest_i$ to $dest_{i+1}$ by Dijkstra's algorithm, where the vertexes in the middle of the path are the relay locations. If a relay is already deployed in a location, we don't drop again. Thus, we ignore all places where relays have already been dropped by time t. In line 7, LDP contains all the additional deploying locations, whose visiting order should be reserved from path p. Guided by online navigation, after the mobile group drops relays at all the places in LDP in sequence, it moves to $dest_{i+1}$. We call either a deploying point or a destination a *stop point*.

An example is given in Figure 6.3. Relays in $\mathcal{R}_t$ are r_1, r_2 and r_3. If a group wants to move from $P1$ to $P2$, it can proceed directly. However, if it wants to move from $P2$ to $P3$, extra relays must be dropped. From the graph, the shortest path from $P2$ to $P3$ is $\{P2, r_3, r_4, P3\}$. Hence, the group first moves to location r_4, drop a relay and continues to $P3$. We have $LDP = \{r_4\}$.

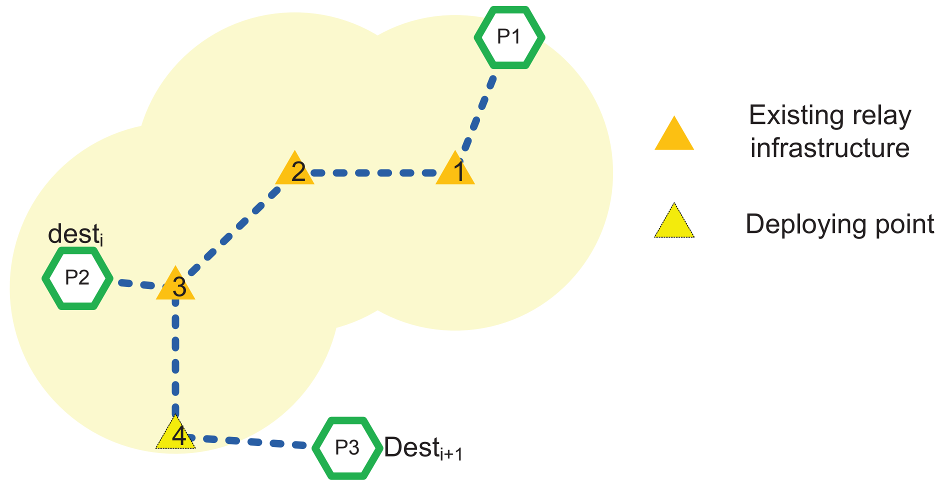


Figure 6.3: Deploying Point Planning

6.3.2 Trajectory Planning

If a group moves based on the deploying point planning algorithm, it is always under the cover of communication infrastructure. This is because whenever the connectivity to BSs is going to lose, a relay is dropped at the deploying location. We have already discussed where to stop when a FR group moves to $dest_{i+1}$. Now, we will discuss how to obtain the shortest trajectory between the current and next stop points. Not all the network area is covered by the communication infrastructure. Hence, the shortest path may not necessarily be a straight line. As shown in Figure 6.4 (a), the shortest path from $P1$ to $P2$ is a straight line, while from $P1$ to r_4 is not. Because of the complex coverage region formed by the communication infrastructure, it is hard to obtain the shortest path using geometric analysis. Our solution is to divide the network area into grid, over which to plan the moving trajectory. Another advantage for grid-based approaches is that they can easily adapt to obstacles and irregular transmission range other than circles.

When planning the movement over the grid, not all the grid vertexes are viable. We only consider the vertexes covered by the infrastructure. Let's say the set of covered grid vertexes, spaced by l, is $\mathcal{G}_c$. We assume that a group can move, in straight line, between *adjacent* grid vertexes or move between a stop point and a grid vertex, if adjacent, while maintaining persistent BS-connectivity. This assumption is generally true, if l is small enough. $Adj(u, v)$ is used to test whether u and v are adjacent ($Adj(u, v) = 1$). $Adj(u, v) = 1$, only if $|uv| \le txR$ to guarantee connectivity. The denser is the grid, the closer to the shortest path is the estimated trajectory. However, it incurs high computation overhead. For example, in Figure 6.4 (b), a group can move to 16 different neighbors. The corresponding Adj function is in Equation (6.9), where $\angle uv$ denotes the angle formed by line uv and x-axis, and $p_{sd} = \{p_s, p_d\}$ (the set of source and destination stop points). $x|\angle uv$ ($x \nmid \angle uv$) means the angle is (not) divisible by x.

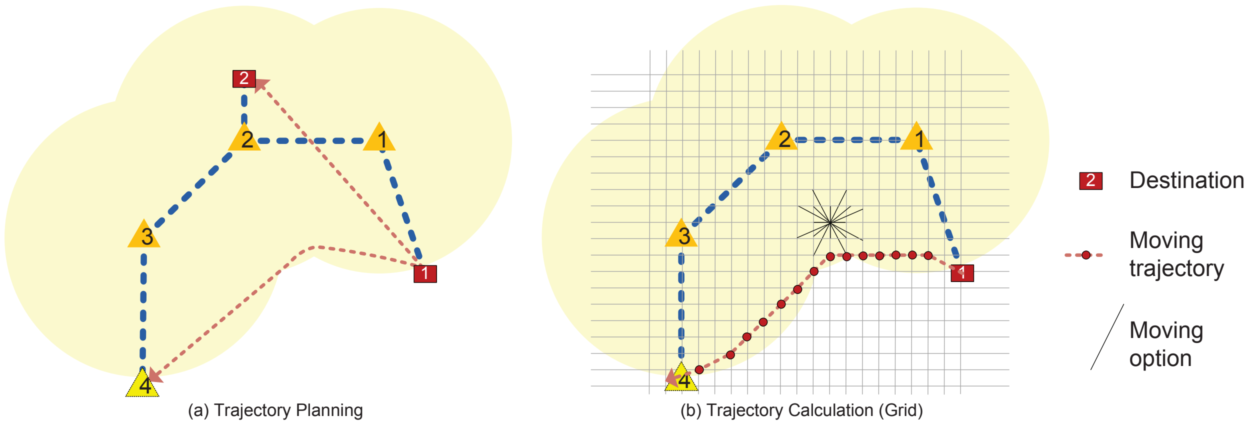


Figure 6.4: Trajectory Planning

$$Adj(u,v) = \begin{cases} 1, & u,v \in \mathcal{G}_c, 30|\angle uv, 90 \nmid \angle uv, |uv| < 2l; \\ 1, & u,v \in \mathcal{G}_c, 30|\angle uv, |uv| \le \sqrt{2}l; \\ 1, & \{u,v\} \bigcap \mathcal{G}_c \neq \varnothing, \{u,v\} \bigcap p_{sd} \neq \varnothing, |uv| \le l; \\ 0, & \text{otherwise.} \end{cases} \tag{6.9}$$

The trajectory planning algorithm is described in Algorithm 9. In line 1, the network area is divided into grid. In line 2, we construct a graph for trajectory calculation. Only the set of grid vertexes covered by the existing communication infrastructure and the source and destination stop points $\{p_s, p_d\}$ are included. The edge set is defined by $Adj(u,v)$ function. In line 3, an edge weight is set to be the Euclidean distance to minimize the total traversal distance. In line 4, we calculate the shortest path $\mathcal{P}$ from p_s to p_d using Dijkstra's algorithm. The FR group moves between adjacent locations in $\mathcal{P}$ in straight line. We show an exemplary trajectory in Figure 6.4(b) assuming Equation (6.9).

