1-1-2015

Energy Cost Optimization for Strongly Stable Multi-Hop Green Cellular Networks

Weixian Liao

Energy cost optimization for strongly stable multi-hop green cellular networks

By

Weixian Liao

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Electrical and Computer Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2015

Energy cost optimization for strongly stable multi-hop green cellular networks

By

Weixian Liao

Approved:

Qian (Jenny) Du
(Major Professor)

James E. Fowler
(Committee Member/Graduate Coordinator)

Pan Li
(Committee Member)

Jason M. Keith
Dean
Bagley College of Engineering

Name: Weixian Liao

Date of Degree: December 11, 2015

Institution: Mississippi State University

Major Field: Electrical and Computer Engineering

Major Professor: Prof. Qian (Jenny) Du

Title of Study: Energy cost optimization for strongly stable multi-hop green cellular networks

Pages of Study: 47

Candidate for Degree of Master of Science

Last decade witnessed the explosive growth in mobile devices and their traffic demand, and hence the significant increase in the energy cost of the cellular service providers. One major component of energy expenditure comes from the operation of base stations. How to reduce energy cost of base stations while satisfying users' soaring demands has become an imperative yet challenging problem. In this dissertation, we investigate the minimization of the long-term time-averaged expected energy cost while guaranteeing network strong stability. Specifically, considering flow routing, link scheduling, and energy constraints, we formulate a time-coupling stochastic Mixed-Integer Non-Linear Programming (MINLP) problem, which is prohibitively expensive to solve. We reformulate the problem by employing Lyapunov optimization theory and develop a decomposition based algorithm which ensures network strong stability. We obtain the bounds on the optimal result of the original problem and demonstrate the tightness of the bounds and the efficacy of the proposed scheme.

DEDICATION

To my family.

ACKNOWLEDGEMENTS

First of all, I would like to thank Dr. Jenny Q. Du and Dr. Pan Li, my master thesis advisors, for their professional guidance and spiritual encouragement through my journey of pursuing the Master's Degree. The unforgettable research experience at MSU is my invaluable lifelong asset that will continuously light up my future life.

At the same time, I would like to thank my other committee member, Dr. James E. Fowler, for serving on my supervisory committee and for his sincere help in various stages of my graduate study.

Last but not the least, I would like to extend my grateful thanks to all my colleagues in our NEST group for providing me with a warm, family-like environment, for their encouragement and collaboration and for countless valuable discussions and all the good memories.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

In the last few years, with the proliferation of smart phones, tablets, etc., we have witnessed tremendous growth in the number of cellular subscribers and in their traffic demand [10]. In parallel with the rapidly growing demand for cellular services, the number of cellular base stations (BSs) all over the world has increased from a few hundred thousands to more than 4 million, and each of them consumes an average of 25 MWh per year [13]. Studies show that the radio network itself adds up to 80% of an operator's entire energy consumption, which represents a significant portion of a network operator's overall expenditures [5]. Therefore, it is in dire need to find effective solutions to reducing the energy costs of cellular networks while satisfying subscribers' ever-increasing traffic demand.

1.2 Related Work

The rising energy costs of cellular networks have led to both academical and industrial efforts to address the energy efficiency issues and develop the "green cellular networks" [2, 4]. In particular, the energy consumption of a BS can be reduced by improving the BS hardware design, for example, the efficiency of power amplifiers (PAs) [21]. We can also reduce BSs' energy consumption by including additional software and system features to balance between energy consumption and network performance, e.g., shutting down BSs during low traffic hours or cell zooming [26–28, 30]. In particular, Niu *et al.* [27] propose algorithms for cell zooming to avoid coverage holes when BSs are turned off. Niu [26] also studies cell deployment when cell zooming is not sufficient. Oh *et al.* [28] propose to switch BSs off by considering a newly introduced notion of network-impact. Peng *et al.* [30] propose to turn underutilized BSs off when traffic is low and turn them on when traffic is high. However, such system level approaches may degrade the cellular network performance and some cellular users can get disconnected.

Beyond the advance of BS development and control itself, it is crucial to consider various paradigm-shifting technologies, such as multi-hop relaying and renewable energy integration, in order to enhance the energy efficiency of cellular networks. Particularly, multi-hop relaying has been introduced into cellular networks to improve network throughput [9, 18, 20]. In fact, since multi-hop communications divides direct paths between mobile terminals and BSs into shorter links [19], in which wireless channel impairments such as path loss are less destructive, lower transmission power can be assigned to the BSs and relays and hence network energy consumption can also be saved. It has been shown [32] that using multi-hopping in CDMA cellular networks can reduce the average energy consumed per call. In addition, renewable energy integration has attracted intense attention [22]. Sustainable energy resources such as sustainable biofuels, solar and wind energy seem to be promising options to reduce the overall network energy expenditure and the CO_2 footprint since they are significantly cheaper to maintain in the long run. Erission and Nokia [1, 6] have developed a green BS that is based on solar power and wind power, respectively, without using any grid electricity. Han *et al.* [12] try to take advantage of green BSs by maximizing the green energy usage. For subscribers, mobile manufacturers like Samsung and Nokia have released a series of future phones which contain solar panels [3].

In this paper, we investigate how to minimize the energy cost of cellular networks while still satisfying users' traffic demand by considering energy-efficient wireless architectures, renewable energy integration, and network stability. Specifically, we consider a multi-hop cellular network consisting of a number of cellular users, a group of base stations, and a set of available spectrum bands. We envision that each node is equipped with a renewable energy resource, for example, a solar panel (e.g., for each mobile user) or a wind turbine (e.g., for each base station), as well as an energy storage unit [1,3,6]. Both spectrum bandwidths and renewable energy resource outputs are modeled as random processes. In this network, mobile users may communicate with each other or with base stations via multiple hops, rather than a single hop as in traditional cellular networks. Thus, the communications can take advantage of locally available spectrums and link rate adaptivity, and hence provide much higher network capacity.

We first formulate an offline energy cost minimization problem, by jointly exploring renewable energy resource allocation, routing, and link scheduling, which turns out to be a time-coupling stochastic Mixed-Integer Non-Linear Programming (MINLP) problem. Previous approaches usually solve such problems based on Dynamic Programming and suffer from the "curse of dimensionality" problem [7]. Full statistical information of the random variables is required to solve the problem, which may be difficult to obtain in practice. Therefore, we reformulate the problem by employing Lyapunov optimization theory [24] and propose an online finite-queue-aware energy cost minimization problem. In the literature, Lyapunov optimization techniques have been adopted to investigate optimization problems in wireless networks [16, 17, 23–25, 33, 34]. Unfortunately, [17, 34] cannot guarantee that all queues are finite. [16, 23, 25] develop opportunistic scheduling schemes, which maintain finite queue sizes by dropping some packets. [24, 33] propose joint stability and utility optimization algorithms, but assume that the users' input data rate is interior to the network capacity region. Thus, in spite of these existing studies, none of the developed algorithms can be adopted to solve our problem, nor to keep all queues finite.

Considering that the previously formulated online finite-queue-aware energy cost minimization problem is an MINLP problem, which is in general NP-hard [31] and needs to be solved in each time slot, we reformulate it and propose an approximation algorithm to solve it efficiently. Specifically, by introducing virtual queues, we are able to decompose the reformulated problem into four subproblems: link scheduling, resource allocation, routing, and energy management. We develop three algorithms to solve the first three subproblems, respectively, based on current network states only. After the first three subproblems are solved, the fourth subproblem can be easily solved as well. We prove that the proposed decomposition based approximation algorithm guarantees that all queues in the network are finite, i.e., network strong stability. Moreover, while the approximation algorithm leads to an upper bound on the optimal result of the original problem, a lower bound is also found by solving a relaxed online Linear Programming (LP) problem.

The main contributions of this paper are briefly summarized as follows:

- We formulate an offline energy cost minimization problem by considering dynamic spectrum and renewable energy resource availability, routing, link scheduling, and energy resource allocation.
- We formulate an online finite-queue-aware energy cost minimization problem and propose a decomposition based algorithm to solve the problem efficiently while guaranteeing the strong stability of all queues in the network, i.e., network strong stability.
- We obtain and prove the lower and upper bounds on the optimal result of the original offline energy cost minimization problem.
- Simulation results demonstrate that the obtained lower and upper bounds are very tight, and that the proposed scheme results in noticeable energy cost savings.

CHAPTER 2

SYSTEM MODELS

2.1 Network Model

As shown in Fig. 2.1, we consider a multi-hop cellular network that consists of $\mathcal{U} = \{1, 2, \cdots, u, \cdots, U\}$ users and $\mathcal{B} = \{1, 2, \cdots, b, \cdots, B\}$ base stations. Let $\mathcal{N} = \mathcal{U} \cup \mathcal{B}$. We denote the set of available spectrum bands by $\mathcal{M} = \{1, 2, \cdots, m, \cdots, M\}$, and assume that the bandwidth of band m is a random process denoted by $\{W^m(t)\}_{t=0}^{\infty}$ which can be observed at the beginning of each time slot. In addition, due to their different geographical locations, different nodes may have different available spectrum bands. Denote by $\mathcal{M}_i \subseteq \mathcal{M}$ the set of available spectrum bands that node $i \in \mathcal{N}$ can access. Thus, it is possible that $\mathcal{M}_i \neq \mathcal{M}_j$ for $i \neq j$, $i, j \in \mathcal{N}$. Assume the system operates in a time-slotted manner. Suppose there are a set of downlink Internet service sessions denoted by $\mathcal{S} = \{1, 2, \cdots, s, \cdots, S\}$, each of which is denoted as a tuple $\{d_s, v_s(t), s_s(t)\}$ where d_s stands for the destination of service session s, $v_s(t)$ is the required throughput (in terms of the number of packets) in time slot t, and $s_s(t)$ stands for the source base station of service session s in time slot t.

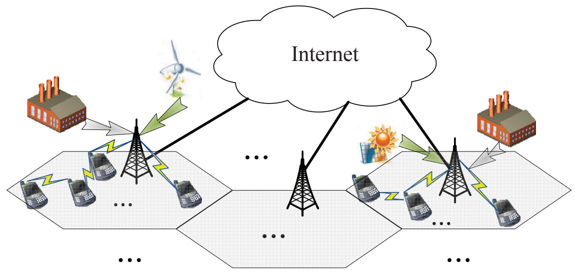

Figure 2.1

System architecture for green multi-hop cellular networks

2.2 Link Capacity

A widely used model [15, 29] employed for power propagation gain between node i and j, denoted by g_{ij}, is $g_{ij} = C \cdot [d(i,j)]^{-\gamma}$, where i and j denote their locations, $d(i,j)$ is the Euclidean distance between i and j, γ refers to the path loss exponent, and C is a constant related to the antenna profiles of the transmitter and the receiver, wavelength, etc.

