A Study of Optimized Caching and User Mobility in Wireless Cache-enabled Networks

by

Bitan Banerjee

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Communications

Department of Electrical & Computer Engineering
University of Alberta

Abstract

With increasing data rate demand, caching popular multimedia content at base stations or routers is emerging as a promising technology for 5G. The primary benefits of wireless caching in cellular networks are traffic offload from macro base stations and decreased delays at the user equipment. However, there are few points of concern that have yet to be addressed, such as optimal caching strategy, impacts of mobility and interference characterization. We address these concerns with three contributions. Firstly, we propose a caching strategy to enhance delay performance by utilizing the core routers. Secondly, we develop a mobility analysis for a small-cell network with caching abilities at small-cell base stations. Finally, the interference characterization of a cache network deviates from a traditional cellular network due to the opportunistic nature of association, and therefore, we develop an asymptotic analysis for the generalized fading channel that can be used for considered networks.

Our analysis leads to following conclusions: firstly, static caching or simple content placement strategy can outperform state-of-the-art dynamic caching strategies; secondly, association with a cache source is ineffective for high velocity users; and finally, our developed asymptotic approximation for a Generalized-Gamma fading model provides better performance measurements over existing methods.

Preface

This thesis is an original work conducted by Bitan Banerjee.

Chapter 3 of the thesis was presented as B. Banerjee, A. Seetharam, and C. Tellambura, "Greedy Caching: A Latency-aware Caching Strategy for Information-centric Networks," IFIP Networking, June 2017.

Chapter 4 of the thesis was presented in part as B. Banerjee and C. Tellambura, "Study of Mobility in Cache-enabled Wireless Heterogeneous Networks," IEEE WCNC, Mar. 2017.

Chapter 5 of the thesis was presented in part as B. Banerjee and C. Tellambura, "Generalized Asymptotic Measures for Wireless Fading Channels with a Logarithmic Singularity," IEEE WCNC, Mar. 2017.

I was responsible for the concept formation, technical apparatus, simulation data collection and manuscript composition. C. Tellambura was the supervisory author and was involved with concept formation and manuscript composition.

Dedicated to my beloved parents and my Jhil

Acknowledgements

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted.

My deep gratitude first goes to Prof. Chintha Tellambura, who expertly guided me through graduate education and who shared the excitement of discovery. His unwavering enthusiasm kept me focused on research and his personal generosity made my time at University of Alberta enjoyable.

I would also like to thank Prof. Krzymien and Prof. Khabbazian for their valuable comments for improving the quality of the thesis.

My appreciation also extends to my lab-mates for their valuable support. Thanks also go to the departmental faculty and administrative staffs for their continued support throughout the period of my M.Sc.

I am indebted to my maa, Bani Banerjee and baba, Karunamoy Banerjee. Their value grows with age. They blessed me with my life and supported me through all my ups and downs.

Above all, thanks to Priyanka Ray (Jhil) for all her love and support. She is the constant source of light that illuminates me to get over all the dark spots of life, and makes me smile when nothing is going in my direction. This space falls short to acknowledge her, so I will just say, she is my greatest strength and she is the person without whom I would never be able to write this thesis.

Contents

List of Figures

List of Tables

Acronyms

Acronyms	Meaning
ICN	Information-centric Networking
CDN	Content Distribution Networks
SDN	Software Defined Networks
D2D	Device-to-device
PPP	Poisson Point Process
BPP	Binomial Point Process
SBS	Small-cell Base Station
RWP	Random Waypoint Model
RSS	Received Signal Strength
PDF	Probability Density Function
CDF	Cumulative Distribution Function
MGF	Moment Generating Function
DAG	Directed Acyclic Graph
LCE	Leave Copy Everywhere
LCD	Leave Copy Down
CL4M	Cache Less For More
LRU	Least Recently Used
GG	Gamma-Gamma
LS	Logarithmic Singularity
SNR	Signal-to-noise Ratio
BER	Bit Error Rate
CEP	Conditional Error Probability
BPSK	Binary Phase Shift Keying
DBPSK	Differential Binary Phase Shift Keying
M-PSK	M-ary Phase Shift Keying
LOS	Line-of-sight

Chapter 1

Introduction

1.1   Problem Statement

Although content caching is a promising technology to support the increasing demands of data rate for Fifth Generation (5G) wireless, there are several inherent challenges, such as developing an efficient caching strategy, and the impacts of user mobility on delay performance. Moreover, quantifying and modeling of interference in a wireless cache-enabled network is highly challenging due to the opportunistic nature of user association with base stations. To alleviate these challenges, in this thesis we propose a novel caching strategy, establish a mathematical framework for mobility analysis, and develop an asymptotic analysis for a generalized fading model to characterize the interference of a cache network.

1.2   5G Wireless Communication

1.2.1   Introduction

The rapid proliferation of mobile devices over the last five years has resulted in an 18-fold growth in mobile data. Cisco has predicted that mobile data traffic will increase to 49 exabytes by 2021 [1]. These requirements will challenge every technology in current communication systems, from protocols to internet architecture. Revolutionary changes will be needed to support the increased data traffic. Thus, 5G wireless [2] has

been proposed to provide very high data rates reaching Gigabytes per seconds, a 1000-fold increase over 4G systems. 5G systems are expected to be implemented by the year 2020 [2]. Fig. 1.1 represent the expected growth of mobile data traffic per month.
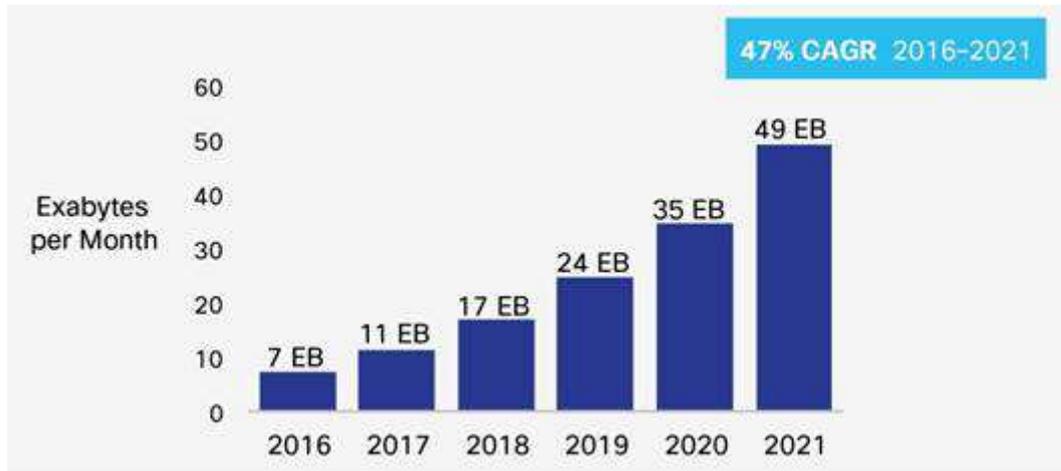
FIGURE 1.1: Mobile data traffic growth by 2021 [1]

5G systems are expected to support a wide range of applications, especially tactile navigation, self-driving cars, and virtual and augmented reality features. Primary applications of 5G are listed below, however, the list is by no means exhaustive.

- Mobile broadband

- Smart cities and smart homes

- Smart grids

- Health monitoring systems

- Augmented/virtual reality

- Tactile navigation systems

- Autonomous transport systems

- Machine type communications (MTC)

1.2.2   5G Requirements

Supporting such a broad spectrum of applications will require a dramatic improvement in a number of performance benchmarks. A brief parametric comparison between 5G and 4G is illustrated in Table 1.1 [3].

These revolutionary targets cannot be achieved by mere evolutionary steps. There are several technologies proposed for 5G, such as Massive MIMO, C-RAN, millimeter wave communication, cognitive radio, small-cell densification and cache-enabled networks. In this thesis, we integrate small-cell and cache-enabled networks and study different scenarios and their impacts on performance.

TABLE 1.1: 5G vs 4G

Parameter	Gain
Data rate	10-1000
Latency	1/10
Energy efficiency	100
Traffic	1000
Capacity	100

1.3 Wireless Cache-enabled Network

As the term suggests, a wireless cache-enabled network refers to a wireless system where caching abilities are enabled at certain nodes of the network. Generally, caching ability is incorporated at routers and/or base stations (BSs), and popular multimedia content is cached to serve user requests with a smaller amount of delay. Thus, content caching directly impacts the delay and throughput performance of a network and has the ability to offload traffic from servers to routers/BSs located closer to users.

1.4 Small-cell Networks

A small-cell network is an umbrella term for micro, pico and femto cells. The primary motivation behind small-cell architecture is the dense deployment of cells within a given geographical area to increase capacity and spectral efficiency while decreasing power usage [4]. Small-cell densification is highly effective in metropolitan environments where wireless data rates are significantly high. However, administration, organization and maintenance of small-cell networks are a challenge. Furthermore, backhaul links can act as a bottleneck for small-cell networks and limit their potential. Enabling caching abilities at BSs is a potential solution for this problem.