Algorithm 9: Trajectory Planning Algorithm

Input : $\mathcal{R}_t$; source and destination stop points p_s, p_d
Output: Moving trajectory $\mathcal{P}$
1 Divide the network area into grids;
2 Construct graph $G = (V, E)$, where $V = \{p | \exists r \in \mathcal{R}_t, p\text{: a grid vertex covered by } r\} \bigcup \{p_s, p_d\}$, and $E = \{uv | Adj(u,v) = 1\}$;
3 Set the weight of edge uv: $w(uv) = |uv|$;
4 Calculate the shortest path $\mathcal{P}$ from p_s to p_d

Algorithm 9 is also used in $Conn(\mathcal{R}_t, dest_i, dest_{i+1})$. If it returns an empty path, then $dest_i$ and $dest_{i+1}$ is not connected by existing infrastructure $\mathcal{R}_t$ ($Conn = false$). Otherwise, $Conn = true$.

6.3.3 Discussion

The description of our online navigation algorithm is meant to give readers a whole picture about how it interacts with mission-aware relay placement. We leave out some technique details for special cases. Now, let us discuss how to amend the algorithm to handle the corner cases and explain why the online navigation only reduces the traversal overhead, not minimizing the overhead.

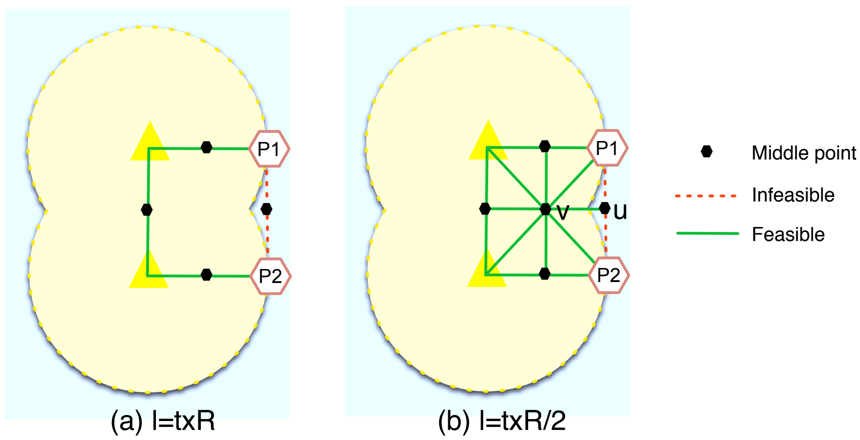


Figure 6.5: Uncovered Lines Between Adjacent Covered Grid Vertexes

One type of corner cases needing special treatment is an uncovered line between adjacent covered grid vertexes. The problems occurs due to our simplified assumption that a terminal can move between adjacent grid vertexes in straight line. We have shown a counterexample in Figure 6.5 (a). Relays are deployed at locations occupied by triangles. Triangles, $P1$ and $P2$ are covered grid vertexes. $l = txR$ and correspondingly,

$$Adj(u, v) = \begin{cases} 1, & u, v \in \mathcal{G}_c, |uv| \leq l; \\ 1, & \{u, v\} \bigcap \mathcal{G}_c \neq \varnothing, \{u, v\} \bigcap p_{sd} \neq \varnothing, |uv| \leq l; \\ 0, & \text{otherwise.} \end{cases} \quad (6.10)$$

. Since $P1$ and $P2$ are within the transmission range of the existing infrastructure, $\{P1, P2\} \subseteq \mathcal{G}_C$. Based on $Adj(u, v)$ definition, $Adj(P1, P2) = 1$. However, the line from $P1$ to $P2$ is not covered except two ends. If a terminal moves between $P1$ and $P2$ in straight line, it will lose connectivity to BSs. In order to address the problems of uncovered lines, we include the α-convex-combination test. $Adj(u, v) = 1$, only if point $\alpha u + (1 - \alpha)v$, the convex combination of points u and v, is also covered by the existing infrastructure. If α-convex-combination tests are done for all $\alpha, \alpha \in (0, 1)$, all the cases of uncovered lines will be detected and eliminated. Nevertheless, for efficiency reason, we will perform the tests for $\alpha = 1/2$ and $\alpha = 1/3$ only. In this way, a majority of uncovered lines are eliminated with little computational overhead. We want to emphasize that α-convex-combination tests are most useful when $l = txR$. If $l < txR$, an appropriately defined *Adj* function may have solved the problem already. We show an example in Figure 6.5 (b), where $l = txR/2$ and *Adj* function is defined in Equation (6.9). $P1$ and $P2$ is not adjacent, even though their distance is equal to txR. Grid vertex u, the middle point of $P1$ and $P2$, does not belong to $\mathcal{G}_c$. The output shortest path from $P1$ to $P2$ is $\{P1, v, P2\}$, which is completely covered.

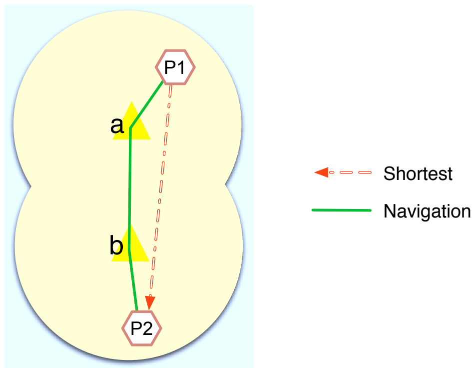


Figure 6.6: Approximation for the Shortest Path

Finally, we want to point out that our online navigation uses various heuristics to reduce the total traversal distance, not minimizing it. Minimizing the traversal overhead requires tremendous geometric analysis, especially when the covered region formed by the existing communication infrastructure reveals complex shapes. We give an sub-optimal example in Figure 6.6. The shortest path from $P1$ to $P2$ is the red dotted straight line. However, we will output $\{P1, a, b, P2\}$ due to the constraints that a terminal can only move between adjacent covered grid vertexes and task destinations.