We adopt the Physical Model [11,15] as the interference model, i.e., a data transmission is successful only if the received signal-to-interference plus noise ratio (SINR) is no less than a threshold Γ. Specifically, if node i sends data to node j on band m in time slot t, the capacity can be calculated as

$$c_{ij}^m(t) = \begin{cases} W^m(t)\log_2(1+\Gamma), \text{ if } SINR_{ij}^m(t) \geq \Gamma \\ 0, \text{ otherwise.} \end{cases} \tag{2.1}$$

where $SINR_{ij}^m(t)$, the SINR of the signal sent from i to j on band m in time slot t, is

$$SINR_{ij}^m(t) = \frac{g_{ij}P_{ij}^m(t)}{\eta_j W^m(t) + \sum_{j\in\mathcal{N}, j\neq i} g_{kj}P_{kv}^m(t)} \tag{2.2}$$

Here, η_j is the thermal noise power density at the receiver j, $P_{ij}^m(t)$ is the transmission power of node i to node j on band m in time slot t, and $P_{kv}^m(t)$ is the transmission power of an interfering node k to its receiver v on band m in time slot t. We also denote the maximum transmission power of node i by P^i_{max}.

2.3 Energy Consumption

For a node i ($i \in \mathcal{N}$), its consumed energy in time slot t, denoted by $E_i(t)$, is attributed to the energy needed to feed the antenna denoted by E_i^{const}, the energy consumed when staying in idle mode denoted by E_i^{idle}, and the energy for serving the traffic $E_i^{TX}(t)$, i.e. [8],

$$E_i(t) = E_i^{const} + E_i^{idle} + E_i^{TX}(t). \tag{2.3}$$

$E_i^{TX}(t)$ will be introduced later.

2.4 Renewable Energy Generation and Energy Storage

We assume that each node $i \in \mathcal{N}$ has a renewable energy resource, for example, a solar panel (e.g., for each mobile user) or a wind turbine (e.g., for each base station). The output of node i's renewable resource, denoted by $R_i(t)$, is an i.i.d. stochastic process that satisfies $0 \leq R_i(t) \leq R_i^{max}$, where R_i^{max} is the maximum energy output and a constant. This is because the output of a renewable energy resource mainly depends on meteorological conditions and is dynamic.

We also assume that every node i has its own energy storage unit, e.g., a battery, for storing energy obtained from its renewable energy resource or drawn from the power grid, which can be used at later time slots. Thus, node i's renewable resource output $R_i(t)$ can be used to charge the energy storage device or serve i's energy demand, i.e.:

$$R_i(t) = c_i^r(t) + r_i(t), \tag{2.4}$$

where $c_i^r(t)$ and $r_i(t)$ are the energy used for charging node i's energy storage unit and serving node i's current energy demand, respectively.

In addition, notice that node i's energy storage unit works as an energy buffer, whose energy level, denoted by $x_i(t)$, can be modeled as an energy queue, i.e.,

$$x_i(t+1) = x_i(t) + c_i(t) - d_i(t). \tag{2.5}$$

where $d_i(t)$ is the energy discharged from the energy storage unit for serving node i's energy demand, and $c_i(t)$ is the energy charging the energy storage unit, i.e.,

$$c_i(t) = c_i^r(t) + \omega_i(t) c_i^g(t) \tag{2.6}$$

$$\omega_i(t) = \begin{cases} 1, \text{ if } i \in \mathcal{B} \\ \xi_i(t), \text{ if } i \in \mathcal{U} \end{cases} \tag{2.7}$$

where $c_i^g(t)$ is the energy drawn from the power grid and $\omega_i(t)$ indicates whether node i is connected into the power grid in time slot t. Note that base stations are always connected to the grid while mobile terminals are only occasionally connected. Thus, we assume that $\{\xi_i(t)\}_{t=0}^{\infty}$ is an i.i.d. random process where $\xi_i(t) \in \{0, 1\}$.

Due to the fact that serving node i's energy demand $E_i(t)$ by directly using energy from the grid or from the renewable energy resource, is more efficient than by first charging the energy storage unit and then discharging it, we have the following two constraints

$$\mathbf{1}_{d_i(t)>0} + \mathbf{1}_{c_i^r(t)>0} \leq 1 \tag{2.8}$$

$$\mathbf{1}_{d_i(t)>0} + \mathbf{1}_{c_i^g(t)>0} \leq 1 \tag{2.9}$$

where $\mathbf{1}_A$ is an indicator function that is equal to 1 when the event A is true, and zero otherwise. Notice that the above constraints (2.8) and (2.9) will hold whenever the following inequality holds:

$$\mathbf{1}_{c_i(t)>0} + \mathbf{1}_{d_i(t)>0} \leq 1 \tag{2.10}$$

Besides, denote by x_i^{max} the maximum amount of energy that can be stored by node i's energy storage unit. Then, we need

$$0 \leq x_i(t) \leq x_i^{max}. \tag{2.11}$$

Denote by c_i^{max} and d_i^{max} the maximum amount of energy that node i's energy storage unit can be charged with and that can be discharged from node i's energy storage unit during a single time slot, respectively. Thus, we have

$$c_i(t) \leq \min[c_i^{max}, x_i^{max} - x_i(t)] \tag{2.12}$$

$$d_i(t) \leq \min[d_i^{max}, x_i(t)]. \tag{2.13}$$

From (2.12) and (2.13), we get $c_i(t) + d_i(t) \leq x_i^{max} - x_i(t) + x_i(t) = x_i^{max}$, which should hold for any $c_i(t)$ and $d_i(t)$ that satisfy (2.12) and (2.13). Since $c_i(t) \leq c_i^{max}$ and $d_i(t) \leq d_i^{max}$, we also have the following constraint:

$$c_i^{max} + d_i^{max} \leq x_i^{max}. \tag{2.14}$$

2.5 Energy Serving and Generation Cost

Node i's energy demand is satisfied by the energy from the power grid, its local renewable energy resource, and its own energy storage device. Particularly, we have $E_i(t) =$

$\omega_i(t)g_i(t) + r_i(t) + d_i(t)$, where $g_i(t)$ is the amount of energy drawn from the power grid to satisfy user i's energy demand in time slot t.

Besides, the amount of energy that node i draws from the power grid in time slot t, denoted by $p_i(t)$, satisfies

$$0 \leq p_i(t) = \omega_i(t)\big(g_i(t) + c_i^g(t)\big) \leq p_i^{max} \tag{2.15}$$

where p_i^{max} is a constant determined by the physical characteristics of user i's connection to the grid.

Since the energy needed from the power grid for mobile terminals is negligible compared to that required for base stations, the total amount of energy supplied by power grid in time slot t, denoted by $P(t)$, is $P(t) = \sum_{i \in \mathcal{B}} \big(g_i(t) + c_i^g(t)\big)$. Thus, the energy cost of the cellular service provider in time slot t can be calculated as $f(P(t))$, where $f(\cdot)$ is assumed to be a non-negative, non-decreasing, and convex function.

2.6 Definitions

Next, we introduce some definitions and theorems that would be used later in this paper [24].

2.6.1 Definition 1: Time Average of Random Process

The time average of a random process $a(t)$, denoted by $\overline{a}$, is $\overline{a} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[a(t)]$.

2.6.2 Definition 2: Rate Stability

A discrete time process $a(t)$ is rate stable if $\lim_{t\to\infty} \frac{a(t)}{t} = 0$ with probability 1, and strongly stable if $\lim_{T\to\infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[|a(t)|] < \infty$.

2.6.3 Theorem 1: Queue Rate Stability

Let $Q(t)$ denote the queue length of a single-server discrete time queueing system, whose initial state $Q(0)$ is a non-negative real-valued random variable, and future states are driven by stochastic arrival and server processes $a(t)$ and $b(t)$ according to the following dynamic equation:

$$Q(t+1) = \max\{Q(t) - b(t), 0\} + a(t) \text{ for } t \in \{0, 1, 2, ...\}. \tag{2.16}$$

Then $Q(t)$ is rate stable if and only if $\overline{a} \leq \overline{b}$.

2.6.4 Theorem 2: Necessity for Queue Strong Stability

If a queue $Q(t)$ is strongly stable, and there is a finite constant c such that either $a(t) + b^{-}(t) \leq c$ with probability 1 for all t (where $b^{-}(t) \triangleq -\min[b(t), 0]$), or $b(t) - a(t) \leq c$ with probability 1 for all t, then $Q(t)$ is rate stable, i.e., $\overline{a} \leq \overline{b}$.

Besides, we say that a network is rate stable or strongly stable if all queues in this network are rate stable or strong stable as described above.

CHAPTER 3

DYNAMIC ENERGY COST OPTIMIZATION

In this section, we investigate the dynamic energy cost minimization problem in a multi-hop cellular network.

3.1 Network Layer Design

Recall that we consider downlink traffic in the network. Specifically, the destination nodes are served by the base stations via multiple hops, with the help of other nodes. Therefore, as a network layer buffer, each node i maintains a data queue Q_i^s for each service session s. The queueing law for Q_i^s is as follows:

$$Q_i^s(t+1) = \max\{Q_i^s(t) - \sum_{j \in \mathcal{N}, j \neq i} l_{ij}^s(t), 0\} + \sum_{j \in \mathcal{N}, j \neq i} l_{ji}^s(t) + k_s(t) \cdot \mathbf{1}_{i=s_s(t)}, \qquad (3.1)$$

where $l_{ij}^s(t)$ is the number of packets transmitted from i to j for service session s in time slot t, and $k_s(t)$ ($0 \leq k_s(t) \leq K_s^{max}$) is the number of packets that the source base station of service session s receives from the Internet. Note that the destination node d_s does not need to maintain a data queue for its own service since data will be directly passed on to the upper layers.

Besides, at the source and destination nodes, we have the following routing constraints:

$$\sum_{j\in\mathcal{N},j\neq i} l_{ji}^s(t) = 0, \text{ if } i = s_s(t), s \in \mathcal{S}, \tag{3.2}$$

$$\sum_{j\in\mathcal{N},j\neq i} l_{ij}^s(t) = 0, \text{ if } i = d_s, s \in \mathcal{S}, \tag{3.3}$$

$$\sum_{j\in\mathcal{N},j\neq i} l_{ji}^s(t) = v_s(t), \text{ if } i = d_s, s \in \mathcal{S}, \tag{3.4}$$

$$\sum_{i\in\mathcal{B}} \mathbf{1}_{i=s_s(t)} = 1. \tag{3.5}$$

Constraints (3.2) and (3.3) indicate that there is no incoming data and outgoing data at the source node and the destination node for service session s, respectively. Constraint (3.4) models the throughput requirement of service session s, where $v_s(t)$ is the number of packets required by session s. Constraint (3.5) indicates that there is only one source base station for session s in any time slot t.

3.2 Link Layer Design

Next, we illustrate the channel allocation and link scheduling constraints on data transmissions.

Assume that band m is available at both node i and node j, i.e., $m \in \mathcal{M}_i \cap \mathcal{M}_j$. We denote

$$\alpha_{ij}^m(t) = \begin{cases} 1, \text{ if node } i \text{ transmits to node } j \text{ using band } m \text{ in time slot } t, \\ 0, \text{ otherwise.} \end{cases} \tag{3.6}$$

Since a node is not able to transmit to or receive from multiple nodes on the same frequency band, we have

$$\sum_{j\in\mathcal{N},j\neq i} \alpha_{ij}^m(t) \leq 1, \text{ and } \sum_{i\in\mathcal{N},i\neq j} \alpha_{ij}^m(t) \leq 1. \tag{3.7}$$

Besides, a node cannot use the same frequency band for transmission and reception, due to "self-interference" at physical layer, i.e.,

$$\sum_{i\in\mathcal{N},i\neq j}\alpha_{ij}^{m}(t)+\sum_{q\in\mathcal{N},q\neq j}\alpha_{jq}^{m}(t)\leq 1. \tag{3.8}$$

Moreover, we consider that each node is only equipped with one single radio, which means that each node can only transmit or receive on one frequency band at a time. Thus, we have

$$\sum_{m\in\mathcal{M}_j}\sum_{i\in\mathcal{N},i\neq j}\alpha_{ij}^{m}(t)+\sum_{m\in\mathcal{M}_j}\sum_{q\in\mathcal{N},q\neq j}\alpha_{jq}^{m}(t)\leq 1. \tag{3.9}$$

Notice that (3.7) and (3.8) will hold whenever (3.9) holds.