1.5   System Model

The architecture of a typical wireless cache-enabled network is illustrated in Fig. 1.2. At the bottom of the hierarchy, there are small-cell base stations (SBSs) or edge caches, which are connected with the users over a wireless medium. SBSs are connected to to the servers via a multi-layered backbone network. The backbone network consists of wired routers, termed core caches. For sake of simplicity let us assume that each node except users has caching abilities.
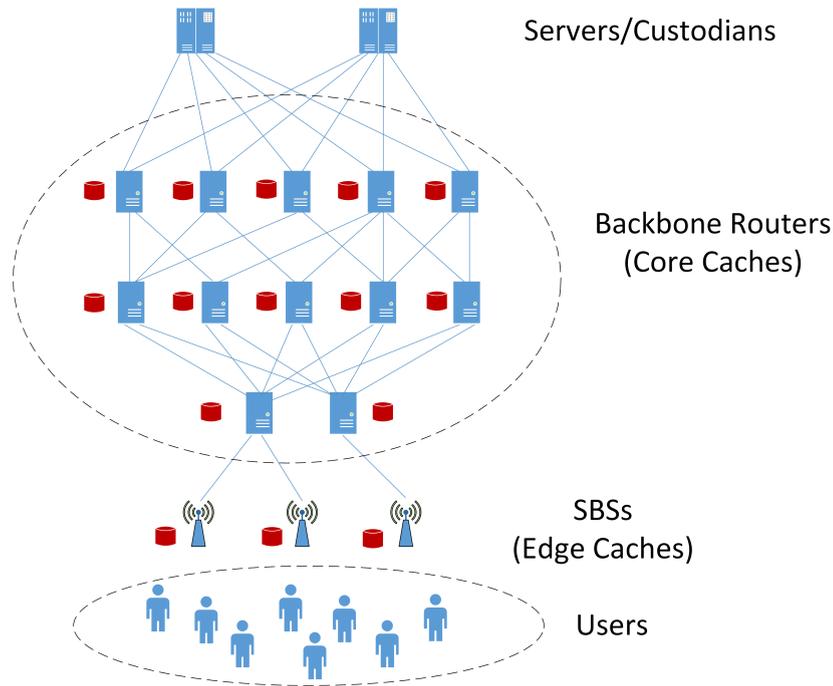
FIGURE 1.2: Considered Network Model

A content request from a user is initially forwarded to the associated SBS and the request is forwarded towards the server on the shortest path. The shortest path is determined using Dijkstra's shortest path algorithm. Enroute to the server, a request can be served by any of the underlying nodes if the requested content is cached at that node. The goal of a cache-enabled network is to reduce the download delay by serving the request from a cache. In this thesis, we consider special cases of this system model and study its user-level performance.

Firstly, the system in Fig. 1.2 is considered with the assumption that content request rate or content popularity at each SBS is available. An optimized caching policy is developed under this assumption. Intuitively it seems that caching most popular content at caches is the optimal solution, however, it actually increases content duplication at

caches, reduces cache utilization and does not provide an optimal solution at all. Several papers in the literature suggest caching most popular content at edge caches, however, this approach does not utilize the core routers efficiently. In our proposed optimized solution, primary emphasis is given to core routers to utilize them efficiently to maximize the number of requests served by them. Secondly, a network with only edge caches is considered and how user mobility affects the performance is studied.

1.6   Contributions and Outline

The main contributions of the thesis are listed as follows:

- *Greedy Caching*, a caching policy that greedily caches the most popular content at each cache based on their relative content popularity is proposed. Performance of *Greedy Caching* is studied via extensive simulations in Icarus [5], a simulator built exclusively for implementing and testing new ICN routing and caching policies to demonstrate the efficacy of *Greedy Caching*. Performance is compared against state-of-the-art caching and routing policies.

- A stochastic geometry-based analytical framework is developed for a K-tier small-cell network with varying transmission power and caching ability at each tier. A random waypoint mobility model is considered to characterize user mobility, and to study the effects of user mobility on delay performance.

- A new asymptotic performance measure for wireless channels is proposed. It includes a logarithmic term, which leads to a generalization of the classical diversity gain and coding gain relationship. We derive new asymptotic expressions for BER and outage probability. Closed-form solutions of BER are derived for different modulation schemes and antenna diversity models to obtain numerical results.

The outline of the thesis is as follows:

- Chapter 2: This chapter contains the necessary background information. Various caching strategies for both wired and wireless network, mobility models and handover management and wireless channel modeling are briefly discussed.

- Chapter 3: This chapter proposes and analyzes an optimized caching policy. First, an optimization problem to maximize cache hit rate is discussed and thereafter the proposed greedy solution is discussed with simple test cases. Furthermore, performance of the proposed strategy is studied for various real-life network topologies and scale-free networks.

- Chapter 4: This chapter investigates the effect of user mobility over delay performance for a cache network. The expression for handover probabilities and delay is derived, and performance for different handover management policies are discussed.

- Chapter 5: This chapter develops a new asymptotic measure for wireless channels that includes channel models with logarithmic singularity. Different modulation schemes and antenna diversity models are considered and the performance of the new asymptotic measure is compared with the existing ones.

- Chapter 6: This chapter presents the conclusions of the thesis and future research directions.

Chapter 2

Background & Motivation

This chapter provides some brief mathematical background and the key concepts used in this thesis. These include caching strategies for wired and wireless systems, terminologies for a cache-enabled network, spatial distribution and mobility models and wireless channel characterization.

2.1 Content Caching

The explosive increase in content in recent years has lead to the proposal of a new internet architecture called information-centric networking (ICN) which aims to evolve the current internet from a host-centric model to a content-centric one. By caching content at storage-enabled network nodes, requests for content can be served from the content custodians (origin servers), as well as from intermediate caches. With the primary emphasis being on content, if a cache enroute to the server has the requested content, the content will be returned to the user from the cache itself, thereby improving user performance. Serving a request from an intermediate cache has several benefits such as reduced content download delay, increased throughput and decreased network congestion.

Video delivery companies (e.g., YouTube, Netflix) already use simple forms of popularity based in-network caching in today's Internet to improve user performance. These video delivery applications primarily determine the popularity of multimedia content based on parameters such as release date, viewership of past series of a show and push popular content to the network edge [6, 7]. In recent years, caching policies proposed

for ICN have also identified that caching popular content within the network is essential to improve performance [8]. However, existing policies focus mainly on the network edge and fail to effectively leverage caches in the network core [9]. In ICN, deployment of network-wide caches is likely to be expensive. Therefore, it is important to design efficient caching and routing policies that maximize cache utilization, both at the network edge and in the network core and minimize unnecessary content duplication. While it is tempting to think that determining what content to cache at a node only requires local information, content cached at downstream nodes drastically impacts the request stream seen by upstream caches and may ultimately reduce network-wide cache utilization.

2.1.1 Caching Strategies

Various caching strategies have been proposed in the literature, both in the context of Content Delivery Networks (CDN) as well as ICN. In this section, we mainly focus on existing literature in ICN and demonstrate how *Greedy Caching* differs from prior work. Some of the most widely accepted caching policies are LCD [10], CL4M [11], and ProbCache. All these caching policies aim to reduce cache redundancy by caching content based on parameters such as content popularity and connectivity of nodes. A modified version of LCD with chunk caching and searching (CLS) is proposed in [12], where a piece of content is cached one level downstream or upstream depending on whether a request is a cache hit or a cache miss. Similarly, a modified version of Prob-Cache, namely ProbCache+ [13] incorporates a new variable called cache weight to enforce fairness between content.

PopCache [14] primarily uses content popularity to determine whether to cache a particular content or not. Authors in [15] propose a caching strategy ProbPD, where the dynamic popularity of a content determines its caching probability. This dynamic popularity is calculated by incorporating the distance of a cache from a user and the incoming content request for a certain time interval. In MPC [16], content popularity is dynamically calculated locally at each cache by maintaining a popularity table. Topology-dependent caching strategies have also been proposed in the literature. Authors in [17] develop a caching strategy called a Progressive Caching Policy (PCP), where content is cached at the node one hop downstream of the serving node and at another intermediate node that has number of incoming links greater than a threshold. Wang *et al.* propose CRCache [18], a caching strategy based on the correlation between content popularity

and network topology information. Hop-based Probabilistic Caching (HPC) [19] probabilistically caches content depending on the distance between the user and the cache.

Badov *et al.* propose a caching strategy to avoid congested links by caching content at the edge of a congested link. A cooperative caching strategy, where off-path caching is explored by controlling the routing algorithm is proposed in [20]. Authors in [21] develop a hash function-based joint routing and caching strategy that helps i) caches decide whether or not to cache a particular content and, ii) routers route requests to relevant caches. Similarly, authors in [22] propose CPHR, a collaborative caching strategy which also uses hash functions. Each content is partitioned according to the hash function and these partitions are then assigned to network caches. Hash function-based strategies generally require centralized control which results in high overhead. To overcome the shortfalls of centralized control, distributed cache management (DCM) [23] was proposed to improve cache utilization by sharing holistic information about request patterns and cache configuration. In an earlier work, Banerjee *et al.* proposed a routing algorithm that leverages characteristic time information to forward content requests [24].