6.4 Evaluation

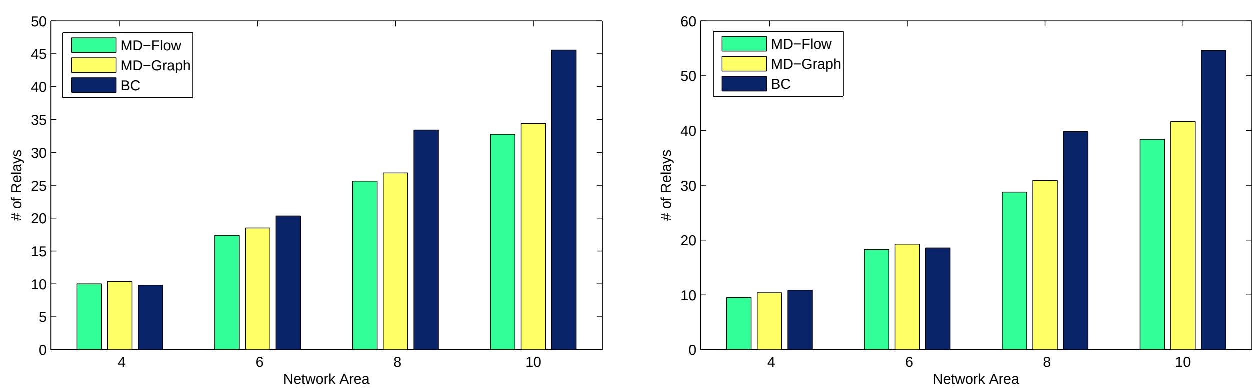
In this section, we evaluate the performance of three proposed algorithms:M-Breadcrumb with flow approach (*MBC-Flow*), M-Breadcrumb with graph approach (*MBC-Graph*) and breadcrumb approach (*BC*). The metrics considered are the number of relays deployed and relative walking distance, under various parameters, e.g., transmission range, the number of mobile groups, network area, and the size of destination list. Relative walking distance of MBC-Flow (MBC-Graph) is the ratio of walking distance of MBC-Flow (MBC-Graph) over BC, which quantitatively measures the traversal overhead. Every figure is averaged over 20 DDN instances randomly generated from a DDN configuration. A *DDN configuration* is defined by the network area, the number of mobile groups, the size of destination list and the locations of BSs, assuming each group has the same number of destinations to visit. We generate a *DDN instance* by generating the desired number of destinations randomly in the network area and the starting location randomly selected from the set of BSs for each group. In all DDN configurations, two BSs are placed, one at the top-right corner and one at the bottom-left corner. For simplicity, transmission range of all the nodes is set to be 100m by default. The relay placement scheme is a regular grid with 100m between adjacent locations.

The number of relays deployed by breadcrumb approach highly depends on mobility patterns, particularly the order by which each group visits the set of destinations. Also mobility patterns affect walking distance for mission-aware relay deployment algorithms. Hence, the performance of three algorithms is further averaged over 3 movement instances. A movement instance is a random permutation of the locations specified in the destination list. With multiple groups, it is unnecessary to simulate all the instantaneous terminal topologies, when examining the mobility pattern related performance. For simplicity, we assume that all groups take turns to move from the current task destination to the next one, while the turns are also randomly generated. The grid granularity in trajectory planning equals the granularity of relay placement scheme unless specified. A terminal can move to 8 directions, which are multiples of 45^o.

We use a 22-approx algorithm $\mathcal{A}$ for single-cover problem, whose implementation is obtained from [9] and a 2-approx algorithm $\mathcal{B}$ for Steiner forest problem, found in [35].

6.4.1 Size of Destination List

The first set of experiments is to examine the number of deployed relays and traversal overhead for different network sizes, with a fixed number of destinations for 3 mobile groups. Two destination list sizes are evaluated, 6 and 10.



(a) 6 Destinations (b) 10 Destinations

Figure 6.7: Number of Relays

First, let us look at the performance with 6 destinations. Figure 6.7(a) depicts the total number of deployed relays. Network size 4 means 400m*400m. Readers can deduce the meaning for the other network sizes in a similar way. The performance of MBC-Flow is consistently better than MBC-Graph, which is better than BC for large network areas on average. The performance of BC is the best when network area is small, i.e., 400m*400m, due to the approximation error in MBC-Flow and MBC-Graph. When network size becomes larger, the approximation error is hidden by the optimization performance; hence MBC-Flow and MBC-Graph outperform BC. As the network area increases, more relays are expected to cover all the destination locations.

Similar results are obtained for 10 destinations, shown in Figure 6.7(b). It is worthwhile noting that the gap between MBC-Flow and MBC-Graph is larger, compared to the result for 6 destinations. The improvement of MBC-Flow over MBC-Graph is more prominent when the network area becomes larger, 1000m*1000m for example. This is because of the accumulated approximation error for single-cover and Steiner forest problems is magnified with more destinations in a larger network area. Similarly, advantage of MBC-Flow and MBC-Graph is more pronounced for large network areas. The number of relays for 10 destinations is greater than that for 6 destinations. On all different settings, MBC-Flow beats MBC-Graph on average.

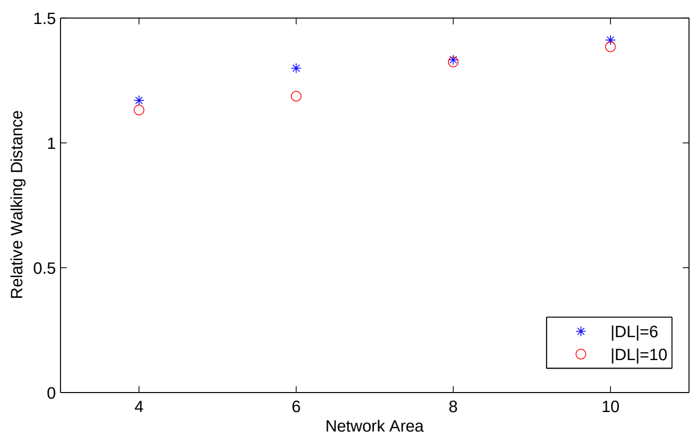


Figure 6.8: Relative Walking Distance of MBC-Flow

In Figure 6.8, we show the relative walking distance of MBC-Flow. The worst ratio is 1.41 for a network area of 1000m*1000m with 6 destinations. Value 1.41 means that on average, FRs have to travel approximately 41% more distance for mission-aware relay deployment algorithms. This value is generally acceptable, when relays are the most critical resources with a limited number of supplies. Though the total cost for communication infrastructure is dramatically reduced in MBC-Flow compared with BC, FRs do not need to walk too much extra distance, regarding deploying relays at the specific locations and walking under the coverage of the existing communication infrastructure.

6.4.2 Number of Mobile Groups

In this section, we study the effect of the number of mobile groups. Each mobile group visits 8 locations, in a network area 900m*900m. We vary the number of mobile groups from 1 to 4. Figure 6.9 draws the number of deployed relays for three algorithms. As the total number of mobile groups increases, more relays are deployed since their destinations are more scattered around the network area. Both MBC-Flow and MBC-Graph keep the increasing rate for the number of relays relatively low, thanks to the optimized relay reuse within and across mobile groups. The extra overhead regarding relative walking distance is shown in Figure 6.10, which falls upon an acceptable range again.