Recall that in (2.3), $E_i^{TX}(t)$ is node i's consumed energy for serving its traffic. Thus, it can be calculated as follows:

$$E_i^{TX}(t)=\sum_{m\in\mathcal{M}_i}\sum_{j\in\mathcal{N},j\neq i}\alpha_{ij}^{m}(t)P_{ij}^{m}(t)\Delta t+\sum_{m\in\mathcal{M}_i}\sum_{j\in\mathcal{N},j\neq i}\alpha_{ji}^{m}(t)P_i^{recv}\Delta t, \tag{3.10}$$

where node i's receiving power, i.e., P_i^{recv}, is a constant, and Δt is the time duration of one time slot.

In addition to the above constraints at a certain node, there are also constraints due to potential interference among different nodes. In particular, according to the Physical Model discussed in Section 2.2, if node i uses a frequency band m for transmitting data to another node, the cumulative interference from all the other nodes transmitting on m at the same time plus the noise power level should be low enough so that the SINR of node i's transmission is above the threshold Γ, i.e., $g_{ij}P_{ij}^{m}(t)\geq\Gamma\big(\eta_j W^m(t)+\sum_{k\neq i,v\neq j}g_{kj}P_{kv}^{m}(t)\big)$.

Rewriting the above expression in the form of a constraint that accommodates all the link-band pairs in the network, we have

$$g_{ij}P_{ij}^m(t)\alpha_{ij}^m(t) + M_{ij}^m\big(1-\alpha_{ij}^m(t)\big) \geq \Gamma\Big(\eta_j W^m(t) + \sum_{k\neq i, v\neq j} g_{kj}P_{kv}^m(t)\alpha_{kv}^m(t)\Big), \quad (3.11)$$

where $M_{ij}^m\big(1-\alpha_{ij}^m(t)\big)$ is set as the sum of interferences from all the other nodes and the noise, i.e., $M_{ij}^m\big(1-\alpha_{ij}^m(t)\big) = \Gamma\big(\eta_j W^m(t) + \sum_{k\neq i} g_{kj}P_{max}^k\big)$.

Moreover, the flow rate over link (i,j) should satisfy the following inequality, i.e.,

$$\delta \sum_{s\in\mathcal{S}} l_{ij}^s(t) \leq \sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t \quad (3.12)$$

where δ is the number of bits per packet. (3.12) indicates that the total number of bits transmitted on a link during one time slot cannot exceed the link's capacity multiplied by the duration of one time slot.

3.3 Offline Finite-Queue-Aware Energy Cost Minimization

Our objective is to minimize the time-averaged expected energy cost of the cellular service provider given the routing, link scheduling and energy capabilities, while guaranteeing the strong stability of the network. Thus, the offline finite-queue-aware energy cost optimization problem can be formulated as follows:

P1: **Minimize** $\psi = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[f(P(t))]$,

s.t. Constraints (2.10)-(2.15), (3.2)-(3.5), (3.9)-(3.12), $\forall t \geq 0$

$\mathbf{Q}(t)$ and $\mathbf{x}(t)$ are strongly stable, (3.13)

where $\mathbf{Q}(t) = \{Q_i^s(t), \forall i \in \mathcal{N}, s \in \mathcal{S}\}$ and $\mathbf{x}(t) = \{x_i(t), \forall i \in \mathcal{N}\}$. We denote the optimal result of **P1** by ψ_{P1}^*. We can see that without the constraint (3.13), **P1** is a time-coupling stochastic Mixed-Integer Non-Linear Programming (MINLP) problem, which is already prohibitively expensive to solve. Previous approaches usually solve such problems based on Dynamic Programming and suffer from the "curse of dimensionality" problem [7]. They also require detailed statistical information on the random variables in the problem, i.e., the available spectrums, and output of renewable energy resources at each node, which may be difficult to obtain in practice. In addition, the constraint (3.13) makes **P1** an even more complicated problem. Next, we will reformulate this problem into an online optimization problem using Lyapunov optimization to break the time coupling in **P1**, and find a feasible solution to it only based on the current network states.

CHAPTER 4

ONLINE FINITE-QUEUE-AWARE ENERGY COST MINIMIZATION

In this section, we exploit Lyapunov optimization techniques to design an online finite-queue-aware algorithm to solve the energy cost minimization problem without requiring any priori knowledge of the network parameters.

Before we delve into the details, we first reformulate **P1** into an equivalent offline optimization problem **P2**. In particular, summing the inequality (3.12) over all $t \in \{0, 1, ...T-1\}$, and taking expectation and limitation on both sides, we get

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\delta \sum_{s\in\mathcal{S}} l_{ij}^{s}(t)] \leq \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^{m}(t)\alpha_{ij}^{m}(t)\Delta t]. \tag{4.1}$$

Thus, we define $\overline{P} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[f(P(t)) - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} k_s(t) \cdot \mathbf{1}_{i=s_s(t)}]$, where λ is a coefficient that can be determined by the system operator. We then formulate the following optimization problem **P2**:

P2: **Minimize** $\psi = \overline{P}$

s.t. Constraints (2.10)-(2.15), (3.2)-(3.5), (3.9)-(4.1), $\forall t \geq 0$.

We denote the optimal result of P2 by ψ_{P2}^{*}. We formulate **P2** in such a way to help ensure the strong stability of the network, which will be clear later. Besides, note that similar to **P1**, **P2** is also a time-coupling MINLP problem. In what follows, we will formulate a drift-plus-penalty problem based on **P2**, which we call **P3**.

4.1 Modeling Virtual Queues

In order to guarantee that all queues in the network are stable, we introduce virtual queues as follows. Consider a virtual queue $G_{ij}(t)$ at node i for each of its one-hop neighbor j with the following queueing law:

$$G_{ij}(t+1) = \max\left\{G_{ij}(t) - \frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t, 0\right\} + \sum_{s\in\mathcal{S}} l_{ij}^s(t). \quad (4.2)$$

This virtual queue can be understood as the link-layer buffer for link (i, j). The queue backlog $G_{ij}(t)$ represents the total number of packets stored at node i to be transmitted to node j at the beginning of time slot t[1].

For queue $G_{ij}(t)$, we have

$$\frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t - \sum_{s\in\mathcal{S}} l_{ij}^s(t) \le \frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t \le \frac{1}{\delta}c_{ij}^{max}\Delta t \quad (4.3)$$

where $\frac{1}{\delta}c_{ij}^{max}\Delta t$ is a constant. Therefore, if we can guarantee the strong stability of this queue, we can ensure its rate stability, i.e., constraint (4.1), according to Theorem 2.6.4. Besides, the virtual queue backlog is always nonnegative according to the queueing law (4.2).

Instead of utilizing $G_{ij}(t)$ directly, we build another virtual queue $H_{ij}(t) = \beta G_{ij}(t)$, where $\beta = \max_{i,j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ij}^{max}\Delta t\}$. Thus, the queueing law of $H_{ij}(t)$ is

$$H_{ij}(t+1) = \max\left\{H_{ij}(t) - \frac{\beta}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t, 0\right\} + \beta\sum_{s\in\mathcal{S}} l_{ij}^s(t). \quad (4.4)$$

Note that the strong stability of $H_{ij}(t)$ implies the strong stability of $G_{ij}(t)$, and hence (4.1) would directly follow.

[1] In order to guarantee that the queue size of $G_{ij}(t)$ is an integer in each time slot, the service rate of the queue should in fact be $\lfloor\frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t\rfloor$. Here, we assume $\frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t$ to be integers for simplicity.

4.2 Reformulation of Dynamic Energy Cost Minimization Using Lyapunov Optimization

We first define a shifted energy level $z_i(t)$ for any node $i \in \mathcal{N}$ to better control its energy storage unit, i.e.,

$$z_i(t) = x_i(t) - V\gamma^{max} - d_i^{max}, \tag{4.5}$$

where γ^{max} is the maximum first-order derivative of $f(P(t))$ with respect to $P(t)$, and V is a positive constant to be defined later. Thus, according to (2.5), $z_i(t)$ is updated following the queueing law below:

$$z_i(t+1) = z_i(t) + c_i(t) - d_i(t). \tag{4.6}$$

Note that $x_i(t)$ is stable as long as $z_i(t)$ is stable.

Next, we define a Lyapunov function [24] as

$$L(\mathbf{\Theta}(t)) \triangleq \frac{1}{2}[\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{N}}(Q_i^s(t))^2 + \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N},j\neq i}(H_{ij}(t))^2 + \sum_{i\in\mathcal{N}}(z_i(t))^2] \tag{4.7}$$

where $\mathbf{\Theta}(t) = \{\mathbf{Q}(t), \mathbf{H}(t), \mathbf{z}(t)\}$. We assume $\mathbf{Q}(0) = \mathbf{0}$, $\mathbf{H}(0) = \mathbf{0}$, and $\mathbf{z}(0) = \mathbf{0}$. This function represents a scalar measure of queues in the system. $L(\mathbf{\Theta}(t))$ being small indicates that all queue backlogs are low, while $L(\mathbf{\Theta}(t))$ being large implies that at lease one queue backlog is high. Meanwhile, its one-slot conditional Lyapunov drift is defined as

$$\Delta(\mathbf{\Theta}(t)) \triangleq \mathbb{E}\big[L(\mathbf{\Theta}(t+1)) - L(\mathbf{\Theta}(t))|\mathbf{\Theta}(t)\big]. \tag{4.8}$$

In order to minimize the long-term time-averaged expected total cost of energy from UC, instead of directly minimizing $\Delta(\mathbf{\Theta}(t))$, we intend to minimize the upper bound of the drift-plus-penalty function, which is defined as:

$$\Delta(\mathbf{\Theta}(t)) + V\mathbb{E}\big[f(P(t)) - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} k_s(t) \cdot \mathbf{1}_{i=s_s(t)} | \mathbf{\Theta}(t)\big] \tag{4.9}$$

where $V \geq 0$ is a constant that represents the weight on how much we emphasize on the energy cost minimization. Such a scheduling decision can be explained as follows: we want to make $\Delta(\mathbf{\Theta}(t))$ small to push queue backlog towards a lower congestion state, but we also want to make $\big(f(P(t)) - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} k_s(t) \cdot \mathbf{1}_{i=s_s(t)}\big)$ small in each time slot so that the energy cost can be low.

4.2.1 Lemma 1: Upper Bound of Drift-plus-penalty Function

We can have the following lemma.