In contrast to existing literature, in this thesis, a caching policy that adopts a locally-optimal approach at each node to determine the set of content to be cached at each network node is proposed. *Greedy Caching* caches content at each node based on relative content popularity, which is calculated based on the request miss stream from downstream nodes. This approach not only maximizes hit rate at each network node, but it also increases cache utilization by reducing content duplication. Note that greedy algorithms for replica management in CDNs have been previously proposed in the literature [25–27]. However, these papers assume that the incoming request rate at each cache is available, and thereafter develop a greedy algorithm subject to several parameters such as cache size, distance from users, and access cost and do not consider the request miss streams from downstream nodes.

2.1.2   Wireless Cache-enabled Network

Primarily, content caching techniques, such as information-centric networking [28], named-data networking [29], and content-centric networking are restricted to wired backbone networks, and studies reveal approximately $30 - 50\%$ IP traffic offload [30]. To utilize these benefits, Golrezaei *et al.* in their seminal paper first explored the concept of caching content at femtocells to improve wireless video streaming experience [31].

Thereafter, fundamental performance parameters of a cache-enabled wireless network, such as SINR, capacity, outage probability are studied in [32–34].

Authors in [35] study a small-cell network with caching abilities, where locations of SBSs follow the Poisson point process (PPP). Their work mainly addresses which parameter has a greater impact on outage probability, cache size or SBS intensity. Authors in [36] analyzed the expected delay for a two file system, i.e., only two files are requested by the users. For obvious reasons, this does not reflect a realistic scenario, and an analysis based on multiple content files and content popularity skewness must be taken into consideration. In a more general work, Yang *et al.* derived expressions of ergodic rate, outage probability, throughput and delay for a K-tier heterogeneous network model [34]. There are several articles in the literature that study the effect of caching on channel capacity. An asymptotic analysis of required link capacity for multi-hop cache-enabled wireless networks is analyzed in [37]. Qiu and Cao developed an analytical framework to study the achievable capacity and request serving rate in a cache enabled wireless network [38]. In [39] authors study the relationship between caching and linear capacity scaling capacity in a backhaul-limited cooperative MIMO system. Capacity scaling laws for a cache-enabled wireless hybrid network is studied in [40].

2.1.3 Caching Strategies for Wireless Networks

Baştuğ *et al.* propose a machine learning based caching strategy for wireless networks [41], however, they did not consider the spatial randomness of the nodes. Considering the spatial randomness of nodes via the Poisson point processes, authors in [42] study optimal caching for cellular networks. Authors in [43] formulate a joint optimization problem for caching and user association to maximize the probability of serving a content request. Similarly, Malak *et al.* develop an optimize content caching in a D2D network to maximize the probability of finding a cache source [44]. In [45], authors develop a greedy algorithm-based optimal content placement strategy. They demonstrate that the optimal content placement problem can be solved using two parameters, file diversity gain and channel diversity gain. Authors in [46] compared hit-rate optimization and throughput optimization for probabilistic caching in a D2D network. Although these caching strategies are developed for wireless networks, effects of user mobility are ignored in these works.

2.2 Spatial Distribution Modeling

Analyzing the performance of a cellular network requires modeling the location of nodes, and it is challenging to determine a mathematical model to describe the randomness in node deployment. Modeling the spatial randomness of nodes is especially important for interference characterization and mobility analysis [47, 48]. To capture spatial randomness, point process is an effective statistical tool. Point process is a collection of points located in a measured space, which for a cellular network is a $d-$dimensional Euclidean space with $d \geq 1$. There are two popular point processes for modeling node deployment in a wireless network, namely the Poisson point process (PPP) and the Binominal point process (BPP). In Chapter 4 we use PPP to model the initial locations of the nodes in a 2-dimensional space.

2.2.1 Poisson point process

Poisson point process is the most popular spatial distribution method, and is used in several foundational works [47, 49, 50]. PPP can be classified as homogeneous and non-homogeneous, where homogeneous PPP follows constant intensity in a given area and non-homogeneous PPP models intensity as a function of the location. Formally PPP can be defined as follows.

Definition 1. *A point process $\Phi = \{X_i, i \in \mathbb{N}\}$ over an area $\mathcal{A}$ with expectation measure $\mu(\cdot)$ is defined as a PPP if*

1. *$\Psi(\mathcal{A})$, number of points in area $\mathcal{A}$ follows a Poisson distribution with a mean $\mu(\mathcal{A})$ for every area $\mathcal{A}$.*

2. *For any m disjoint sets $\mathcal{A}_1, \cdots, \mathcal{A}_m$, the random variables $\Psi(\mathcal{A}_1), \cdots, \Psi(\mathcal{A}_m)$ are independent.*

A homogeneous PPP follows uniform intensity λ such that $\mu(\mathcal{A}) = \lambda l(\mathcal{A})$, where $l(\mathcal{A})$ is the Lebesgue measure (i.e., size) of the area $\mathcal{A}$ [47]. Therefore, for a 2-dimensional homogeneous PPP, the probability of having n nodes in area $\mathcal{A}$ is given by

$$P\left(\Psi(\mathcal{A}) = n\right) = \frac{(\lambda \, l(\mathcal{A}))^n}{n!} e^{-\lambda \, l(\mathcal{A})} \tag{2.1}$$
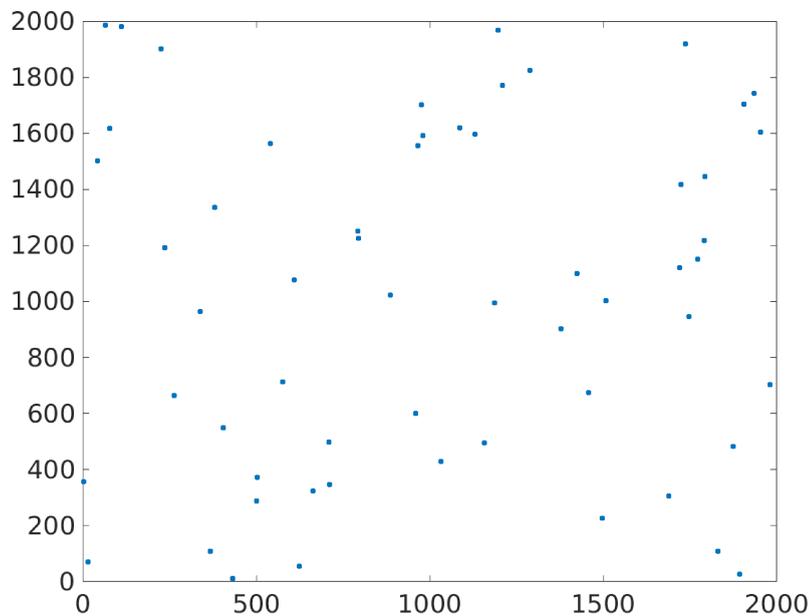
FIGURE 2.1: Homogeneous PPP with $\lambda = 2 \times 10^{-5}$

Fig. 2.1 illustrates the distribution of points over a 2-dimensional space following PPP. Increasing the intensity of λ (2.1) increases the number of points in an area. Therefore, to model a cellular network with BSs and users, different λ values are used. Now consider a small-cell network with several types of BSs, e.g., macro base stations (MBSs), femto base stations (FBSs), pico base stations (PBSs). Simply using PPP with different λ values for each type of BS is inefficient due to possible overlapping locations, and therefore, we need to use thinning to model this scenario [51].

Thinning is a process to discard some of the points from a set of points. Once the points are removed, a thinned point process is obtained. For a thinned point process, each point in $\Psi(\mathcal{A})$ is independently removed (or kept) with probability p (or $1 - p$). Therefore, thinning can be used to determine the location of different types of BSs. Thinned PPP is employed in Chapter 4 to develop an analytical framework for a small-cell network.

2.2.2 Binomial point process

Although PPP is the most popular spatial model due to its tractability and simplicity, it is inefficient for a small network with a given number of nodes. Binomial point process (BPP) is a popular choice for such cases [51]. For a given area $\mathcal{A}$ and fixed number of nodes K, a BPP is set of independent and identically distributed (iid) random variables

12

$\{x_1, \cdots, x_K\}$, each with PDF f. The intensity measure is $\rho = n \int_{\mathcal{A}} f(x) dx$, and number of points in $\mathcal{S} \in \mathcal{A}$ are binomially distributed,

$$P(N(\mathcal{S}) = n) = \binom{K}{n} \rho^n (1 - \rho)^{K-n}. \tag{2.2}$$

2.3 Mobility Analysis in Wireless Communication

Since mobility is a key attribute of wireless systems, several analytical models are available in the literature [52–54]. However, according to [55], human movement has extremely complicated spatial and temporal correlations, and it is extremely difficult to develop a precise analytical model. Nevertheless, Lin *et al.* developed an analytical model where nodes are initially modeled as Poisson spatial randomness and their mobility is modeled considering transition lengths to be Rayleigh distributed [56]. According to their study, the results match well with real-life trace results. Therefore, in this thesis, the model in [56] is utilized to characterize user mobility. Generally, mobility models can be classified into individual mobility models and group mobility models. In an individual mobility model, the mobility pattern of a node in the network is considered, whereas in a group mobility model, several nodes form a group and move in synchrony [57]. Two of the most popular individual mobility models are discussed below.