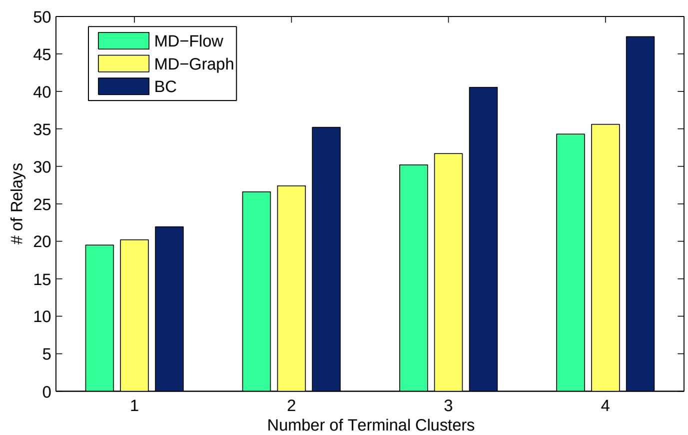


Figure 6.9: Number of Relays

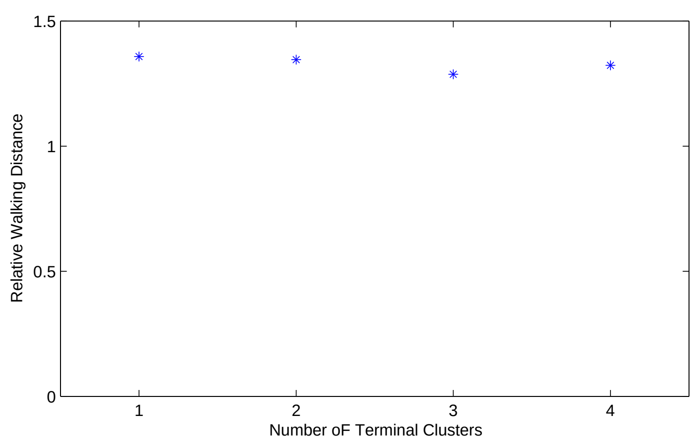


Figure 6.10: Relative Walking Distance

6.4.3 Granularity of Relay Placement Scheme

Both MBC-Flow and MBC-Graph are constrained relay placement algorithms. Their performance is correlated with the granularity of relay placement scheme. Now, let us investigate in which degree the granularity of relay placement scheme will impact the performance of two proposed algorithms for persistent BS-connectivity. We vary network areas, while keeping the destination density at the same level, at approximately 1 destination per $40,000m^2$. The mapping between network sizes and the number of destinations is shown in Table 6.1. 2 mobile groups are dispatched.

Table 6.1: Network Setup

DDN configuration	Network Area	Number of Destinations
1	400m*400m	4
2	600m*600m	9
3	800m*800m	16

The number of relays under different DDN configurations is shown in Figure 6.11. txR means that the grid size is the same as the transmission range txR, while $txR/2$ means it is half of txR. It is interesting to see with denser relay placement scheme (smaller grid size), more relays are used in MBC-Graph. We suspect that the single-cover algorithm may calculate some covers, whose pairwise distance is not exactly multiples of transmission range, thus incurring more relays. Certain local adjustment algorithms might help in this case. For MBC-Flow, denser relay placement scheme excels only with large network areas, DDN configuration 3 for example. In a summary, it is acceptable and reasonable to use a relay placement scheme with grid size equal to the transmission range, due to both its simplicity and preservation of optimization performance.

6.4.4 Grid Granularity in Trajectory Planning

Relative walking distance of M-Breadcrumb depends on the grid granularity of trajectory planning. Denser is the grid, more accurately does the trajectory planning algorithm approximate the shortest path from a location to another.

In Figure 6.12, we verify the relative walking distance of MBC-Flow for 3 mobile groups with 8 destinations each. We change the network area from 4, 6, 8, to 10. As expected, as the grid granularity becomes more compressed, the walking distance is closer to the shortest distance in straight line (at value 1). As the network area increases, the relative walking distance increases as well. This is because the number of deployed relays is dramatically reduced, and FRs may need to walk further away from the optimal path

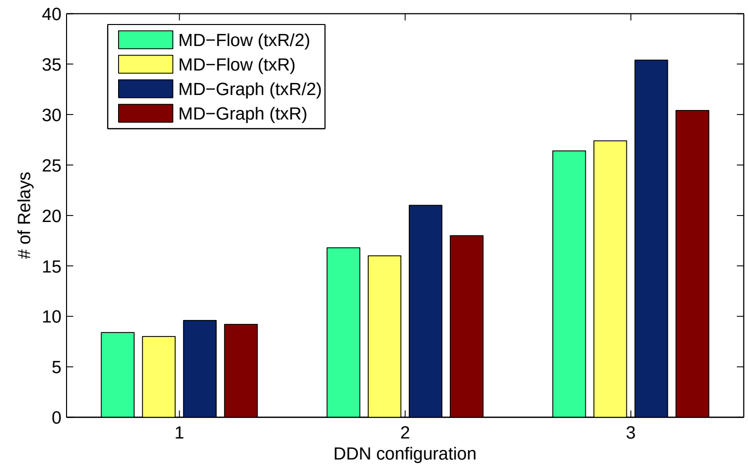


Figure 6.11: Number of Deployed Relays

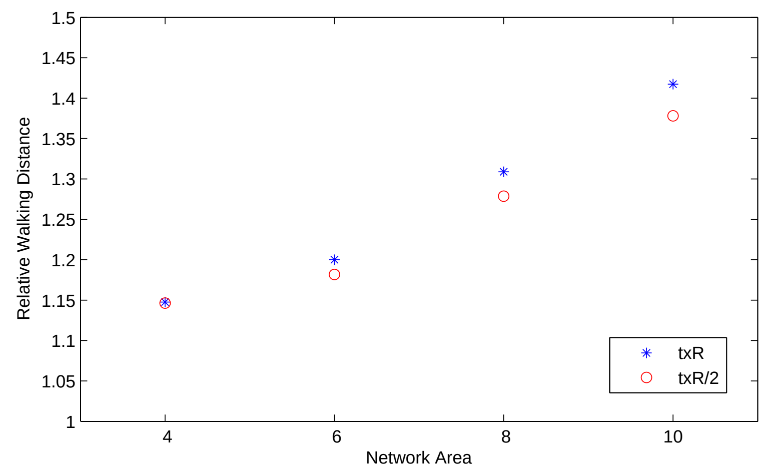


Figure 6.12: Relative Walking Distance of MBC-Flow

towards the relay deploying locations.