Given $\Delta(\mathbf{\Theta}(t))$ defined in (4.8), we have

$$\begin{aligned}&\Delta(\mathbf{\Theta}(t)) + V\mathbb{E}\big[f(P(t)) - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} k_s(t) \cdot \mathbf{1}_{i=s_s(t)} | \mathbf{\Theta}(t)\big] \\ &\leq B + \Psi_1(t) + \Psi_2(t) + \Psi_3(t) + \Psi_4(t),\end{aligned} \tag{4.10}$$

where B is a constant, i.e.,

$$\begin{aligned}B=&\frac{1}{2}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{N}}\big[\big(\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ij}^{max}\Delta t\}\big)^2 \\ &+\big(\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ji}^{max}\Delta t\} + l_s^{max}\cdot\mathbf{1}_{i=s_s(t)}\big)^2\big] \\ &+\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N},j\neq i}\big[\frac{\beta}{\delta}(c_{ij}^{max}\cdot\Delta t)\big]^2 \\ &+\frac{1}{2}\sum_{i\in\mathcal{N}}\max\big\{(c_i^{max})^2,(d_i^{max})^2\big\}\end{aligned} \tag{4.11}$$

$\Psi_1(t)$ is only related to the link scheduling variables $\alpha_{ij}^m(t)$'s, i.e.,

$$\Psi_1(t) = -\frac{\beta}{\delta}\mathbb{E}\big[\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N},j\neq i}\big(H_{ij}(t)\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j}c_{ij}^m(t)\alpha_{ij}^m(t)\cdot\Delta t\big)|\mathbf{\Theta}(t)\big], \tag{4.12}$$

$\Psi_2(t)$ is related to the resource allocation variables $k_s(t)$ and $\mathbf{1}_{i=s_s(t)}$'s, i.e.,

$$\Psi_2(t)=\mathbb{E}\big[\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}}\big((Q_i^s(t)-\lambda V)(k_s(t)\cdot\mathbf{1}_{i=s_s(t)})\big)|\mathbf{\Theta}(t)\big], \tag{4.13}$$

$\Psi_3(t)$ is only related to the routing variables $l_{ij}^s(t)$'s, i.e.,

$$\begin{aligned}\Psi_3(t)=&\mathbb{E}\big[\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{N}}Q_i^s(t)\big(\sum_{j\in\mathcal{N}j\neq i}l_{ji}^s(t)-\sum_{j\in\mathcal{N},j\neq i}l_{ij}^s(t)\big)\\ &\mid\mathbf{\Theta}(t)\big]+\mathbb{E}\big[\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N},j\neq i}\big(\beta H_{ij}(t)\sum_{s\in\mathcal{S}}l_{ij}^s(t)\big)|\mathbf{\Theta}(t)\big],\end{aligned} \tag{4.14}$$

and $\Psi_4(t)$ is related to the energy management variables $c_i(t)$, $d_i(t)$ and $P(t), \forall i\in\mathcal{N}$, i.e.,

$$\Psi_4(t) = \mathbb{E}\big[\sum_{i\in\mathcal{N}}\big(z_i(t)(c_i(t)-d_i(t))\big)|\mathbf{\Theta}(t)\big]+V\mathbb{E}\big[f(P(t))|\mathbf{\Theta}(t)\big]. \tag{4.15}$$

Proof: Note that $\forall x, y, z$ with $x\geq 0$, $0\leq y\leq y_{max}$, $0\leq z\leq z_{max}$, we have

$$\begin{aligned}(\max\{x-y,0\}+z)^2 &\leq x^2+y^2+z^2+2x(z-y)\\ &\leq x^2+y_{max}^2+z_{max}^2+2x(z-y).\end{aligned} \tag{4.16}$$

Due to (3.12), we know that $\sum_{j\in\mathcal{N},j\neq i}l_{ij}^s(t)\leq\frac{1}{\delta}\sum_{j\in\mathcal{N},j\neq i}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j}c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t$. Denote by c_{ij}^{max} the maximum possible link capacity of link (i,j). Since $c_{ij}^m(t)$ depends on $d(i,j)$ and $W^m(t)$, among which $d(i,j)$ is constant, then c_{ij}^{max} is determined by W^{max}, i.e., the maximum bandwidth that the channels available on link (i,j) can have. Thus, according to (3.9), i.e., one node can transmit to at most one neighbor on at most one band at a time, we can get $\sum_{j\in\mathcal{N},j\neq i}l_{ij}^s(t)\leq\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\alpha}c_{ij}^{max}\Delta t\}$. Similarly, we also have

$\sum_{j\in\mathcal{N},j\neq i} l_{ji}^s(t) \leq \max_{j\in\mathcal{N},j\neq i} \frac{1}{\delta} c_{ji}^{max}\Delta t$. Besides, we have $k_s(t) \leq K_s^{max}$. Thus, based on (3.1), (4.4), (4.5) and (4.16), we can obtain the following inequalities:

$$\begin{aligned}\big(Q_i^s(t+1)\big)^2 \leq & \big(Q_i^s(t)\big)^2 + \big(\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ij}^{max}\Delta t\}\big)^2 \\ & +\big(\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ji}^{max}\Delta t\} + K_s^{max}\cdot \mathbf{1}_{i=s_s(t)}\big)^2 \\ & +2Q_i^s(t)\cdot\big(\sum_{j\in\mathcal{N},j\neq i} l_{ji}^s(t) + k_s(t)\cdot\mathbf{1}_{i=s_s(t)} - \sum_{j\in\mathcal{N},j\neq i} l_{ij}^s(t)\big)\end{aligned} \quad (4.17)$$

$$\begin{aligned}(H_{ij}(t+1))^2 \leq & \big(H_{ij}(t)\big)^2 + 2\big(\frac{\beta}{\delta}c_{ij}^{max}\Delta t\big)^2 + 2H_{ij}(t)\cdot \\ & \big(\sum_{s\in\mathcal{S}}\beta l_{ij}^s(t) - \frac{\beta}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t\big)\end{aligned} \quad (4.18)$$

$$\big(z_i(t+1)\big)^2 \leq \big(z_i(t)\big)^2 + \max\{(c_i^{max})^2, (d_i^{max})^2\} + 2z_i(t)\big(c_i(t) - d_i(t)\big) \quad (4.19)$$

Applying these inequalities to the drift-plus-penalty function, we have

$$\begin{aligned}& \Delta(\mathbf{\Theta}(t)) + V\mathbb{E}[f(P(t)) - \lambda\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}} k_s(t)\cdot\mathbf{1}_{i=s_s(t)}|\mathbf{\Theta}(t)] \\ \leq & \frac{1}{2}\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{N}}\big[\big(\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ij}^{max}\Delta t\}\big)^2 + \big(\max_{j\in\mathcal{N},j\neq i}\{\frac{1}{\delta}c_{ji}^{max}\Delta t\} \\ & +K_s^{max}\cdot\mathbf{1}_{i=s_s}\big)^2\big] + \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N},j\neq i}\big(\frac{\beta}{\delta}c_{ij}^{max}\Delta t\big)^2 + \frac{1}{2}\sum_{i\in\mathcal{N}}\max\big\{(c_i^{max})^2, (d_i^{max})^2\big\} \\ & +\mathbb{E}\big[\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{N}} Q_i^s(t)\big[\sum_{j\in\mathcal{N},j\neq i} l_{ji}^s(t) + k_s(t)\cdot\mathbf{1}_{i=s_s(t)} - \sum_{j\in\mathcal{N},j\neq i} l_{ij}^s(t)\big]|\mathbf{\Theta}(t)\big] \\ & +\mathbb{E}\big[\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N},j\neq i} H_{ij}(t)\big[\sum_{s\in\mathcal{S}}\beta l_{ij}^s(t) - \frac{\beta}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t\big]|\mathbf{\Theta}(t)\big] \\ & +\mathbb{E}\big[\sum_{i\in\mathcal{N}}\big(z_i(t)(c_i(t) - d_i(t))\big)|\mathbf{\Theta}(t)\big] \\ & +V\mathbb{E}\big[f(P(t)) - \lambda\sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}} k_s(t)\cdot\mathbf{1}_{i=s_s(t)}|\mathbf{\Theta}(t)\big]\end{aligned} \quad (4.20)$$

Thus, Lemma 1 directly follows. ■

Based on the drift-plus-penalty framework, our objective is to minimize the right-hand-side of (4.10), and hence to minimize $\Psi_1(t) + \Psi_2(t) + \Psi_3(t) + \Psi_4(t)$ since B is a constant, given the current system status $\mathbf{\Theta}(t) = \{\mathbf{Q}(t), \mathbf{H}(t), \mathbf{z}(t)\}$ in each time slot. We now use the concept of opportunistically minimizing an expectation [24], which is to minimize:

$$\hat{\Psi}_1(t) = -\frac{\beta}{\delta}\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}, j\neq i}\big(H_{ij}(t)\cdot\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t\big) \tag{4.21}$$

$$\hat{\Psi}_2(t) = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}}\big((Q_i^s(t) - \lambda V)(k_s(t)\cdot\mathbf{1}_{i=s_s(t)})\big) \tag{4.22}$$

$$\hat{\Psi}_3(t) = \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{N}} Q_i^s(t)\big(\sum_{j\in\mathcal{N} j\neq i} l_{ji}^s(t) - \sum_{j\in\mathcal{N}, j\neq i} l_{ij}^s(t)\big) + \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}, j\neq i}\big(H_{ij}(t)\sum_{s\in\mathcal{S}}\beta l_{ij}^s(t)\big) \tag{4.23}$$

$$\hat{\Psi}_4(t) = \sum_{i\in\mathcal{N}}\big(z_i(t)(c_i(t) - d_i(t))\big) + Vf(P(t)). \tag{4.24}$$

Therefore, the problem of online finite-queue-aware energy cost minimization can be formulated as follows:

P3: **Minimize** $\hat{\Psi}_1(t) + \hat{\Psi}_2(t) + \hat{\Psi}_3(t) + \hat{\Psi}_4(t)$

s.t. Constraints (2.10)-(2.15), (3.2)-(3.5), (3.9)-(3.12), $\forall t \geq 0$.

$\mathbf{\Theta}(t)$ is strongly stable. (4.25)

Note that the constraint (4.1) has been left out in **P3** (compared to **P2**) since it can be guaranteed if $\mathbf{H}(t)$ is strongly stable as mentioned before.

4.3 A Decomposition Based Approximation Algorithm

In the following we decompose **P3** into four subproblems (from S1 to S4) and solve them respectively. The intuition is that since each subproblem has fewer variables compared with that in P3 and can be solved easily, by solving the subproblems one by one, the later subproblems can treat the variables that have been solved in previous subproblem as constants. Consequently, we can obtain a feasible solution to **P3**.

4.3.1 Link Scheduling

First, we minimize $\hat{\Psi}_1(t)$ by finding the optimal link scheduling policy, i.e., determining the variables $\alpha_{ij}^m(t)$'s ($\forall i, j \in \mathcal{N}, j \neq i, m \in \mathcal{M}_i \cap \mathcal{M}_j$), as follows:

S1: Minimize $\hat{\Psi}_1(t)$

s.t. Constraint (3.9).