2.3.1 Random walk mobility

A random walk can be considered as a sequence of random variables $\{S_i \mid i = 0, 1, 2, \cdots\}$ that obeys the Markovian property,

$$P(S_{t+1} = y \mid S_0 = x_0, S_1 = x_1, \cdots, S_t = x_t) = P(S_{t+1} = y \mid S_t = x_t) . \tag{2.3}$$

The characteristics of a RW model can be summarized as follows [57]:

- Nodes change their velocity (v) and direction (θ) on each movement, and the pause time (t_p) between two movements is zero.

- Each node chooses v randomly from a predefined range $[v_{max}, v_{min}]$ for each movement, where v_{max} and v_{min} are the maximum velocity and minimum velocity, respectively.

- Each node chooses a new θ uniformly from the ranges $[0, 2\pi]$.

- Each movement occurs with either a constant time interval t or with a constant distance traveled d.

- A node bounces off the simulation boundary by an angle of θ or $(\pi - \theta)$ if it reaches the boundary in a movement.

2.3.2 Random waypoint mobility

The RWP mobility model is another popular mobility model which overcomes the limitation of zero pause time in the RW mobility model. In RWP, a node moves from its current position to a new location based on several parameters, velocity (v), transition length ($\mathcal{L}$) and direction (θ). Once the node reaches its destination, it stops for a random pause time (T_p). Parameters v and θ follow the same characteristics as in the RW model, whereas a different distribution for $\mathcal{L}$ is available in the literature. In [56] and Chapter 4 of this thesis $f_{\mathcal{L}}(l)$ is considered to follow the Rayleigh distribution. Formally, RWP for node j can be defined as an infinite sequence of tuple:

$$\{P_i^j, V_i, \mathcal{L}_i, T_{p,i}\}_{i \in \mathbb{N}} = \{P_1^j, V_1, \mathcal{L}_1, T_{p,1}\}, \{P_2^j, V_2, \mathcal{L}_2, T_{p,2}\}, \cdots, \tag{2.4}$$

where P_i denotes $i-$th waypoint. An additional waypoint P_0 is required to initialize the location of the node and it can be obtained from the spatial distribution of the nodes. The vector $\left(p_{i-1}, p_i, v_i, l_i, t_{p,i}\right)$ defines the $i-$th movement period completely in the RWP model.

2.3.3 Mobility Analysis in Cache Networks

Recently, [58] analyzed the effect of mobility on the coverage probability of a device-to-device communication with caching ability. Their work is primarily focused on developing interference analysis in a mobile environment and considers a simplified mobility model. Apart from mobility-related analysis, a caching strategy based on a user's mobility pattern is developed in [59]. Their caching strategy is based on cell sojourn time, i.e., the expected time of a mobile user staying in a cell. Wang *et al.* developed an optimization problem to maximize the data offload ratio in mobile D2D networks where each device caches content depending on its velocity [60]. In general, the authors concluded that high and low velocity users should cache high popularity content

14

and medium velocity users should cache low popularity content to reduce data duplication and maximize the data offload ratio. However, both these works require solving of multiple optimization problems and might result in additional network overhead. Moreover, existing works have yet to analyze effective handover management policies. We can understand that handover management in a wireless cache-enabled network is different from that in a traditional network, as it deals with incomplete downloads from a cached source.

2.3.4 Handover / Mobility Management

Mobility management policies for 5G can be classified into distributed mobility management and centralized mobility management. Authors in [61] describe a centralized policy and also a local or distributed policy. For the centralized policy, a local access server controls the handover between two SBSs. Whereas for the localized handover management, SBS manages handover events by using a local access sever as a mobile anchor. Giust *et al.* formally defined a management policy without using a mobility anchor and re-establishing a new connection [62]. Authors in [62] also analyzed two other distributed mobility management policies using mobile IP protocols and SDN based mobility management.

2.4 Wireless Channel Modeling

Analysis of the relevant performance metrics such as outage and error rates is contingent upon the proper statistical modeling of the wireless channel. For instance, statistical wireless channel models are used to design and optimize transmitters and receivers and their antenna configurations, to determine performance limits and to perform many other wireless system design tasks [63]. Characterization of a wireless channel depends on several channel impairments [64], including:

- path loss

- small-scale fading

- shadowing

In the following sections, these impairments are briefly described.

2.4.1 Simplified Path Loss Model

Attenuation of a signal as it propagates from the transmitter to the receiver is defined as path loss. Apart from the distance between the receiver and the transmitter, several other factors affect path loss, such as the transmitter and the receiver heights, atmospheric conditions and the physical properties of the antennas and the signal frequency. Therefore, modeling the path loss is a complex task. Several complex path loss models, COST 231 [65], Okumura [66], and Hata [67] models are presented in [64]. For analytical simplicity the following simplified path loss model is extensively used [64],

$$P_R = P_T k \left(\frac{r_0}{r}\right)^\eta,$$
(2.5)

where P_R, P_T are the received and transmitted powers and r_0 and r are the reference distance for the antenna far-field and the distance between the transmitter and the receiver, respectively. The constant η is known as the path loss exponent and varies depending on the environment.

2.4.2 Small Scale Fading Models

Small scale fading is the attenuation of signal amplitude due to multipath propagation. It is primarily influenced by rapid changes in signal strength over a small travel distance or time interval, random frequency modulation due to varying Doppler shifts due to user mobility and time dispersion caused by multipath propagation delays. Small scale fading can be characterized by various mathematical models. Several popular fading models are discussed below.

2.4.2.1 Rayleigh Fading

The simplest fading model from the analytical characterization perspective is known as Rayleigh fading. Rayleigh fading occurs in the absence of a line of sight (LOS) signal between the transmitter and the receiver. For a Rayleigh channel, received signal follows the Rayleigh distribution and the probability density function (PDF) of the received signal power follows exponential distribution. So PDF of the received signal

power is given by,

$$f(\beta) = \frac{1}{\overline{\gamma}} e^{-\beta/\overline{\gamma}}, \quad 0 \le \beta < \infty \quad , \tag{2.6}$$

where $\overline{\gamma}$ is the average SNR. It should be noted that while the Rayleigh distribution denotes the envelope amplitude, the power is specified by an exponential distribution.

As mentioned earlier, Rayleigh fading assumes unavailability of the LOS path between the transmitter and the receiver. This assumption is valid for scenarios such as a mobile user in an urban environment communicating with a base station, where the line of sight signal propagation is often blocked due to surrounding buildings.

2.4.2.2 Rician Fading

Whereas Rayleigh fading assumes an absence of LOS propagation, Rician fading is considered when a dominant LOS component is present in the signal. Therefore, the Rician fading model is expecially useful for satellite links. The PDF of the received signal power is given as follows:

$$f(\beta) = \frac{K+1}{\overline{\gamma}} \exp\left(-K - \frac{\beta(1+K)}{\overline{\gamma}}\right) I_0\left(2\sqrt{\frac{\beta K(K+1)}{\overline{\gamma}}}\right), \quad 0 \le \beta < \infty \quad , \tag{2.7}$$

where K is the ratio between the power in the LOS path and the power in the scattered path, and $I_0(\cdot)$ is the modified Bessel function of the first kind.

2.4.2.3 Nakagami-m Fading

Based on empirical measurements, a Nakagami-m model has been proposed [68]. This covers Rayleigh fading as a special case too. PDF of the received signal power in a Nakagami-m fading channel is given by:

$$f(\beta) = \frac{\beta^{m-1}}{\Gamma(m)} \left(\frac{\beta}{\overline{\gamma}}\right)^m e^{-\frac{m\beta}{\overline{\gamma}}}, 0.5 \le m < \infty, \quad 0 \le \beta < \infty \quad , \tag{2.8}$$

where $\overline{\gamma}$ is the average SNR, parameter m describes the level of fading and covers several fading models, for example, $m = 1$ gives Rayleigh fading, and $m = \infty$ denotes no fading scenario.

2.4.3 Shadowing Models

Shadowing is the fluctuation of received signal power due to blockage from large obstacles in the propagation path between transmitter and receiver. Unavailability of several parameters, such as distance from the transmitter, size and the dielectric properties of the obstacle leads to statistical modeling of shadowing.

2.4.3.1 Log-normal Shadowing

The most popular shadowing model is the log-normal shadowing model. The PDF of the ratio between transmit and receive power ψ for a log-normal shadowing model is given by [64]

$$f_\Psi(\psi) = \frac{10/\ln(10)}{\sqrt{2\pi}\psi\sigma_{\psi dB}} \exp\left[-\frac{\left(10\log_{10}(\psi) - \mu_{\psi dB}\right)^2}{2\sigma_{\psi dB}^2}\right], \quad 0 \le \psi < \infty \quad , \tag{2.9}$$

where $\mu_{\psi dB}$ and $\sigma_{\psi dB}$ are mean and standard deviation of ψ in decibels respectively.

As (2.9) is not mathematically tractable, several approximations are proposed, Gamma mixture [69] and K-distribution [70] are two of them.