Chapter 7

Related Work

Recently, relay management problems have drawn great attention in different research areas, such as improving meeting opportunities in vehicular delay-tolerant networks [36], and maximizing path diversity in intra-domain routing [37]. In this chapter, we will discuss relay management problems from three aspects: relay placement problems to maintain connectivity, energy-aware relay placement problems, and relay deployment problems.

7.1 Relay Placement Problems in DPNs

Almost all related work address the relay placement problems for monomorphous networks, which are then divided into constrained and unconstrained relay placement problems. We summarize three popular approaches here.

The first and most popular approach is *Steinerization*, which is first proposed for Steiner tree with minimum number of Steiner points and bounded edge-length [38]. Steinerization method works in general with two steps: (a) construct the subgraph $\mathcal{G}$ with vertex-connectivity requirement from the completed weighted graph among terminals and base stations (weight is proportional to the distance); and (b) break the long edges of the subgraph $\mathcal{G}$ by inserting relays, duplicated for reliability. [4][3] prove the performance bound of Steinerization approach for 1-connectivity. Later work add many features in Steinerization, such as single-tiered or two-tiered networks, multiple base stations, vertex-connectivity (connectivity, survivability or reliability), homogeneous or heterogeneous links, and full or partial connectivity (one-way or two-way path) [7][8][6]. In two-tiered relay placement [39][7], relay can receive and forward packets, while terminals do not forward packets they receive. A minimum geometric disk cover over all terminals is identified before Steinerization on the covering relay set, for two-tiered replacement.

The second approach is based on idea of regular placement [5]. The network area is divided into cells. Relays are placed to cover all the terminals inside cells, and then additional relays are placed in special locations, e.g., intersections of cells, to provide connectivity across cells.

In both Steinerization and regular placement methods, relays are placed wherever they are needed. Essentially, they are both for unconstrained relay placement. Regrettably, unconstrained relay placement oversimplifies our environment, ignoring various issues such as radio irregularity and obstacles. First, link is represented by circular transmission range. It is nontrivial to adapt the algorithm to irregular radio propagation, wherein transmission range has different values along different directions [15]. Second, obstacles are not considered. Here, we mean the areas where relays can't be placed, rather than obstacles to signal propagation. Constrained relay placement can deal with both problems by adjusting the set of relay locations and links between vertexes.

The third approach is based on pure graph analysis for constrained relay placement problems. All the terminals, base stations and candidate locations are represented as vertexes in a graph. Well-known min cost vertex-connectivity algorithms can be applied. Misra et all in [14] consider both connectivity and survivability which requires biconnectivity of the sensor nodes and base stations. Authors approximate the node weight by edge weight and then use the minimum Steiner tree approximation algorithm for connectivity and sequential maximum flow based algorithm for survivability. However, the proposed algorithms cannot be directly applied to DPNs due to multiple topologies. Furthermore, connectivity is maintained for every pair of terminals instead of BS connectivity, which is more valuable for monitoring and controlling applications.

Steiner tree problems have been extensively studied with edge weight in ordinary graphs or Euclidean space. Zachariasen et all [40] propose an exact solution for Euclidean Steiner Tree to avoid obstacles, which can be possibly integrated with Steinerization approach to deal with obstacles. The best approximation ratio for edge-weighted Steiner tree is 1.55 [41]; while the best approximation ratio for general node-weighted Steiner tree is $O(\log n)$ [29], wherein n is the number of terminals. Klein et all [42] propose an approximation algorithm for node-weighted Steiner tree by considering spider-decomposition. Polzin et all [43] study several techniques for Steiner tree relaxation. Our flow-based relaxation is one among the other relaxation techniques discussed in [43], such as cut-based and tree-based relaxation.

7.2 Energy-aware Relay Placement Problems in DPNs

Wang et all [10] study the traffic-aware relay node deployment problems, for the simple case of one source node, both with single and multiple traffic flows, wherein the relay placement scheme is unconstrained. Compared with constrained ones, unconstrained ones are more difficult to model for optimization, thus heuristic rules being applied. Thomas et all [44] consider joint energy provisioning and relay node placement problems, e. g. how to allocate a total amount of additional energy at a set of locations for relays, such that the network lifetime is maximized. Authors' analysis assumes that power control is enforced and overheard packets do not consume any energy, which may introduce discrepancy in the actual network lifetime. Another class of related work is for sensor node deployment. Under the random deployment scheme, Olariu et all [45] theoretically exam how to maximize the network lifetime, while avoiding energy holes for sensor networks with high density. Xu et all [46] study several random deployment strategies for relay nodes in a heterogeneous sensor network, which are connectivity-oriented, lifetime-oriented, and hybrid deployment strategies. In lifetime-oriented strategy, the number of relays at different locations should be proportional to the expected energy dissipation rates at those locations. The hybrid deployment approach tries to balance the concern between connectivity and lifetime.

Related work on energy-aware relay placement problems can be also found in other categories, for example energy-aware routing and energy-aware control. Energy-aware relay placement problems are the most relevant to energy-aware routing problems. Usually routing assumes that the set of nodes and their locations are given, while placement needs to determine both the optimal number of nodes and their locations. Xue et all [47] study the energy-efficient routing for data aggregation in wireless sensor networks to maximize the network lifetime. A fast approximation algorithm is presented to solve the multi-commodity flow formulation. In [48], authors suggest a general model for energy-aware routing to analyze and evaluate various strategies, such as variable-range transmission power control with optimal traffic distribution, mobile-data-sink deployment, multiple-data-sink deployment, nonuniform initial energy assignment, and intelligent sensor/relay deployment. In addition, the factor of extra costs involved in more complex deployment strategies are studied. Energy-aware control is exploited in both topology control and transmission control. Niewiadomska-Szynkiewicz et all [49] propose two energy-efficient topology control protocols based on two sleep and wakeup schemes, synchronous and topology-based power control. In synchronous power control, nodes periodically wake up to exchange data packets. In topology-based power control, a subset of nodes always stay awake, topologically covering the whole network area. Under the context of FR systems, the second control scheme is more appropriate, when relays stay awake all the time, notifying standby terminals for packet reception. Zimmerling et all [50] discuss a novel approach for energy-efficient data routing to a single control center in a linear sensor topology. The target scenarios are suitable for the cases where sensors are deployed along bridges or pipelines.

Finally, we list some interesting papers, which resemble the ideas developed in this thesis. Bouabdallah et all [51] show that distributing the traffic generated by each sensor node through multiple paths allows significant energy savings, compared with distributing the traffic over a single path. The paper presents systematic analysis of energy consumption without transmission control, a.k.a fixed transmission power and data rate. Our multi-path routing for each terminal coincide with the authors' finding. Tokgoz et all [52] propose to improve the reliability and power consumption of a wireless ad hoc network by deploying additional agents with less stringent power constraints than regular mobile nodes. Results show that agents help regular mobile nodes reduce their power consumption, especially when agent-aware routing protocols are used. InTTCIs-TC, relays in relay backbone can be viewed as the special agents. We believe that by provisioning relays with high-capacity batteries can further decrease the number of deployed relays in ERP. It is a tradeoff between battery capacity and the dimensional size of relays, thus portability.