Since the variables $\alpha_{ij}^m(t)$'s can only take value of 0 or 1, the above subproblem is a Binary Integer Programming (BIP) problem. In the following, based on the similar ideas in [14, 29], we propose a heuristic greedy scheme called the sequential-fix (SF) algorithm to find a suboptimal solution to this problem, the solution of which can be obtained in polynomial time. The main idea of SF is to fix the binary variables $\alpha_{ij}^m(t)$'s sequentially through a series of relaxed linear programming problems. Specifically, we first set $\alpha_{ij}^m(t)$'s to 0 if $H_{ij}(t) = 0$, remove all the terms associated with such $\alpha_{ij}^m(t)$'s from the objective function, and eliminate the related constraints in (3.9). Then, in each iteration, we first relax all the 0-1 integer constraints on $\alpha_{ij}^m(t)$'s to $0 \leq \alpha_{ij}^m(t) \leq 1$ to transform the problem to a linear programming (LP) problem. Then, we solve this LP to obtain an optimal solution with each $\alpha_{ij}^m(t)$ being between 0 and 1. Among all the values, we set the largest $\alpha_{ij}^m(t)$ to 1. After that, based on the constraint (3.9), we can fix $\alpha_{pj}^m(t) = 0$ and $\alpha_{jq}^n(t) = 0$ for any $n \in \mathcal{M}_j$ and $p, q \in \mathcal{N}$. Besides, if the result includes several $\alpha_{ij}^m(t)$'s with the value of 1, we can set those $\alpha_{ij}^m(t)$'s to 1 and perform an additional fixing for the largest fractional variable in the current iteration as illustrated above. Having fixed some $\alpha_{ij}^m(t)$'s in the first iteration, we remove all the terms associated with those already fixed $\alpha_{ij}^m(t)$'s from the objective function, eliminate the related constraints in (3.9), and update the problem to a new one for the next iteration. The iteration continues until we fix all $\alpha_{ij}^m(t)$'s to be either 0 or 1.

4.3.2 Resource Allocation

Second, we minimize $\hat{\Psi}_2(t)$ by finding the source base station for each service session s ($s \in \mathcal{S}$) and determining its incoming packet rate $k_s(t)$, i.e.,

$$\textbf{S2:}\quad \textbf{Minimize}\quad \hat{\Psi}_2(t)$$

$$\textbf{s.t.}\quad \text{Constraints (3.5).}$$

We develop the following search algorithm to locally find a resource allocation policy. Specifically, at the beginning of each time slot, given the current queue backlogs $Q_i^s(t)$'s ($\forall i \in \mathcal{B}$) for each service session s, we find the base station with the smallest $Q_i^s(t)$ and choose it as the source base station. If there are multiple variables with the same smallest queue backlog, we randomly pick one of them as the source base station. After that, we can determine the source node's incoming packet rate as follows:

$$k_s(t) = \begin{cases} K_{max}^s, \text{ if } Q_{s_s}^s(t) - \lambda V < 0 \\ 0, \qquad \text{otherwise.} \end{cases} \tag{4.26}$$

4.3.3 Routing

Third, after reorganizing $\hat{\Psi}_2(t)$, we minimize it by finding the optimal routing policy, i.e., determining the variables $l_{ij}^s(t)$'s ($\forall s \in \mathcal{S}, i, j \in \mathcal{N}, j \neq i$), as follows:

$$\textbf{S3:} \quad \textbf{Minimize} \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}, j\neq i} \big(- Q_i^s(t) + Q_j^s(t) + \beta H_{ij}(t) \big) \cdot l_{ij}^s(t)$$

s.t. Constraints (3.2)-(3.4), (3.12).

We can see that S3 is an Integer Linear Programming (ILP) problem with the only variables being $l_{ij}^s(t)$'s. We notice that the total flow rate $\sum_{s\in\mathcal{S}} l_{ij}^s(t)$ over link (i, j) does not affect the flow rates over other links $\{(p, q)|p \neq i \cap q \neq j\}$, and only depends on its link capacity according to the constraint (3.12). Besides, the objective function of S3 can be viewed as a weighted sum of the variables $l_{ij}^s(t)$'s. Therefore, we can determine the flow rate over any link (i, j) at node i locally, based on its current queue backlogs $Q_i^s(t)$ and $H_{ij}(t)$, and the queue backlogs of node j, i.e., $Q_j^s(t)$. In the following, we will propose an algorithm to obtain the optimal solution for $l_{ij}^s(t)$'s.

In particular, we first set the variables l_{ij}^s's ($\forall j = s_s(t), i \in \mathcal{N} \setminus \{j\}, s \in \mathcal{S}$) and those ($\forall i = d_s, j \in \mathcal{N} \setminus \{i\}, s \in \mathcal{S}$) to 0 according to constraints (3.2) and (3.3). Besides, if a node $j = d_s$ ($s \in \mathcal{S}$) in time slot t, then the variable l_{ij}^s ($\forall i \in \mathcal{N} \setminus \{j\}$) with the smallest coefficient in the objective function of S3 is set to $v_s(t)$ due to constraint (3.4). In all the other cases, in order to minimize the objective function, node i also sets the variables $l_{ij}^s(t)$'s ($\forall j \in \mathcal{N}, j \neq i, s \in \mathcal{S}$) to 0 if their coefficients are non-negative. Otherwise, for any $l_{ij}^s(t)$'s ($s \in \mathcal{S}$) over link (i, j), node i sets the variable with the smallest coefficient to $\frac{1}{\delta} \sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t$ while the rest to 0, due to the constraint (3.12). The intuition is that by doing so, the link (i, j) can be fully utilized while minimizing S3. Besides, if there are variables $l_{ij}^s(t)$'s with the same smallest coefficients on link (i, j), node i randomly picks one of them and sets it to $\frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t$ while the rest to 0. Note that $\alpha_{ij}^m(t)$'s are known from the link scheduling optimization problem S1. It is possible that $\frac{1}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^m(t)\Delta t$ is equal to 0 if $\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} \alpha_{ij}^m(t) = 0$. Then, the corresponding variable $l_{ij}^s(t)$ is also equal to 0.

4.3.4 Energy Management

Fourth, in order to minimize $\hat{\Psi}_4(t)$, we try to find the optimal energy management for all $i \in \mathcal{N}$, i.e., determining the variables $P_{ij}^m(t)$'s, $c_i^r(t)$'s, $c_i^g(t)$'s, $r_i(t)$'s, $d_i(t)$'s, and $g_i(t)$'s. This problem can be formulated as follows:

S4: **Minimize** $\hat{\Psi}_4(t)$

s.t. Constraint (2.10)-(2.15), (3.11).

Notice that S4 is a convex optimization problem, which can be easily solved, e.g., using CPLEX, given the system states and shifted energy levels $z_i(t)$.

In summary, in each time slot, the online finite-queue-aware energy minimization problem **P3** can be solved after S1, S2, S3 and S4 are solved. The queues $\mathbf{Q}(t)$, $\mathbf{H}(t)$ and $\mathbf{z}(t)$ are then updated in each time slot according to the queueing laws (3.1), (4.4), and (4.5), respectively. We will show in the next section that all queues are strongly stable. We denote the corresponding time-averaged expected total energy cost by ψ_{P3}.

CHAPTER 5

PERFORMANCE ANALYSIS

In this section, we first prove that the proposed approximation algorithm can guarantee network strong stability. Then, we derive both the lower and upper bounds on the optimal result of **P1**.

5.1 Network Strong Stability

Our proposed algorithm finds a feasible solution to P3 which satisfies the constraints (2.10)-(2.15), (3.2)-(3.5), (3.9)-(3.12). We can have the following theorem.

5.1.1 Theorem 3: Network Strong Stability

Our proposed approximation algorithm guarantees that the queues $\mathbf{Q}(t)$, $\mathbf{H}(t)$ and $\mathbf{z}(t)$ are all strongly stable.

Proof: *First*, we demonstrate the strong stability of $\mathbf{Q}(t)$ by considering an arbitrary queue $Q_i^s(t)$. Specifically, we prove by induction that $Q_i^s(t) \leq \lambda V + K_{max}^s$.

When $t = 0$, we have $Q_i^s(0) = 0 < \lambda V + K_{max}^s$.

Assume that we have $Q_i^s(t) \leq \lambda V + K_{max}^s$ in time slot t ($t \geq 0$). Then, we consider the following two cases to prove the stability of $Q_i^s(t)$.

1) if $i = s_s(t)$: According to the queueing law of $Q_i^s(t)$, we have

$$Q_{s_s}^s(t+1) = \max\{Q_{s_s}^s(t) - \sum_{j \in \mathcal{N}, j \neq s_s} l_{s_s j}^s(t), 0\} + k_s(t). \tag{5.1}$$

Based on the derived solution to the subproblem $S2$, we have the following two sub-cases:

- If $Q_{s_s}^s(t) \geq \lambda V$, according to the optimal solution to S2, we know that $k_s(t) = 0$. Thus, we have

$$Q_{s_s}^s(t+1) \leq Q_{s_s}^s(t) \leq \lambda V + K_{max}^s. \tag{5.2}$$

- If $Q_{s_s}^s(t) < \lambda V$, according to the optimal solution to S3, we get that $k_s(t) = K_{max}^s$. Following (5.1), we have

$$Q_{s_s}^s(t+1) \leq Q_{s_s}^s(t) + K_{max}^s \leq \lambda V + K_{max}^s. \tag{5.3}$$

Therefore, we have $Q_{s_s}^s(t) \leq \lambda V + K_{max}^s$.

2) if $i \neq s_s(t)$ and $i \neq d_s$: We then explore the stability of $Q_i^s(t)$ when $i \neq s_s$ and $i \neq d_s$, whose queueing law is:

$$Q_i^s(t+1) = \max\{Q_i^s(t) - \sum_{j \in \mathcal{N}, j \neq i} l_{ij}^s(t), 0\} + \sum_{j \in \mathcal{N}, j \neq i} l_{ji}^s(t). \tag{5.4}$$

Since only one neighboring node can transmit to node i in time slot t, we denote it by j. Consider the coefficient in front of $l_{ji}^s(t)$ in the objective function of S3.

- If $Q_i^s(t) < Q_j^s(t) - \beta H_{ji}(t)$, according to (5.4), we have

$$\begin{aligned} Q_i^s(t+1) &\leq Q_i^s(t) + l_{ji}^s(t) < Q_j^s(t) - \beta H_{ji}(t) + l_{ji}^s(t) \\ &\leq Q_j^s(t) \leq \lambda V + K_{max}^s, \end{aligned} \tag{5.5}$$

  The third inequality above can be proved in the following two cases.

  - If $H_{ji}(t) = 0$, according to the solution to S1, we can know that $\alpha_{ji}^m(t) = 0$ $\forall m \in \mathcal{M}_j \cap \mathcal{M}_i$, and hence $l_{ji}^s(t) = 0$. Thus, the inequality holds.
  - If $H_{ji}(t) \geq 1$, we have $\beta H_{ji}(t) \geq l_{ji}^s(t)$, as $l_{ji}^s(t) \leq \max_{i,j \in \mathcal{N}, j \neq i}\{\frac{1}{\delta} c_{ij}^{max} \Delta t\} = \beta$ as defined before.

- If $Q_i^s(t) \geq Q_j^s(t) - \beta H_{ji}(t)$, according to our proposed solution to S3, we know that $l_{ji}^s(t) = 0$. Following (5.4), we have

$$Q_i^s(t+1) \leq Q_i^s(t) \leq \lambda V + K_{max}^s. \tag{5.6}$$

  Therefore, we also have $Q_i^s(t) \leq \lambda V + K_{max}^s$.