2.4.3.2 Composite shadowing and fading models

Apart from separate models for fading and shadowing, there are several channel models available in literature that incorporate the effects of both shadowing and fading. Two examples are the Rayleigh-lognormal model and Nakagami-lognormal model [71]. Considering Rayleigh fading and approximating log-normal shadowing by Gamma distribution, author in [72] derived generalized-K distribution. The PDF of received signal power for generalized-K distribution is given by:

$$f(\beta) = \frac{2}{\Gamma(\lambda)}\left(\frac{\lambda\beta}{\Omega_s}\right)^{\frac{1+\lambda}{2}} K_{\lambda-1}\left(2\sqrt{\frac{\lambda\beta}{\Omega_s}}\right), \quad 0 \le \beta < \infty, \tag{2.10}$$

where $\lambda = \frac{1}{e^{\sigma^2}-1}$, $\Omega_s = \sqrt{\frac{\lambda+1}{\lambda}}$, and σ is the variance of the log-normal shadowing, respectively. $K_n(\cdot)$ stands for the modified Bessel function of second kind with order n.

In spite of numerous notable works over years, an important characteristic of wireless channels, logarithmic singularity (LS), has been previously overlooked. Now a fading

18

channel can be classified as LS wireless channel if its PDF can be expressed in the form of $f(\beta) = a\beta^t + b\beta^\mu \log(\beta) + \cdots$ near $\beta = 0$. The importance of analyzing LS channels is manifested by its versatility, as LS property is observed in popular generalized fading models that cover composite fading and shadowing models, e.g., Gamma-Gamma (GG) channel and generalized-K channel.

The GG distribution was introduced as a more flexible model than the K-distribution [73]. It covers Gamma and K-distribution, and the Nakagami-lognormal composite fading model as special cases. GG distribution is also used to model multiple communication scenarios, such as a single point-to-point channel with co-channel interferences [74–77], relay networks with amplify-and-forward (AF) and decode-and-forward (DF) strategies [78], wireless optical channels [79] and radar systems [72]. We realize that the flexibility of GG model makes it a strong contender as a channel model for wireless cache-enabled networks, and we therefore develop an asymptotic analysis for LS channels.

2.4.4 High-SNR Analysis

A very effective and common method to achieve simple yet direct and insightful analytical expressions for fading channels is to develop asymptotic or high signal-to-noise ratio (SNR) analysis. Throughout the thesis, unfaded SNR or average SNR is denoted by $\overline{\gamma}$. Thus, analysis in the region characterized by $\overline{\gamma} \to \infty$ allows us to generate simpler analytical expressions in general. The fading SNR of the received signal can be expressed as $\gamma = \beta\overline{\gamma}$, where β is the random variable in the PDF $f(\beta)$. Useful asymptotics are then developed by extracting the first term of the Taylor series expansion of the PDF $f(\beta)$ near $\beta = 0$. Following this approach, outage probability and BER asymptotics were developed by Wang and Giannakis [80]. To do so, they approximated $f(\beta)$ as $f(\beta) = a\beta^t + R_{t+1}$, for $\beta \to 0^+$, where R_{t+1} is a remainder term that vanishes for $\beta \to 0^+$. $a > 0$ and $t \geq 0$ are the two parameters that determine the SNR (coding) gain and diversity gain. The intuition of their work is that since the asymptotic performance may be given by $\int_0^{1/\overline{\gamma}} g(\beta)f(\beta)d\beta$ where $g(\cdot)$ is a rapidly decreasing function, what matters is a simple but accurate asymptote of $f(\beta)$ near $\beta = 0$. Moreover, they observed that a classical coding and diversity gain model, given by (2.11) can be expressed as a function of a and t.

$$P_e(\overline{\gamma}) = (G_c\overline{\gamma})^{-G_d} + R(\overline{\gamma}) \tag{2.11}$$

where $R(\overline{\gamma})$ is the remainder term that vanishes as $\overline{\gamma} \to \infty$, and G_c and G_d are called the coding gain and the diversity gain, respectively, and are important, widely-used parameters that are useful for wireless system design and optimization. For instance, from (2.11), we observe that on a log-log scale, $P(\overline{\gamma})$ varies linearly with $\overline{\gamma}$, which is a directly insightful representation of the system performance. Error probability of a fading channel using Wang and Giannakis' approach is given by, $P_e(\overline{\gamma}) = c/\overline{\gamma}^{t+1}$ where c and t are constants depending on the fading model. Therefore, G_c and G_d in (2.11) are $c^{-1/(1+t)}$ and $t + 1$. Coding and diversity gain, G_c and G_d can be determined from MGF of the channel model as well [81, 82]. Similarly, error probability of fading channels for Trellis coded modulations is derived using MGF of the channel [83]. MGF of a channel can be used to characterize interference of a wireless network too [84–87].

Since the seminal paper by Wang and Giannakis [80], several incremental works have been published [88–92]. Authors in [88] used two terms from the Taylor series expansion of $f(\beta)$ and approximate it as an expansion of an exponential function. Dhungana and Tellambura used a Mellin transform-based approach over the asymptotic expansion of the PDF, given in (1) to derive a uniform approximation [89] that works for both low and high SNR regimes. Authors in [90, 91] combined a dual exponential sum with the asymptotic model in (1), where the argument of the exponential functions depends on the fading channel model. In [92] authors employed a characteristic function-based approach to derive error probabilities. Annamalai *et al.* used a characteristic function of $f(\beta)$ along with Parseval's theorem to calculate the average error probability [92].

However, the approach used by Wang and Giannakis does not hold for LS channels. Later in the thesis, in Fig. 5.1, we also show how a Taylor series-based asymptote diverges from the GG PDF. Consequently, the models derived using the Taylor series fail to closely approximate the GG fading channel too. Thus, in terms of asymptotic analysis, LS wireless channels are fundamentally different from all the existing fading models. Therefore, in this thesis we also develop an effective asymptotic analysis for LS channels.

Chapter 3

Greedy Caching: A Latency-aware Caching Strategy

3.1 Introduction

In this chapter, *Greedy Caching*, a simple caching policy that determines the optimized set of content to be cached at each network node based on the relative content popularity, with the goal of reducing content download delay (referred to as latency) is proposed. *Greedy Caching* estimates the relative content popularity at each node based on the request stream from directly connected users as well as the request miss stream from downstream nodes and then uses a greedy algorithm to determine the content to be cached. The difficulty of the problem stems from the fact that different pairs of network nodes can forward requests to one another, resulting in interdependencies, and cycles in the underlying graph, thereby making it difficult to estimate the relative content popularity.

The main contributions of this chapter are given below.

- Assuming that the network has an underlying routing policy for forwarding requests for content towards the custodian, *Greedy Caching*, a caching policy that greedily caches the most popular content at each cache based on their relative content popularity is proposed. To estimate relative content popularity, *Greedy Caching* first leverages routes provided by the routing algorithm to create a directed acyclic graph (DAG). For the single custodian case, DAG construction is

relatively straightforward. However, for the multiple custodian case, simply combining the routes provided by the routing algorithm results in a cyclical graph, due to node pairs sending traffic to one another. *Greedy Caching* therefore uses the feedback arc set algorithm to prune this cyclical graph and construct a DAG. *Greedy Caching* then combines the request stream from users with the constructed DAG to determine the set of content to be cached at each network node, starting from the network edge and ending at the custodians.

- Extensive simulations are performed in Icarus [5], a simulator built exclusively for implementing and testing new ICN routing and caching policies to demonstrate the efficacy of *Greedy Caching*. Then the performance of *Greedy Caching* is compared against state-of-the-art caching and routing policies, Leave Copy Everywhere (LCE) [93], Leave Copy Down (LCD) [10], Cache Less for More (CL4M) [11], ProbCache [94], and Random Caching (Random) [95] on real world internet topologies (e.g., GARR, GEANT, and WIDE) [96]. We study the impact of various simulation parameters (e.g., cache size, content-universe, content popularity skewness) on the performance of *Greedy Caching* and demonstrate that it provides approximately 5-28% improvement in latency and 15-50% improvement in hit rate over state-of-the-art strategies.

The rest of the chapter is organized as follows. First, the problem and the proposed *Greedy Caching* algorithm is described in Section 3.2. Second, Experimental results are presented in Section 3.4 and finally, the chapter is concluded in Section 5.5.

3.2 Problem Statement

3.2.1 Network Model

Let us consider an ICN which is represented by an undirected graph $\mathcal{G}(V, E)$, where V consists of all the nodes in the network including the users, caches and custodians and E consists of the set of interconnected links. Assume $\mathbb{U} = \{U_1, U_2, ..., U_N\}$, $\mathbb{R} = \{R_1, R_2, ..., R_M\}$ and $\mathbb{C} = \{C_1, C_2, ..., C_L\}$ to denote the set of users, caches and custodians respectively. Therefore, the network comprises of N users, M caches and L custodians. Also assume that each cache has the same amount of finite storage, C.

The content universe $F = \{f_1, f_2, ..., f_K\}$ is uniformly distributed among the custodians. Each piece of content is available at only one custodian and is permanently stored there. Let F_j be the set of content stored at the jth custodian. Let us consider that content popularity varies according to user and follows a certain probability distribution (e.g., Zipf). User U_i generates requests for content at rate $\Lambda_i = \{\lambda_{i1}, \lambda_{i2}, ..., \lambda_{iK}\}$. Let us consider that these requests are forwarded toward the custodian depending on the underlying routing strategy (e.g., Dijkstra's shortest path routing). Let Λ_i also denote the outgoing request rate at any intermediate network node i (apart from the users). The incoming request rate at node i is denoted by Λ_i'. Note that Λ_i and Λ_i' can differ due to caching at node i.