7.3 Relay Deployment Problems in DDNs

There are many outstanding results for the relay deployment problems, particularly on system and implementation aspects. With deployment problems first studied in [11], simple but effective approach called "breadcrumb" is proposed and used in the testbed evaluation of droppable relay system. All the later work follow the "breadcrumb" approach, where movement are not amendable. Relays are dropped based on various link quality measurements, reliability requirements, etc. In successive work [12], the authors take more extensive measurement from the improved and flexible testbed platform of IEEE 802.11-based radios, utilizing bi-directional SNR. It reveals some areas of improvement, such as route transitions. Liu et all [13] study more realistic and practical system design for relay system and propose solution regarding four aspects: redundancy degree optimization, decision support system, height effect solver, and adaptive power control to maintain link quality over time.

Another body of related work is based on simulation, which helps relay planning and reveals latent problems for large-scale relay communication infrastructures. Wolff et all [53] present different heuristics to deploy relays for reduced interference, such as every 20 meters, at each door or edge of a building, when signal strength falls below some threshold. Realistic homeland security scenarios are analyzed in a dedicated simulation environment, which is able to provide a reliable prediction of the relay number for planning stage. Refaei et all [54] study the impact of interference on reliability of relay communication infrastructures. It is found that the benefit from relays is diminished, in presence of interference. Authors propose cognitive radios to detect interference, identify vacant channels and adapt channel selection.

Unlike relay deployment problems, relay placement problems aim for static networks, where locations of static terminals and BSs are given. There are many outstanding work for the optimal relay placement problems. However, none of them is able to maintain persistent BS-connectivity for mobile terminals.

Another area of research which might be helpful to relay deployment problems is the distributed coordination control of multi-agents to preserve connectivity [55][56][57][58][59], i.e., a group of mobile agents need to stay connected, while achieving certain performance objective. Two fundamental problems are agreement problem and control problem. Agreement problem is concerned with decentralized strategies for convergence and control problem is concerned with the movement controller design. Techniques, such as potential field, controller and distributed auction, can be used in designing mobile relays, which is outside the scope of this thesis.

Chapter 8

Conclusion

8.1 Summary of Dissertation

In this thesis, we study the relay management algorithms to establish on-demand communication infrastructure for BS-connectivity, while meeting various performance requirements.

To begin with, we investigate the constrained relay placement problems to maintain reliable base station communication for disconnected polymorphous networks with multiple terminal topologies. We propose three algorithms. Two heuristic algorithms, topology stitch algorithm and topology iterative algorithm are presented, which are built upon existing constrained relay placement algorithms for monomorphous networks. Also, we propose the Weigh-and-Place algorithm (WPA), which optimizes relay placement with balanced load at BSs across multiple topologies based on integer programming formulation. Our evaluation shows that WPA has better performance than the other heuristic algorithms. In addition, connectivity provided by multiple BSs can be uniformly utilized by enforcing stress constraints in our formulation.

Second, we study the energy-aware constrained relay placement problems to maintain BS-connectivity, while maintaining the expected network lifetime above an acceptable value for DPNs. Our proposed energy-aware relay placement algorithm is able to meet all the lifetime requirements when the battery provision is critical, with respect to the traffic rate and the expected network lifetime. Our formulation is flexible to plug in both power control and data rate control to the energy consumption model. Also we observe interesting results that rate control is more effective than power control to reduce the total battery consumption.

Finally, we research the mission-aware constrained relay deployment problems to maintain persistent BS-connectivity, when FRs are continuously moving around the incident scene. Compared with the state-of-art relay deployment algorithms of “breadcrumb”, we exploit the mission knowledge to reduce the total number of relays by maximally reusing relays within and across mobile groups. Results show that our optimization dramatically reduce the number of relays, at tolerable cost of extra walking distance, compared with the shortest path.

8.2 Discussion

Our multi-commodity formulation can be easily adapted to include installation cost or risk for each candidate relay location. Right now, each relay location receives a unit cost in the objective function. However, we can differentiate locations by assigning to them different weights w_l and transforming the objective to weighted summation of decision variables y_l. One way is to assign weight based on the risk. A location, where it is more likely for a relay to malfunction, is assigned a higher weight, for example locations near fire with high temperature or damp areas with high humidity level. By minimizing the weighted summation, we minimize the total risk of the communication infrastructure, thus improving its reliability.

Another advantage of our formulation is that we can easily capture obstacles. Obstacles may affect relay management in three aspects: candidate locations, radio propagation and moving trajectory. If some obstacle prevents FRs from entering and dropping relays, the affected locations are excluded from the set of candidate locations $\mathcal{L}$. If some obstacle blocks the radio propagation, thus packet transmission, between two relay locations, the link between two relay locations is excluded from the communication graph. The last one is mainly for relay deployment problems. For online navigation, we suppose that FRs can move between adjacent covered grid vertexes and task destinations in straight line. If the obstacle happens to intersect with the straight line, two locations are no longer adjacent. In this way, we can avoid complicated geometric analysis. Moreover, adding obstacles not only makes our model more realistic, but also reduces the problem size by eliminating many candidate locations and links from consideration.

Finally, we want to debate over the relationship between DPNs and DDNs. DPNs are used to model polymorphous networks with multiple terminal topologies, while DDNs are used to model mobile networks. Mission topologies from DPNs may look the same as terminal topologies from DDNs regarding the underlying models. However a set of mission topologies corresponds to an infinite number of terminal topologies, which can be deduced based on all kinds of moving trajectory of different groups.

8.3 Future Work

In this thesis, we make our first attempt to address the relay management problems for disconnected evolving networks of First Responder systems. The most important take-away message is that relay management problems should be considered together with the special properties of FR systems, such as its topology dynamics and coarse-level predictability. These properties should not only be utilized in plotting evaluation scenarios, but also be exploited in devising algorithms and protocols.

We recognize that the solutions presented here are not complete and there are still many improvement dimensions to explore. First, we need to make our approximate algorithm scalable to a large number of terminals and terminal topologies, dense relay placement scheme, and dense movement grid for trajectory planning. Second, we can enforce resource isolation among different domains or organizations. For now, we assume that BSs and relays are the communication infrastructure shared by all the terminals and a single command center. Imaging in the incident scene, there are usually a chain of commands and a command center per organization is setup for better resource management and control. All terminals from an organization may only communicate with their organizational CC through the organizational BSs. However, relays are still shared across organizational boundaries. We can capture this organization boundaries by replacing CC by multiple organizational CCs and replacing the one-to-one link between CC and every BS by the one-to-one links between the organizational CC and the associated BSs. A terminal will send commodities to its organizational CC. Third, we can add k-BS-connectivity to relay deployment algorithms to tolerate backbone relay failure. The added reliability relies on no information of instantaneous terminal topologies, so that different mobile groups can work independently. Finally, we can introduce network coding to further optimize the upload traffic in energy-aware relay placement [60]. Suppose every terminal transmits traffic towards a single CC with rate control, different packet delivery ratios are observed by a set of overhearing neighbors. Because of independent packet error, several neighbors may overhear the packet successfully. With opportunistic packet delivery, like anycast, all the neighbors receiving the right packets can combine the packets with other data and send them toward CC now or in the future. In this way, we waste no energy in overhearing correct packets and save the total energy.