Note that the destination node d_s does not need to maintain a data queue since data will be directly passed on to the upper layers. Consequently, based on the above results, we can see that an arbitrary queue $Q_i^s(t)$ is finite in any time slot. Thus, $\mathbf{Q}(t)$ is strongly stable by Definition 2.

Second, we prove the strong stability of $\mathbf{H}(t)$, and particularly,

$$H_{ij}(t) \leq \max_{0 \leq k \leq t} \sum_{s \in \mathcal{S}} l_{ij}^s(k) \tag{5.7}$$

for any $i, j \in \mathcal{N}$, by induction. We consider an arbitrary queue $H_{ij}(t)$.

When $t = 0$, we have $H_{ij}(0) = 0$, and hence (5.7) holds.

When $t = 1$, we have $H_{ij}(1) = \sum_{s \in \mathcal{S}} l_{ij}^s(1)$ according to the queueing law (4.4), and (5.7) holds.

Assume (5.7) holds in time t, i.e., $H_{ij}(t) \leq \max_{0 \leq k \leq t} \sum_{s \in \mathcal{S}} l_{ij}^s(k)$. Then, at the beginning of time slot $t+1$, we have

$$H_{ij}(t+1) = \max\left\{H_{ij}(t) - \frac{\beta}{\delta} \sum_{m \in \mathcal{M}_i \cap \mathcal{M}_j} c_{ij}^m(t) \alpha_{ij}^m(t) \Delta t, 0\right\} + \beta \sum_{s \in \mathcal{S}} l_{ij}^s(t). \tag{5.8}$$

If $H_{ij}(t) > \frac{\beta}{\delta} \sum_{m \in \mathcal{M}_i \cap \mathcal{M}_j} c_{ij}^m(t) \alpha_{ij}^m(t) \Delta t$, with inequality (3.12), we have

$$H_{ij}(t+1) \leq H_{ij}(t) \leq \max_{0 \leq k \leq t} \sum_{s \in \mathcal{S}} l_{ij}^s(k) \leq \max_{0 \leq k \leq t+1} \sum_{s \in \mathcal{S}} l_{ij}^s(k). \tag{5.9}$$

If $H_{ij}(t) \leq \frac{\beta}{\delta} \sum_{m \in \mathcal{M}_i \cap \mathcal{M}_j} c_{ij}^m(t) \alpha_{ij}^m(t) \Delta t$, then

$$H_{ij}(t+1) = \sum_{s \in \mathcal{S}} l_{ij}^s(t) \leq \max_{0 \leq k \leq t+1} \sum_{s \in \mathcal{S}} l_{ij}^s(k). \tag{5.10}$$

Therefore, (5.7) holds when $t = t + 1$ as well.

Since $\sum_{s \in \mathcal{S}} l_{ij}^s(t) \leq \frac{\beta}{\delta} c_{ij}^{max} \Delta t$, we have that $H_{ij}(t) \leq \frac{\beta}{\delta} c_{ij}^{max} \Delta t$ and hence always finite and strongly stable.

Third, we prove the strong stability of $\mathbf{z}(t)$. Since $z_i(t) \leq x_i(t)$, the strong stability of $\mathbf{z}(t)$ directly follows if we prove the strong stability of $\mathbf{x}(t)$. Firstly, we define the maximum value of V as:

$$V^{max} = \min_{i \in \mathcal{N}} \frac{x_i^{max} - c_i^{max} - d_i^{max}}{\gamma^{max}}. \tag{5.11}$$

Assume that for arbitrary node i, (2.11) holds in time slot t. Then we consider three cases when in the time slot $t + 1$.

- If $0 \leq x_i(t) < d_i^{max}$, Recall that $c_i(t) = \omega_i(t) c_i^g(t) + c_i^r(t)$. In this case, the partial derivative of the objective function of S4, i.e., $\hat{\Psi}_4(t)$, with respect to $c_i^r(t)$, is

$$\frac{\partial \hat{\Psi}_4(t)}{\partial c_i^r(t)} = z_i(t) + V \frac{\partial f(P(t))}{\partial c_i^r(t)} \leq x_i(t) - V\gamma^{max} - d_i^{max} + 0 < 0. \tag{5.12}$$

  Thus, by solving S4, i.e., minimizing $\hat{\Psi}_4(t)$, our energy management scheme leads to the control decisions that maximizes $c_i^r(t)$. Due to constraint (2.8), we have $d_i(t) = 0$. Therefore, according to (2.5), we get $x_i(t+1) = x_i(t) + c_i(t)$ and hence $0 \leq x_i(t+1) \leq d_i^{max} + c_i^{max} \leq x_i^{max}$ due to constraint (2.14).

- If $d_i^{max} \leq x_i(t) \leq V\gamma^{max} + d_i^{max}$, Since $V \leq V^{max} \leq \frac{x_i^{max} - c_i^{max} - d_i^{max}}{\gamma^{max}}$, we have $x_i(t) \leq x_i^{max} - c_i^{max}$. Thus, according to (2.5), we can obtain $x_i(t+1) \leq x_i^{max} - c_i^{max} + c_i(t) - d_i(t) \leq x_i^{max}$ and $x_i(t+1) \geq d_i^{max} + c_i(t) - d_i(t) \geq 0$.

- If $V\gamma^{max} + d_i^{max} < x_i(t) \leq x_i^{max}$. Note that $V \leq \frac{x_i^{max} - c_i^{max} - d_i^{max}}{\gamma^{max}}$, and hence $V\gamma^{max} + d^{max} \leq x_i^{max} - c_i^{max} < x_i^{max}$. The partial derivative of the objective function of S4 with respect to $d_i(t)$ is

$$\frac{\partial \hat{\Psi}_4(t)}{\partial d_i(t)} = -V \frac{\partial f(P(t))}{\partial d_i(t)} - z_i(t) \leq 0 - x_i(t) + V\gamma^{max} + d_i^{max} < 0. \tag{5.13}$$

  Thus, our energy management scheme minimizing $\hat{\Psi}_4(t)$ results in control decisions that satisfy $d_i(t) = d_i^{max}$. Due to constraint (2.10), we have $c_i(t) = 0$. Thus, according to (2.5), we get $x_i(t+1) = x_i(t) - d_i^{max}$ and hence $0 \leq x_i(t+1) \leq x_i^{max} - d_i^{max} \leq x_i^{max}$.

Therefore, we can see that (2.11) holds for all $t \geq 0$. ■

5.2 Lower and Upper bounds on ψ^*_{P1}

In what follows, we obtain both lower and upper bounds on the optimal results of P1, i.e., ψ^*_{P1}.

5.2.1 Theorem 4: Upper Bound

The solution obtained from our proposed algorithm serves as a suboptimal yet feasible solution to **P1**, and the corresponding time-averaged expected amount of energy cost works as an upper bound on the optimal result of **P1**, i.e., $\psi^*_{P1} \leq \psi_{P3}$.

Proof: The proposed decomposition based algorithm finds a solution that satisfies all the constraints in **P3**, i.e., (2.10)-(2.15), (3.2)-(3.5), (3.9)-(3.12), and (4.25). Thus, the solution is also a feasible solution to **P1**, and the corresponding time-averaged expected energy cost, i.e., ψ_{P3}, is no less than the optimal result of **P1**, i.e., $\psi_{P3} \geq \psi^*_{P1}$. ■

Next, we find a lower bound on ψ^*_{P1}. We first present a lemma as follows.

5.2.2 Lemma 2: Lower Bound

The time-averaged expected amount of energy cost achieved by optimally solving **P3**, denoted by ψ^*_{P3}, is within a constant gap $\frac{B}{V}$ from the minimum time-averaged expected energy cost achieved by **P2**, i.e., ψ^*_{P2}. Particularly, we have $\psi^*_{P3} - \frac{B}{V} \leq \psi^*_{P2}$ where B and V are defined in Section 4.2.

Proof: Denote by $\widehat{\alpha^m_{ij}}(t)$, $\widehat{k_s}(t)$, $\widehat{\mathbf{1}}_{i=s_s(t)}$, $\widehat{l^s_{ij}}(t)$, $\widehat{c_i}(t)$, $\widehat{d_i}(t)$, and $\widehat{f(P(t))}$ the results obtained by our proposed scheme in time slot t, i.e., based on the optimal solution to **P3**. We also denote by $\alpha_{ij}^{m*}(t)$, $k_s^*(t)$, $\mathbf{1}_{i=s_s(t)}^*$, $l_{ij}^{s*}(t)$, $c_i^*(t)$, $d_i^*(t)$, and $f^*(P(t))(t)$ the results that we get for time slot t based on the optimal solution to **P1**. Thus, from Lemma 4.2.1, we can have

$$\begin{aligned}
&\Delta(\mathbf{\Theta}(t)) + V\mathbb{E}[\widehat{f(P(t))} - \lambda \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}} \widehat{k_s(t)} \cdot \widehat{\mathbf{1}}_{i=s_s(t)}|\mathbf{\Theta}(t)] \\
\leq& B + \widehat{\Psi_1(t)} + \widehat{\Psi_2(t)} + \widehat{\Psi_3(t)} + \widehat{\Psi_4(t)} \\
\leq& B + \Psi_1^*(t) + \Psi_2^*(t) + \Psi_3^*(t) + \Psi_4^*(t) \\
=& B + V\mathbb{E}[f^*(P(t)) - \lambda \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}} k_s^*(t)\cdot \mathbf{1}_{i=s_s(t)}^*] \\
&+ \hat{\Psi}_1^*(t) + \hat{\Psi}_3^*(t) + \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}} \big(Q_i^s(t)(k_s^*(t)\cdot {\mathbf{1}^*}_{i=s_s(t)})\big) \\
&+ \sum_{i\in\mathcal{N}} \big(z_i(t)(c_i^*(t) - d_i^*(t))\big) \\
=& B + V\mathbb{E}[f^*(P(t)) - \lambda \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{B}} k_s^*(t)\cdot \mathbf{1}_{i=s_s(t)}^*] \\
&+ \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}, j\neq i} Q_i^s(t) \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\big(l_{ji}^{s*}(t) + k_s^*(t)\cdot \mathbf{1}_{i=s_s(t)}^* \\
&- l_{ij}^{s*}(t)\big) + \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}, j\neq i} H_{ij}^s(t) \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\big(\sum_{s\in\mathcal{S}} \beta l_{ij}^{s*}(t) \\
&- \frac{\beta}{\delta}\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t)\alpha_{ij}^{m*}(t)\Delta t\big) \\
&+ \sum_{i\in\mathcal{N}} z_i(t) \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\big(c_i^*(t) - d_i^*(t)\big)
\end{aligned} \tag{5.14}$$

Note that the third step is due to the fact that the optimal solutions to **P1** are obtained independent of the current queues $\mathbf{\Theta}(t)$. The fourth step is due to the strong law of large numbers: If $\{a(t)\}_{t=0}^{\infty}$ are i.i.d. random variables, we have $\Pr(\frac{1}{T}\lim_{t\to\infty}\sum_{t=0}^{T-1} a(t) =$