Let $\mathcal{P}_{ij}(V_{ij}, E_{ij})$ denote the shortest path from U_i to the C_j with V_{ij} denoting the set of nodes on that path and E_{ij} denoting the set of directed edges tracing the path from U_i to the C_j. Additionally, for all edges $e_{ij} \in E_{ij}$ connecting nodes i and j, an indicator variable, I_{ij}^k is set to 1 if e_{ij} lies on the shortest path for content k, otherwise it is set to 0. We assume that the shortest path algorithm returns $\mathcal{P}_{ij}(V_{ij}, E_{ij})$ and also sets I_{ij}^k. Note that each request can traverses multiple caches en route to the custodian. If a cache en route to the custodian has the requested content, the cache serves it, otherwise the content is served by the custodian.

3.2.2 Caching Problem

Let x_{km} be a binary variable that denotes if the kth piece of content is cached at the mth node (including users, caches and the custodian). It takes a value 1 if the content is cached and 0 otherwise. For users x_{km} will always take values 0 while for a custodian x_{km} will take a value 1 for content that is housed at the custodian and 0 otherwise. Let $\mathcal{H}$ denote the hit rate for all user requests. A request is said to be a hit if it is served by any in-network cache apart from the custodian. For ease of representation, let us assume that for each path $\mathcal{P}_{ij}(V_{ij}, E_{ij})$ between user U_i and custodian C_j, $(U_i, n_1, \cdots, C_j)$ comprises of the topological ordering of all vertices in V_{ij}. A content request for f_k will be served by node n_l in path $\mathcal{P}_{ij}$ only if the request is not served by any of the preceding nodes n_i. Formally, we denote "$n_l > n_i$" to represent that n_l comes after n_i in the ordered list $\mathcal{P}_{ij}$. Using the above notation, the hit rate can be expressed by (3.1). The equation takes into account the traffic for different content from the N users and the fact that if the requested content is cached on any node between the user and custodian, then it results in a hit.

$$\mathcal{H} = \frac{1}{\displaystyle\sum_{i=1}^{N}\sum_{k=1}^{|F|}\lambda_{ik}} \sum_{i=1}^{N}\sum_{k=1}^{|F_j|}\sum_{j=1}^{L}\sum_{\substack{m\in\mathcal{P}_{ij}-C_j: \\ C_j>m>U_i}}\prod_{\substack{l\in\mathcal{P}_{ij}: \\ m>l\geq U_i}}\lambda_{ik}x_{km}\left(1-x_{kl}\right) \tag{3.1}$$

Now the problem of maximizing the hit rate as an optimization problem is expressed as:

$$\begin{aligned} \max \quad & \mathcal{H} \\ \text{s.t} \quad & \sum_{k=1}^{|F|} x_{km} \leq C \ \ \forall m \\ & x_{km} \in \{0,1\} \end{aligned} \tag{3.2}$$

The goal is to present a solution to the optimization problem presented above. As the objective function is non-linear, a solution to the optimization problem cannot be easily obtained using a solver. Therefore, in the next section, we present *Greedy Caching*, an optimized content placement policy for ICN that aims to maximize the hit rate. *Greedy Caching* adopts a greedy approach that maximizes the hit rate at each network node and eliminates unnecessary content duplication. The algorithm estimates the relative content popularity at each node based on the miss request stream from downstream nodes. *Greedy Caching* then employs a simple greedy algorithm that caches the most popular content at each node based on the relative content popularity at that node. We demonstrate in Section 3.4 that *Greedy Caching* outperforms state-of-the-art caching policies both in terms of network metrics such as hit rate and link load as well as user-facing metrics such as latency.

3.3 Proposed Solution

In this section, at first a simple example to motivate the need for *Greedy Caching* and to illustrate the importance of caching content based on the concept of relative content popularity instead of absolute content popularity is presented. Then the concept of relative content popularity is leveraged to design the *Greedy Caching* algorithm.

3.3.1 Motivating Example

Let us consider a network of 3 users (U_1, U_2, U_3), 2 caches (R_1, R_2), and one custodian (C_1) as shown in Fig. 3.1(a). Let us assume that the delay on each link is 1 second. Consider that there are only two unique pieces of content A and B, with probability of requesting content A and B being 0.6 and 0.4 respectively. Let us assume that R_1 and R_2 can cache only one piece of content and all users generate requests at same rate λ.

If content is cached based on absolute popularity, then this will result in content A being cached at both R_1 and R_2. At first glance, this appears to be a good idea, but if one considers the miss request stream from R_1 to R_2, the total incoming rate at R_2 for content A and B is 0.6λ and 0.8λ respectively. This is referred as *relative content popularity*, which is calculated at each node based on the miss request stream from downstream nodes. From this discussion it is clear that it is better to cache content A at R_1 and content B at R_2. In fact, for this simple network this is the optimal caching policy. Caching content A at R_1 and content B at R_2 decreases overall content download delay to 1.47 seconds in comparison to 1.67 seconds when content A is cached at both R_1 and R_2.

3.3.2 Greedy Caching

In this subsection, the details of the *Greedy Caching* algorithm, that leverages this concept of relative content popularity is presented. Estimating the relative content popularity at network nodes lies at the heart of *Greedy Caching*. At the highest level, the *Greedy Caching* algorithm starts by caching the most popular content at the network edge and then iteratively determines the content to be cached at the nodes in the network core by estimating the relative popularity. This iterative process stops when all network nodes have been visited. We first discuss *Greedy Caching* for the relatively simple scenario of an ICN with a single custodian and then move on to the more challenging multiple custodian case. Estimating the relative content popularity, especially for the multiple custodian case is non-trivial because of the interdependencies arising from pairs of network nodes forwarding requests to one another.

3.3.2.1 Single Custodian

To determine the relative content popularity with respect to a cache, *Greedy Caching* first combines the routes provided by the underlying shortest path routing algorithm for all users to generate a directed acyclic graph (DAG) $\Psi(V', E')$, where V' and E' are the number of vertices and edges in the DAG respectively. As there is only a single custodian in the network, it is easy to observe that combining these paths will result in a DAG. This is because if a node (say R_1) forwards requests through a node (say R_2) toward the custodian, R_2 lies on the shortest path from R_1 to the custodian. Therefore, R_2 cannot route the requests it receives through R_1. For each node i in Ψ, let $\mathbb{N}'_i$ denote the set of neighbors from which there is an incoming edge to i.

Greedy Caching then performs a topological sort on $\Psi(V', E')$ to determine Θ, an ordering of the vertices in Ψ. Let Θ_i denote the ith vertex in Θ. For a DAG, topological sort provides a linear ordering of the vertices such that for every directed edge from vertex u to vertex v, u comes before v in the ordering. It is evident that the the users and the custodian will be first and last nodes in this topological ordering. *Greedy Caching* then visits the nodes in order.

Greedy Caching then caches the set of content with the highest incoming request rate (i.e., the content with the highest relative popularity). Note that nodes at the network edge will only have incoming edges from the users and thus will be the first group of nodes visited by the algorithm. Therefore, the *Greedy Caching* algorithm will cache the C most popular content at each edge node. Now, these nodes at the network edge will only forward requests for uncached content along their outgoing edges as determined by the routing algorithm. As a result, any node v, which appears in the topological ordering after the edge nodes will take into account the request stream from directly connected users and the request miss stream from nodes that appear earlier than it in the ordering to calculate the relative popularity. Node v will thus cache the C most popular content based on the relative content popularity. Details of the *Greedy Caching* algorithm for a single custodian are provided in Algorithm 1.

Let us now revisit Fig. 3.1(a) and see how *Greedy Caching* ends up caching content A at R_1 and content B at R_2. For this network, the DAG obtained by combining the shortest paths will be similar to the network itself and is given in Fig. 3.1(b). The topological sort is given by $U_1, U_2, U_3, R_1, R_2, C_1$. Therefore, the algorithm visits R_1 first and caches A. Accounting for the miss stream from R_1 to R_2 and the request stream from U_3, it is easy to see that *Greedy Caching* will cache content B at R_2.