References

[1] Fact sheet: Achieving first responder communications interoperability: a local, state, and federal partnership. U.S. Department of Homeland Security, Sep 2004.

[2] Incident management successful practices. U.S. Department of Transporatation. `http://www.itsdocs.fhwa.dot.gov//JPODOCS/REPTSTE/8V001!.PDF`.

[3] Xiuzhen Cheng, Ding zhu Du, Lusheng Wang, and Baogang Xu. Relay sensor placement in wireless sensor networks. *IEEE Transactions on Computers*, 2001.

[4] Donghui Chen, Ding-Zhu Du, Xiao-Dong Hu, Guo-Hui Lin, Lusheng Wang, and Guoliang Xue. Approximations for steiner trees with minimum number of steiner points. *Journal of Global Optimization*, 2000.

[5] Jian Tang, Bin Hao, and Arunabha Sen. Relay node placement in large scale wireless sensor networks. *Computer Communications*, 2006.

[6] Xiaofeng Han, Xiang Cao, E.L. Lloyd, and Chien-Chung Shen. Fault-tolerant relay node placement in heterogeneous wireless sensor networks. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications.*, May 2007.

[7] Weiyi Zhang, Guoliang Xue, and S. Misra. Fault-tolerant relay node placement in wireless sensor networks: Problems and algorithms. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications.*, May 2007.

[8] Jonathan L. Bredin, Erik D. Demaine, MohammadTaghi Hajiaghayi, and Daniela Rus. Deploying sensor networks with guaranteed capacity and fault tolerance. In *MobiHoc '05: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, 2005.

[9] Dejun Yang, Satyajayant Misra, Xi Fang, and Junshan Zhang. Two-tiered constrained relay node placement in wireless sensor networks: Efficient approximations. In *SECON'10: Proceedings of the 7th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2010.

[10] Feng Wang, Dan Wang, and Jiangchuan Liu. Traffic-aware relay node deployment for data collection in wireless sensor networks. In *SECON'09: Proceedings of the 6th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pages 351–359, Piscataway, NJ, USA, 2009. IEEE Press.

[11] Michael R. Souryal, Johannes Geissbuehler, Leonard E. Miller, and Nader Moayeri. Real-time deployment of multihop relays for range extension. In *MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 85–98, New York, NY, USA, 2007. ACM.

[12] Michael R. Souryal, Andreas Wapf, and Nader Moayeri. Rapidly-deployable mesh network testbed. In *GLOBECOM'09: Proceedings of the 28th IEEE conference on Global telecommunications*, pages 5536–5541, Piscataway, NJ, USA, 2009. IEEE Press.

[13] Hengchang Liu, Jingyuan Li, Zhiheng Xie, Shan Lin, Kamin Whitehouse, John A. Stankovic, and David Siu. Automatic and robust breadcrumb system deployment for indoor firefighter applications. In *MobiSys '10: Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 21–34, New York, NY, USA, 2010. ACM.

[14] Satyajayant Misra, Seung Don Hong, Guoliang Xue, , and Jian Tang. Constrained relay node placement in wireless sensor networks to meet connectivity and survivability requirements. In *Infocom'2008. 27th IEEE International Conference on Computer Communications.*, 2008.

[15] Gang Zhou, Tian He, Sudha Krishnamurthy, and John A. Stankovic. Impact of radio irregularity on wireless sensor networks. In *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, 2004.

[16] National response plan. U.S. Department of Homeland Security. `http://www.dhs.gov/xlibrary/assets/NRPBrochure.pdf`.

[17] Ying Huang, Wenbo He, Klara Nahrstedt, and Whay C. Lee. Requirements and system architecture design consideration for first responder systems. In *2007 IEEE Conference on Technologies for Homeland Security: Enhancing Critical Infrastructure*, 2007.

[18] Safecom statement of requirements (sor). U.S. Department of Homeland Security, 2006. `http://www.safecomprogram.gov/NR/rdonlyres/8930E37C-C672-48BA-8C1B-8378%4D855C1E/0/SoR1v1210182006.pdf.`

[19] Fan Bai, Narayanan Sadagopan, and Ahmed Helmy. Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. In *INFOCOM 2003. 22th IEEE International Conference on Computer Communications.*, 2003.

[20] Ming Zhao and Wenye Wang. Design and applications of a smooth mobility model for mobile ad hoc networks. In *MILCOM 2006. Military Communication Conference*. IEEE, 2006.

[21] Xiaoyan Hong, Mario Gerla, Guangyu Pei, and Ching-Chuan Chiang. A group mobility model for ad hoc wireless networks. In *MSWiM '99: Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, pages 53–60, New York, NY, USA, 1999. ACM.

[22] Amit Jardosh, Elizabeth M. Belding-Royer, Kevin C. Almeroth, and Subhash Suri. Towards realistic mobility models for mobile ad hoc networks. In *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 217–229, New York, NY, USA, 2003. ACM.

[23] C. Tuduce and T. Gross. A mobility model based on wlan traces and its validation. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 1, pages 664–674 vol. 1, 2005.

[24] Ying Huang, Wenbo He, Klara Nahrstedt, and Whay C. Lee. Incident scene mobility analysis. In *2008 IEEE Conference on Technologies for Homeland Security: Enhancing Critical Infrastructure*, 2008.

[25] Ying Huang, Wenbo He, Klara Nahrstedt, and Whay C. Lee. Corps: Event-driven incident mobility model for first responders. *Technical report of Computer Science Department in UIUC*, 2009.

[26] Ying Huang, Wenbo He, K. Nahrstedt, and W.C. Lee. Corps: Event-driven mobility model for first responders in incident scene. In *Military Communications Conference, 2008. MILCOM 2008.*, pages 1–7, Nov. 2008.

[27] Samuel C. Nelson, Albert F. Harris, and Robin H. Kravets. Event-driven, role-based mobility in disaster recovery networks. In *CHANTS '07: Proceedings of the second ACM workshop on Challenged networks*, page 27–34. ACM, ACM, 2007.

[28] David R. Choffnes and Fabián E. Bustamante. An integrated mobility and traffic model for vehicular wireless networks. In *VANET '05: Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*, pages 69–78, New York, NY, USA, 2005. ACM.

[29] Sudipto Guha and Samir Khuller. Improved methods for approximating node weighted steiner trees and connected dominating sets. *Information and Computation*, 1999.