$\mathbb{E}\{a(t)\}) = 1$ almost surely. Consequently, taking expectation of the above inequality yields:

$$\begin{aligned}
&\mathbf{E}[\widehat{L}(\boldsymbol{\Theta}(t+1))] - \mathbf{E}[\widehat{L}(\boldsymbol{\Theta}(t))] + V\mathbb{E}[\widehat{f(P(t))} - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} \widehat{k_s(t)} \cdot \widehat{\mathbf{1}}_{i=s_s(t)}] \\
\leq & B + V\mathbb{E}[f^*(P(t)) - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} k_s^*(t) \cdot \mathbf{1}^*_{i=s_s(t)}] \\
+ & \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}, j\neq i} Q_i^s(t) \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\big[l_{ji}^{s*}(t) + k_s^*(t) \cdot \mathbf{1}^*_{i=s_s(t)} \\
- & l_{ij}^{s*}(t)\big] + \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}, j\neq i} H_i^s(t) \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\big[\sum_{s\in\mathcal{S}} \beta l_{ij}^{s*}(t) \\
- & \frac{\beta}{\delta} \sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t) \alpha_{ij}^{m*}(t) \Delta t\big] + \sum_{i\in\mathcal{N}} z_i(t) \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\big[c_i^*(t) - d_i^*(t)\big]
\end{aligned} \tag{5.15}$$

Since we have prove the strong stability of $\mathbf{Q}(t)$, $\mathbf{H}(t)$ and $\mathbf{z}(t)$ are all strongly stable, we know that $\mathbf{Q}(t)$, $\mathbf{H}(t)$ and $\mathbf{z}(t)$ are also rate stable, according to Theorem 2. So we can have:

$$\overline{l_{ji}^{s*}(t)} + \overline{k_s^*(t) \cdot \mathbf{1}_{i=s_s}} - \overline{l_{ij}^{s*}(t)} \leq 0 \tag{5.16}$$

$$\overline{\sum_{s\in\mathcal{S}} \beta l_{ij}^{s*}(t)} - \frac{\beta}{\delta} \overline{\sum_{m\in\mathcal{M}_i\cap\mathcal{M}_j} c_{ij}^m(t) \alpha_{ij}^{m*}(t) \Delta t} \leq 0 \tag{5.17}$$

$$\overline{c_i^*(t)} - \overline{d_i^*(t)} \leq 0 \tag{5.18}$$

Therefore, we can obtain

$$\begin{aligned}
&\mathbf{E}[\widehat{L}(\boldsymbol{\Theta}(t+1))] - \mathbf{E}[\widehat{L}(\boldsymbol{\Theta}(t))] + V\mathbb{E}[\widehat{f(P(t))} - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} \widehat{k_s(t)} \cdot \widehat{\mathbf{1}}_{i=s_s(t)}] \\
\leq & B + V\mathbb{E}[f^*(P(t)) - \lambda \sum_{s\in\mathcal{S}} \sum_{i\in\mathcal{B}} k_s^*(t) \cdot \mathbf{1}^*_{i=s_s(t)}]
\end{aligned} \tag{5.19}$$

Summing the above over $t \in \{0, 1, 2, ..., T-1\}$ for any positive integer T yields

$$\mathbb{E}[\widehat{L}(\mathbf{\Theta}(T))] - \mathbb{E}[\widehat{L}(\mathbf{\Theta}(0))] + V \sum_{t=0}^{T-1} \mathbb{E}[\widehat{f(P(t))} - \lambda \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{B}} \widehat{k_s(t)} \cdot \widehat{\mathbf{1}}_{i=s_s(t)}]$$
$$\leq TB + V \sum_{t=0}^{T-1} \mathbb{E}[f^*(P(t)) - \lambda \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{B}} k_s^*(t) \cdot \mathbf{1}^*{}_{i=s_s(t)}]. \quad (5.20)$$

Since all queues are finite in all time slots, dividing both sides of (5.20) by VT and taking limits as $T \to \infty$ leads to

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\widehat{f(P(t))} - \lambda \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{B}} \widehat{k_s(t)} \cdot \widehat{\mathbf{1}}_{i=s_s(t)}]$$
$$\leq \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[f^*(P(t)) - \lambda \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{B}} k_s^*(t) \cdot \mathbf{1}^*{}_{i=s_s(t)}] + \frac{B}{V}, \quad (5.21)$$

which means $\psi^*_{P3} - B/V \leq \psi^*_{P2}$ ■

Recall that **P1**, **P2** and **P3** are both Mixed-Integer Programming problems. We relax **P2** to a Linear Programming (LP) problem without the strong stability constraint (3.13) denoted by $\overline{P2}$, and formulate a corresponding online energy cost minimization problem denoted by $\overline{P3}$. We can see that $\overline{P3}$ is a relaxed LP problem based on **P3** without the strong stability constraint (4.25), which can be easily solved. Denoted by $\psi^*_{\overline{P1}}$ and $\psi^*_{\overline{P3}}$ the time-averaged expected amount of energy cost obtained by optimally solving $\overline{P1}$ and $\overline{P3}$, respectively, based on Lemma 2, we can know that $\psi^*_{\overline{P3}} - \frac{B}{V} \leq \psi^*_{\overline{P2}}$. Since obviously we also have $\psi^*_{\overline{P2}} \leq \psi^*_{P2} \leq \psi^*_{P1}$, we can arrive at the following result.

The optimal result of **P1** is lower bounded by $\psi^*_{\overline{P3}} - B/V$, where $\psi^*_{\overline{P3}}$ can be obtained by optimally solving $\overline{P3}$.

CHAPTER 6

SIMULATION RESULTS

6.1 System Settings

In order to complement the analysis in the previous sections, we carry out extensive simulations to evaluate the performance of our proposed scheme. Our goals are to obtain the lower and upper bounds on the optimal result of **P1**, to examine the tradeoff between energy cost and queue size, and to demonstrate the energy efficiency of our scheme compared with that of other similar energy management strategy. Simulations are conducted under CPLEX 12.4 on a computer with a 3.00 GHz CPU and 4 GB RAM.

6.2 Experiment Results

Specifically, we consider a square network of area $2000m \times 2000m$, where 2 base stations are located at coordinates (500m, 500m), (1500m, 500m), respectively, and 20 users are randomly distributed. Besides, we assume there is one cellular band with bandwidth 1 MHz and four other spectrum bands whose bandwidth are independently and uniformly distributed within [1, 2] MHz in each time slot. Only a random subset of the spectrum bands are available at each mobile user while base stations can access all the bands. Each service session has a traffic demand of 100 Kbps. Some other important simulation parameters are listed as follows. The path loss exponent is 4 and $C = 62.5$. The SINR