(a) Considered network   (b) Resulting DAG
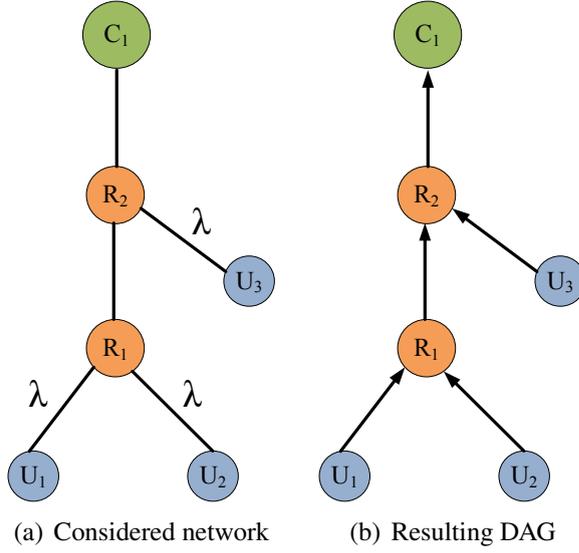
FIGURE 3.1: Greedy Caching illustration for single custodian

Algorithm 1 Greedy caching for single custodian

1: **Input** network $\mathcal{G}(V, E)$
2: **for** $U_i \in \mathbb{U}$ **do**
3:      $\mathcal{P}_{i1}(V_{i1}, E_{i1}) = \textbf{ShortestPath}(U_i, \mathbb{C}_1)$
4: **end for**
5: $\Psi(V', E') = \bigcup\limits_{i=1}^{N} \mathcal{P}_{i1}(V_{i1}, E_{i1})$
6: $\Theta = \textbf{TopologicalSort } \Psi(V', E')$
7: **procedure** GREEDY CACHING(Θ, $\mathbb{R}$)
8:      **for** $i = 1, i \leq |V'|, i + + $ **do**
9:          $r = \Theta_i$
10:          **if** $r \in \mathbb{R}$ **then**
11:              **for** each content k **do**
12:                  $\lambda'_{rk} = \sum_{j \in \mathbb{N}'_r} I^k_{jr} \lambda_{jk}$
13:              **end for**
14:              $\Lambda'_{r_{sort}}$: Sort Λ'_r in descending order
15:              Cache top C content in $\Lambda'_{r_{sort}}$
16:              **for** each content k cached at r **do**
17:                  $\lambda_{rk} = 0$
18:              **end for**
19:          **end if**
20:      **end for**
21: **end procedure**

3.3.2.2   Multiple Custodian

The multiple custodian scenario is more challenging, primarily due to the fact that simply combining the shortest paths from the users in the network may result in a cyclic

(a) Considered network   (b) Cyclic graph   (c) Resulting DAG
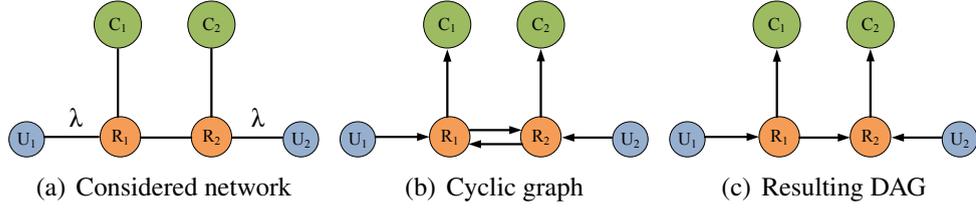
FIGURE 3.2: Greedy Caching illustration for multiple custodian

graph ($\mathcal{G}'$), as shown in Algorithm 2. Let us consider Fig. 3.2(a) to understand this. In this network, there are two users U_1 and U_2, two caches R_1 and R_2 and two custodians C_1 and C_2. Let us assume that one half of the content universe F_1 is available at C_1 and the other half F_2 is available at C_2. As is evident from the figure, R_1 will forward requests for content in F_2 from U_1 to R_2 and R_2 will forward requests for content in F_1 from U_2 to R_1. Combining the shortest paths will result in a cyclic graph as shown in Fig. 3.2(b).

Greedy Caching attempts to eliminate this problem by leveraging the feedback arc set algorithm [97, 98] that provides the set of edges to be removed from the cyclic graph $\mathcal{G}'$ to create a DAG $\Psi(V', E')$. Therefore, applying the feedback arc set algorithm to Fig. 3.2(b) will result in Fig. 3.2(c). Fig. 3.2(c) demonstrates DAG construction for the multiple custodian network given by Fig. 3.2(a). Once the DAG is constructed, the relative content popularity at each node is determined in a manner similar to Algorithm 1.

Algorithm 2 Greedy caching for multiple custodian

1: **Input** network $\mathcal{G}$
2: **for** $U_i \in \mathbb{U}$ **do**
3:      **for** $C_j \in \mathbb{C}$ **do**
4:          $\mathcal{P}_{ij}(V_{ij}, E_{ij}) = \textbf{ShortestPath}(U_i, \mathbb{C}_j)$
5:      **end for**
6: **end for**
7: $\mathcal{G}' = \bigcup\limits_{i=1}^{N} \bigcup\limits_{j=1}^{L} \mathcal{P}_{ij}(V_{ij}, E_{ij})$
8: Apply feedback arc set over $\mathcal{G}'$ to generate $\Psi(V', E')$
9: $\Theta = \textbf{TopologicalSort } \Psi(V', E')$
10: **procedure** GREEDY CACHING($\Theta, \mathbb{R}$)
11: **end procedure**

Also note that *Greedy Caching* may not always provide the optimal solution for both single and multiple custodian scenarios. However, it will be observed in the following section that *Greedy Caching* performs well in practice and outperforms state-of-the-art policies. The primary reason for its superior performance, especially for user facing

metrics such as latency is that it tries to maximize hit rate by pushing content closer to the users and thus lowers overall latency.

3.4 Performance Evaluation

In this section, first describe the experimental setup and then present simulation results. Then the performance of *Greedy Caching* is compared with state-of-the-art caching and routing policies, namely Leave Copy Everywhere (LCE), Leave Copy Down (LCD), Cache Less for More (CL4M), ProbCache, Random Caching (Random), and basic Hash Routing (HR). These strategies are explained below.

(a) $\alpha = 0.6$

(b) $\alpha = 0.8$

(c) $\alpha = 1.1$
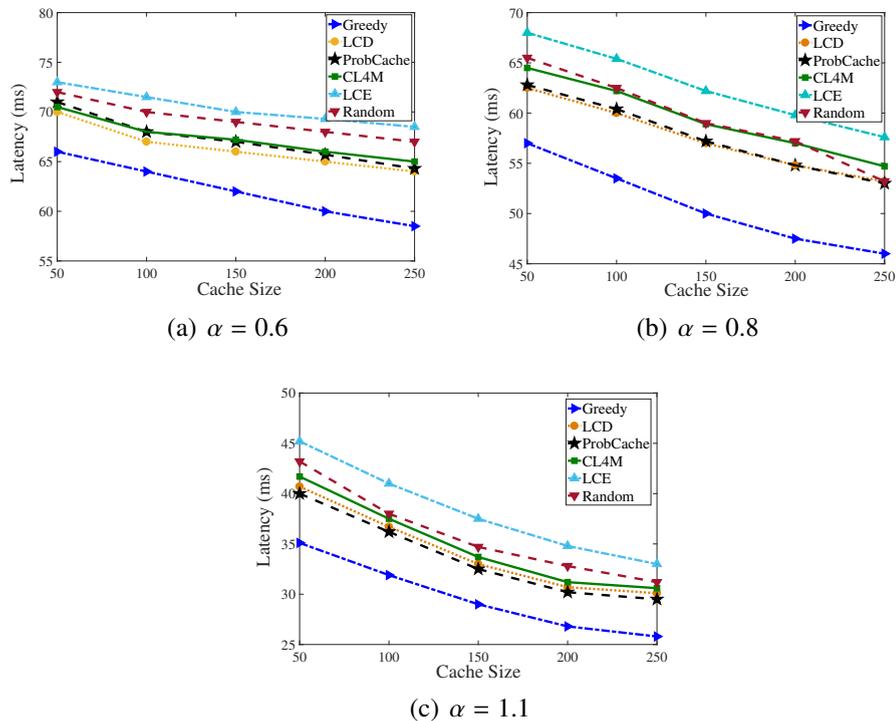
FIGURE 3.3: Latency for GARR with single custodian

- **LCE:** In this strawman approach, content is cached at every node along the path as it is being downloaded[93].

- **LCD:** In this policy whenever there is a cache hit, the content is replicated at the cache which is one hop downstream toward the requester [10].

- **CL4M:** This policy leverages the concept of betweenness centrality (i.e., the number of shortest paths traversing a cache) to make caching decisions [11]. This

29

(a) $\alpha = 0.6$

(b) $\alpha = 0.8$

(c) $\alpha = 1.1$
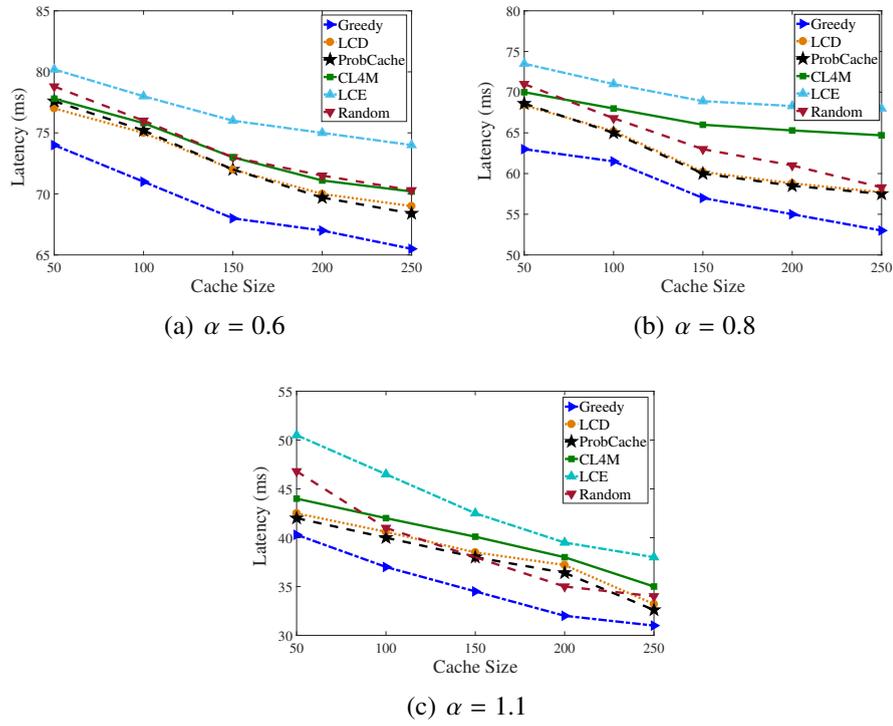
FIGURE 3.4: Latency for GARR with 2 custodians