[30] Richard T. Wong. A dual ascent approach for steiner tree problems on a directed graph. *Mathematical Programming*, 1984.

[31] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, February 1997.

[32] Joseph Cheriyan, Santosh Vempala, and Adrian Vetta. Network design via iterative rounding of setpair relaxations. *Combinatorica*, 2006.

[33] Julia Chuzhoy and Sanjeev Khanna. Algorithms for single-source vertex connectivity. In *FOCS '08: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008.

[34] dei80211mr: a new 802.11 implementation for ns-2. Website, 2008. `http://www.dei.unipd.it/wdyn/?IDsezione=5090`.

[35] Vijay V. Vazirani. *Approximation Algorithms.* Springer, July 2001.

[36] F. Farahmand, I. Cerutti, A.N. Patel, Q. Zhang, and J.P. Jue. Relay node placement in vehicular delay-tolerant networks. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1 –5, nov. 2008.

[37] Meeyoung Cha, Sue Moon, Chong dae Park, and Aman Shaikh. Placing relay nodes for intra-domain path diversity. Technical report, in Proc. IEEE INFOCOM, 2005.

[38] Guo-Hui Lin and Guoliang Xue. Steiner tree problem with minimum number of steiner points and bounded edge-length. *Information Processing Letters*, 1999.

[39] Errol L. Lloyd and Guoliang Xue. Relay node placement in wireless sensor networks. *IEEE/ACM Transactions on Networking*, 2007.

[40] Martin Zachariasen and Pawel Winter. Obstacle-avoiding euclidean steiner trees in the plane: An exact algorithm. *Lecture Notes in Computer Science*, 1999.

[41] Gabriel Robins and Alexander Zelikovsky. Tighter bounds for graph steiner tree approximation. *SIAM Journal on Discrete Mathematics*, 2005.

[42] Philip Klein and R. Ravi. A nearly best-possible approximation algorithm for node-weighted steiner trees. *Journal of Algorithms*, 1995.

[43] Tobias Polzin and Siavash Vahdati Daneshmand. A comparison of steiner tree relaxations. *Discrete Applied Mathematics*, 2001.

[44] Y. Thomas, Houy Yi, Shiy Hanif, D. Sheralit, and Scott F. Midkiffy. Prolonging sensor network lifetime with energy provisioning and relay node placement. In *IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, pages 295–304, Santa Clara, CA, USA, September 2005.

[45] Stephan Olariu and Ivan Stojmenovi? Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting. In *IEEE INFOCOM*, pages 1–12. Society Press, 2006.

[46] Kenan Xu, Hossam Hassanein, Glen Takahara, and Quanhong Wang. Relay node deployment strategies in heterogeneous wireless sensor networks. *IEEE Transactions on Mobile Computing*, 9:145–159, 2010.

[47] Yuan Xue, Yi Cui, and Klara Nahrstedt. Maximizing lifetime for data aggregation in wireless sensor networks. *Mob. Netw. Appl.*, 10(6):853–864, 2005.

[48] Zhao Cheng, Mark Perillo, and Wendi B. Heinzelman. General network lifetime and cost models for evaluating sensor network deployment strategies. *IEEE Transactions on Mobile Computing*, 7:484–497, 2008.

[49] Ewa Niewiadomska-Szynkiewicz, Piotr Kwasniewski, and Izabela Windyga. Comparative study of wireless sensor networks energy-efficient topologies and power save protocols, march 2009.

[50] Marco Zimmerling, Waltenegus Dargie, and Johnathan M. Reason. Energy-efficient routing in linear wireless sensor networks, oct. 2007.

[51] F. Bouabdallah, N. Bouabdallah, and R. Boutaba. Load-balanced routing scheme for energy-efficient wireless sensor networks. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1 –6, nov. 2008.

[52] Y. Tokgoz and A. Acampora. Improving connectivity and power efficiency in wireless ad hoc networks through agent nodes. In *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005.*, pages 751–756. IEEE, 2005.

[53] Sebastian Subik Andreas Wolff and Christian Wietfeld. Performance analysis of highly available ad hoc surveillance networks based on dropped units. In *2008 IEEE Conference on Technologies for Homeland Security: Enhancing Critical Infrastructure*, 2008.

[54] M.T. Refaei, M.R. Souryal, and N. Moayeri. Interference avoidance in rapidly deployed wireless ad hoc incident area networks. In *INFOCOM Workshops 2008, IEEE*, pages 1 –6, 13-18 2008.

[55] M. Ani Hsieh, Anthony Cowley, Vijay Kumar, and Camillo J. Taylor. Maintaining network connectivity and performance in robot teams: Research articles. *J. Field Robot.*, 25(1-2):111–131, 2008.

[56] M Bernardine Dias, Robert Michael Zlot, Nidhi Kalra, and Anthony (Tony) Stentz. Market-based multirobot coordination: A survey and analysis. Technical Report CMU-RI-TR-05-13, Robotics Institute, Pittsburgh, PA, April 2005.

[57] Michael M. Zavlanos and George J. Pappas. Distributed connectivity control of mobile networks. In *Proceedings of the 46th IEEE Conference on Decision and Control*, Dec. 2007.

[58] Ji and M. Egerstedt. Distributed coordination control of multiagent systems while preserving connectedness. *Robotics, IEEE Transactions on [see also Robotics and Automation, IEEE Transactions on]*, 23(4):693–703, 2007.

[59] Nathan Michael, Michael M Zavlanos, Vijay Kumar, and George J Pappas. Maintaining connectivity in mobile robot networks. 2009.

[60] Sachin Katti, Hariharan Rahul, Wenjun Hu, Dina Katabi, Muriel Médard, and Jon Crowcroft. Xors in the air: practical wireless network coding. *SIGCOMM Comput. Commun. Rev.*, 36(4):243–254, 2006.

Author's Biography

Ying was born and grew up in Wuhan, China. She received her Bachelor of Science degree from Department of Electronics Engineering and Computer Science at Peking University, Beijing, China in 2004. In August 2004, she was admitted to Department of Computer Science at University of Illinois at Urbana-Champaign (UIUC). Ying received her Master of Science degree in Computer Science in 2008. Since summer 2005, Ying has worked in the MONET (Multimedia Operating System and Networking) research group led by Professor Klara Nahrstedt, while pursuing her Ph.D. degree. Her research interests include vehicular networks, proximity social networks, mobile wireless networks, smartphone applications, distributed systems, content sharing and data dissemination, Quality of Service (QoS), topology control, mobility model, and security. Ying hold two internships at Broadband Network Research Lab, Motorola, in summer 2005 and 2006, respectively. She is a recipient of Qualcomm "Q" Award of Excellence Scholarship in 2008. She received both IEEE MASS'09 Travel Grant Award and CS Department Conference Travel Grant Award from UIUC in 2009. After graduation, Ying joined Google in Kirkland, WA.