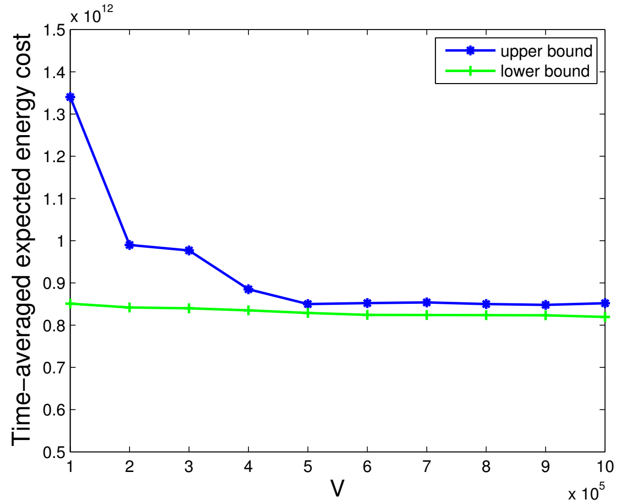

Figure 6.1

Time-averaged expected energy cost

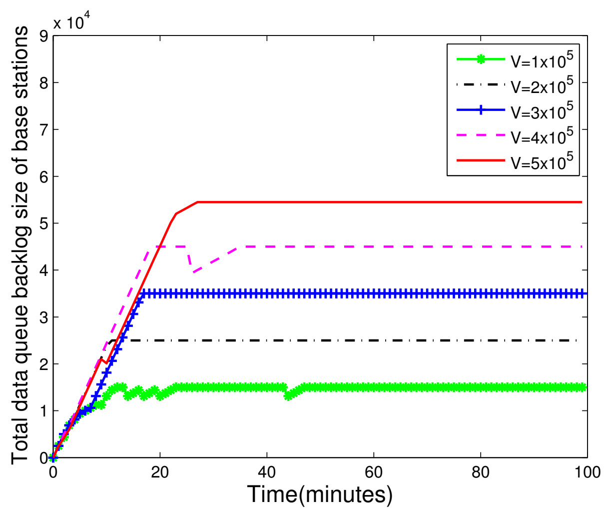

Figure 6.2

Total data queue backlog size of base stations over time

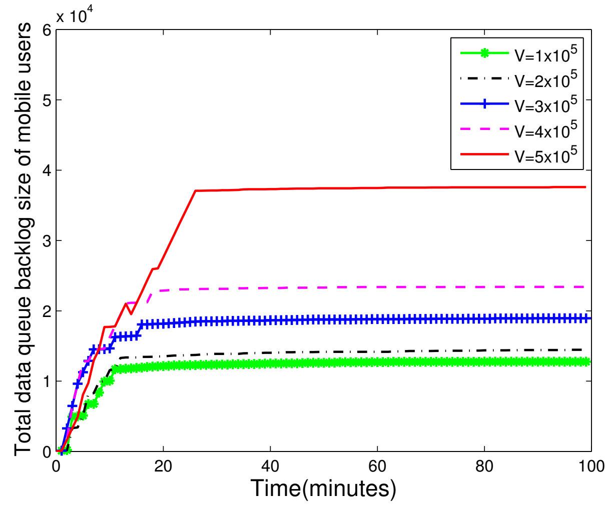

Figure 6.3

Total data queue backlog size of mobile users over time

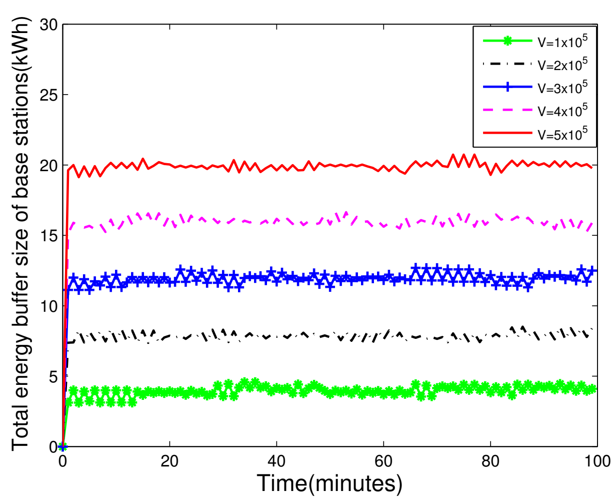

Figure 6.4

Total energy buffer size of base stations over time

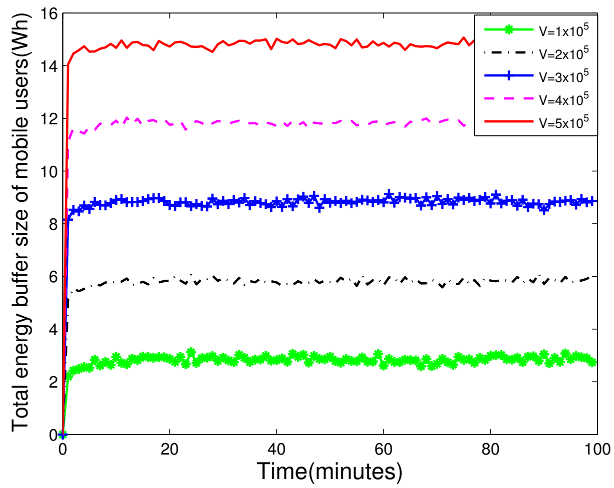

Figure 6.5

Total energy buffer size of mobile users over time

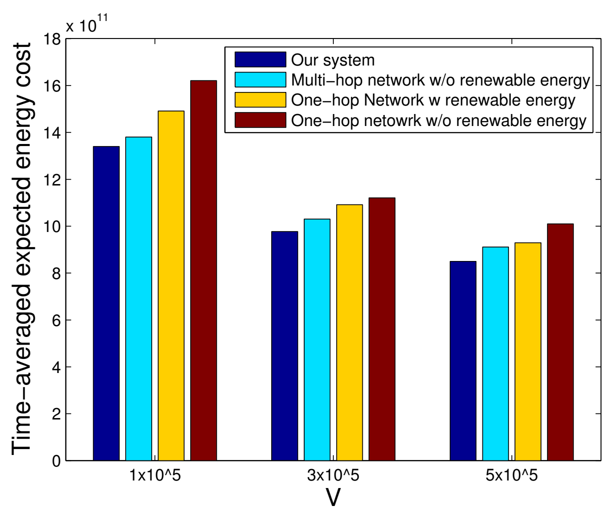

Figure 6.6

Performance comparison of different architectures threshold is $\Gamma = 1$. The noise power spectral density is $\eta = 10^{-20}$ W/Hz at all nodes. All nodes ($\forall i \in \mathcal{U}$) have the same maximum transmission power, which is $P^i_{max} = 1$ W while base stations have a much larger transmission power, i.e., 20 W. In addition, the outputs of mobile users' renewable energy resources and that of base stations' are assumed to be independently and uniformly distributed within [0, 1] W and [0, 15] W, respectively, in each time slot. The maximum charging and discharging limits on each user's energy storage device in a time slot, i.e., c_i^{max} and d_i^{max}, are both set to 0.06 kWh for mobile users and 0.1 kWh for base stations. The maximum amount of energy that each node can draw from the power grid in a time slot, i.e., p_i^{max}, is set to 0.2 kWh. The energy generation cost function, i.e., $f(P(t))$, is defined as $f(P(t)) = aP^2(t) + bP(t) + c$, where $a = 0.8, b = 0.2$ and $c = 0$. All our results presented below are collected after the experiments run for a period of $T = 100$ time slots, the duration of each of which is set to 1 minute.

In Figure 6.1, we show both the upper and lower bounds on the optimal result of $\mathbf{P1}$. Recall that the upper bound is achieved by our proposed algorithm, i.e., ψ_{P3}, and the lower bound is obtained by optimally solving the relaxed problem $\overline{P3}$, i.e., $\psi^*_{\overline{P3}} - B/V$. We can find that the lower and upper bounds get closer to each other as V increases.

Then, we examine the tradeoff between energy cost and the queue backlog sizes incurred by our scheme. We find that in Figure 6.2 and Figure 6.3, the data queue backlog sizes of base stations and mobile users increase as time goes by and are bounded. We can also get similar results in Figure 6.4 and Figure 6.5 for energy queues. Since the expected total sizes of all data queues and energy buffers of both mobile users and base stations are all finite, each single data queue and energy buffer in the network are finite in each time slot, therefore guaranteeing the strong stability of the network. Besides, a larger V results in a larger queue backlog size. This is because a larger V means more emphasis on the energy cost minimization than on the queue size and that the system needs to have a larger queue buffer so as to save more energy cost. The results in Figure 6.1-Figure 6.5 together show the tradeoff between energy cost minimization and queue length in our proposed algorithm.

Lastly, we compare the time-averaged expected energy cost of our proposed architecture with other cellular network architectures, i.e., multi-hop network without renewable energy, one-hop network with renewable energy, and one-hop network without renewable energy. As shown in Figure 6.6, our system has the lowest time-averaged expected energy cost among these four network systems when V goes from 1×10^5 to 5×10^5. Specifically, compared with the multi-hop network without renewable energy, our system can take advantage of the renewable energy and the energy stored locally and hence save energy cost. In addition, by comparing one-hop and multi-hop networks, we can find that the latter have lower energy cost. This is because multi-hop technology enables nodes in the network to use lower transmission powers to help each other with the transmissions and reduce energy consumption.

CHAPTER 7

CONCLUSIONS AND FUTURE WORKS

7.1 Main Contributions

In this paper, we propose an energy cost minimization framework for downlink data communication in multi-hop cellular networks. In particular, with the objective of minimizing the long-term time-averaged expected energy cost of cellular service provider while guaranteeing the strong stability of the network, we construct a time-coupling stochastic Mixed-Integer Non-Linear Programming (MINLP) problem, which is prohibitively expensive to solve. By employing Lyapunov optimization theory, we reformulate the problem and develop a decomposition based scheme to solve the problem in each time slot without priori knowledge of the network statistics. The proposed scheme can ensure the network strong stability. Both lower and upper bounds on the optimal result of the original optimization problem are obtained. Extensive simulation results validate the energy cost savings of the proposed scheme.

7.2 Future Research

In our future work, we are going to continue on this topic by investigating the delay problems while guaranteeing the energy constraints in multi-hop cellular networks. On the other hand, as the cloud computing has become a new paradigm that attracts increasing attentions for both academia and industry, the security and privacy issues in cloud computing become more critical. Therefore, in our future works, we plan to design an efficient, secure and private cloud computing framework by incorporating current energy efficient algorithms in this dissertation and new secure and private techniques.

REFERENCES

[1] “Sustainable Energy Use in Mobile Communications,”, Ericsson Inc., August 2007.

[2] “Overview of Energy Saving of Networks,”, ITU-T Focus Group on Future Networks, 2010.

[3] “Samsung’s solar-powered Blue Earth touchscreen phone review,”, mobileburn, 2010.

[4] “Telecommunication management; Study on Energy Savings Management (ESM),”, 3GPP TR 32.826, March 2010.

[5] “Telecommunication Predictions 2010, Deloitte,”, Technology, Media and Telecommunications Industry Group, 2010.

[6] “E-Plus, Nokia Siemens Networks build Germanys first off-grid base station,”, Nokia Siemens, 2011.

[7] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1 and 2, 2 edition, Athena Scientific, Belmont, Massachusetts, 2007.

[8] A. Bousia, E. Kartsakli, L. Alonso, and C. Verikoukis, “Energy efficient base station maximization switch off scheme for LTE-advanced,” *IEEE 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Barcelona, Spain, September 2012.

[9] J. Cho and Z. J. Haas, “On the throughput enhancement of the downstream channel in cellular radio networks through multihop relaying,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, 2004, pp. 1206–1219.

[10] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, “The global footprint of mobile communications: The ecological and economic perspective,” *IEEE Communications Magazine*, vol. 49, no. 8, August 2011, pp. 55–62.

[11] P. Gupta and P. Kumar, “The Capacity of Wireless Networks,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, March 2000, pp. 388–404.

[12] T. Han and N. Ansari, “On Optimizing Green Energy Utilization for Cellular Networks with Hybrid Energy Supplies,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, August 2013, pp. 3872–3882.

[13] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Communications Surveys and Tutorials*, vol. 13, no. 4, 2011, pp. 524–540.

[14] Y. T. Hou, Y. Shi, and H. D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio Networks," *Proc. of IEEE Conference on Computer Communications (INFOCOM'07)*, Anchorage, AL, May 2007.

[15] Y. T. Hou, Y. Shi, and H. D. Sherali, "Spectrum Sharing for Multi-Hop Networking with Cognitive Radios," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, January 2008, pp. 146–155.

[16] J. J. Jaramillo and R. Srikant, "Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic," *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'10)*, San Diego, USA, March 2010.

[17] H. Li, W. Huang, Z. L. C. Wu, and F. C. M. Lau, "Utility-Maximizing Data Dissemination in Socially Selfish Cognitive Radio Networks," *Proceedings of IEEE MASS*, San Francisco, CA, November 2011.

[18] P. Li, X. Huang, and Y. Fang, "Capacity Scaling of Multihop Cellular Networks," *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'11)*, Shanghai, China, April 2011.

[19] X. J. Li, B.-C. Seet, and P. H. J. Chong, "Multihop cellular networks: Technology and economics," *Computer Networks*, vol. 52, no. 9, June 2008, pp. 1825–1837.

[20] Y. Lin and Y. Hsu, "Multihop cellular: a new architecture for wireless communications," *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM'00)*, Tel Aviv, Israel, March 2000.

[21] J. T. Louhi, "Energy efficiency of modern cellular base stations," *in proceedings of the 29th International Telecommunications Energy Conference (INTELEC 2007)*, Rome, Italy, October 2007.

[22] X. Lu, W. Wang, and J. Ma, "An empirical study of communication infrastructures towards the smart grid: Design, implementation, and evaluation," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, March 2013, pp. 170–183.

[23] M. J. Neely, "Intelligent packet dropping for optimal energy-delay tradeoffs in wireless downlinks," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, March 2009, pp. 565–579.

[24] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, Morgan and Claypool Publishers, 2010.

[25] M. J. Neely, “Opportunistic Scheduling with Worst Case Delay Guarantees in Single and Multi-Hop Networks,” *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM’11)*, Shanghai, China, April 2011.

[26] Z. Niu, “TANGO: traffic-aware network planning and green operation,” *IEEE Wireless Communications*, vol. 18, no. 5, 2011, pp. 25–29.

[27] Z. Niu, Y. Wu, J. Gong, and Z. Yang, “Cell zooming for cost-efficient green cellular networks,” *IEEE Communications Magazine*, vol. 48, no. 11, November 2010, pp. 74–79.

[28] E. Oh, K. Son, and B. Krishnamachari, “Dynamic Base Station Switching-On/Off Strategies for Green Cellular Networks,” *IEEE transactions on Wireless Communications*, vol. 12, no. 5, May 2013, pp. 2126–2136.

[29] M. Pan, C. Zhang, P. Li, and Y. Fang, “Joint Routing and Link Scheduling for Cognitive Radio Networks under Uncertain Spectrum Supply,” *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM’11)*, Shanghai, China, April 2011.

[30] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, “Traffic-driven power saving in operational 3G cellular networks,” *Proc. of the 17th annual international conference on Mobile computing and networking (Mobicom’11)*, Las Vegas, USA, September 2011.

[31] Y. Pochet and L. Wolsey, *Production Planning by Mixed Integer Programming*, Secaucus, 2006.

[32] A. Radwan and H. S. Hassanein, “Does multi-hop communication extend the battery life of mobile terminals?,” *IEEE Global Telecommunications Conference, (GLOBECOM’06)*, San Francisco, USA, November 2006.

[33] R. Urgaonkar and M. J. Neely, “Opportunistic scheduling with reliability guarantees in cognitive radio networks,” *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, June 2009, pp. 766–777.

[34] V. Venkataramanan, X. Lin, L. Ying, and S. Shakkottai, “On scheduling for minimizing end-to-end buffer usage over multihop wireless networks,” *Proceeding of the IEEE International Conference on Computer Communications (INFOCOM’10)*, San Diego, USA, March 2010.