(a) $\alpha = 0.6$

(b) $\alpha = 0.8$

(c) $\alpha = 1.1$
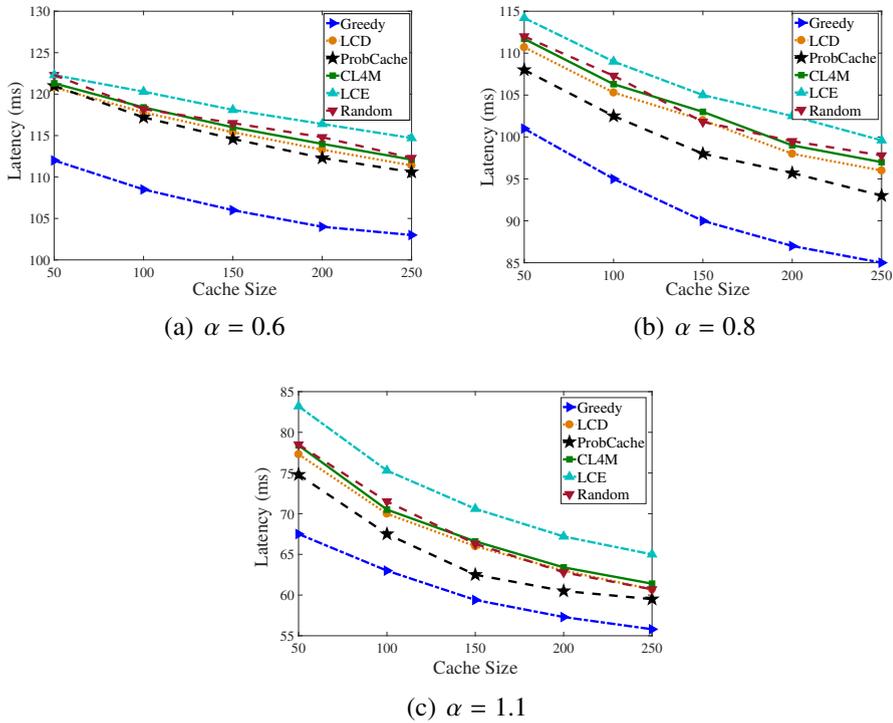
FIGURE 3.5: Latency for GARR with 5 custodians

policy caches content at nodes with the greatest betweenness centrality, so as to maximize the probability of a cache hit.

- **ProbCache:** This policy reduces cache content redundancy [94] by probabilistically caching content at en route caches.

- **Random:** In this caching strategy [95], content is cached at any one of the downstream node. So probability of caching content at a downstream node is inversely proportion to the path length.

- **Hash Routing:** In this approach [21], nodes at the network edge compute a hash function to map the content identifier to a specific cache upon receiving a content request. These edge nodes then forward the request to that particular cache.

(a) Single custodian

(b) Custodian = 2
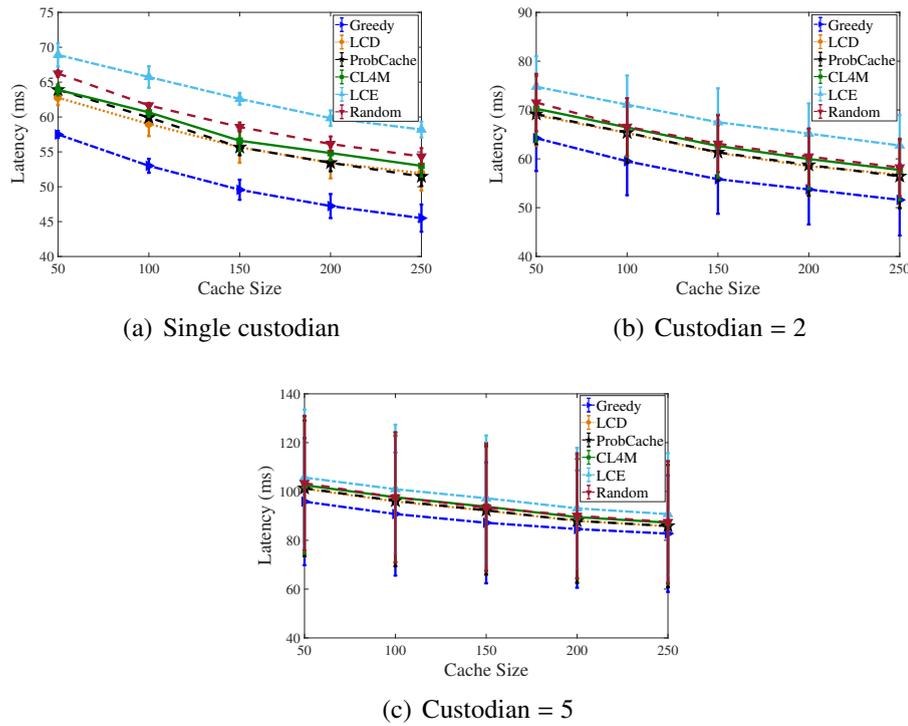
(c) Custodian = 5

FIGURE 3.6: Latency for GARR topology for random custodian selection

Assume that the network uses Dijkstra's weighted shortest path routing to route content requests to custodians, where weights corresponds to the delay on links. If content is found at a cache en route to a custodian, then it is served from that cache. We perform simulations on a discrete event based simulator Icarus [5], a simulator designed exclusively for ICN research. The simulator consists of four building blocks, scenario generation, experiment orchestration, experiment execution, and result collection. Each content request is considered as an event, and whenever an event occurs, a corresponding timestamp is stored. The result collection block gathers the results of the simulation.

(a) Single custodian

(b) Custodian = 2

(c) Custodian = 5
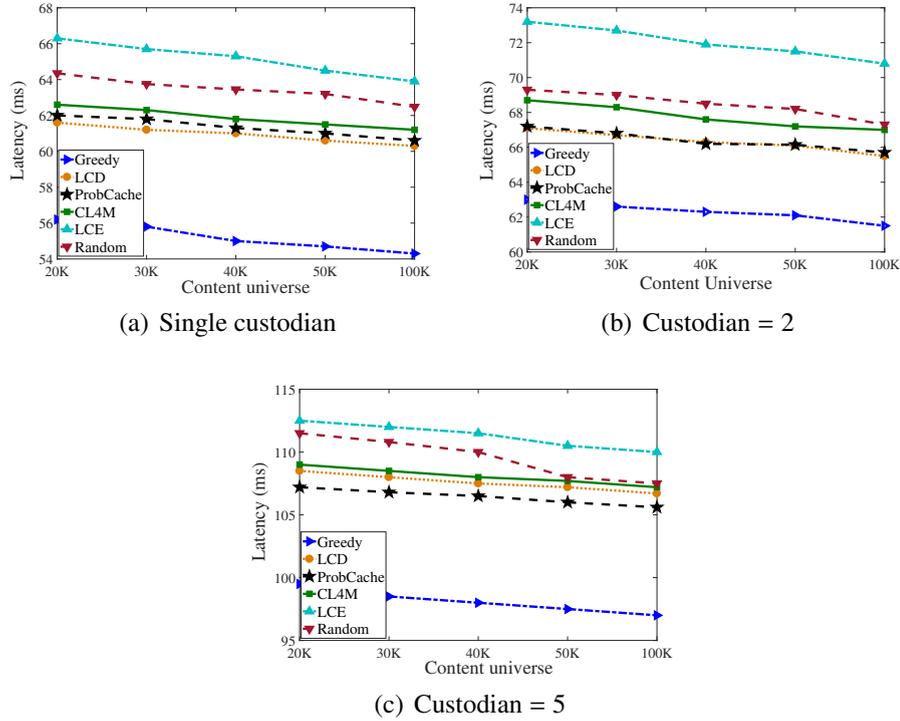
FIGURE 3.7: Latency for GARR topology for varying content universe

In Icarus, latency is calculated as the sum of delays on each link, traversed during content download.

Extensive experiments are performed on various real world networks namely GARR (Italian computer network), GEANT (European academic network), and WIDE (Japanese academic network). GARR is the Italian national computer network for universities with 61 nodes and 89 edges. GEANT is an academic network spread around the world consisting of 40 nodes and 61 edges. The WIDE topology is the first network established in Japan and consists of 30 nodes and 33 edges. To test the scalability of *Greedy Caching*, experiments on scale-free networks are performed as well. To avoid cluttering the chapter with figures of similar nature, unless mentioned otherwise, all results shown in this chapter are for the GARR topology.

In our simulations, prior to evaluation, all caches are warmed up. Except for large content universe, caches are always warmed up with 100000 requests and the subsequent 100000 requests are used for performance evaluation. Error bars in the figures are obtained over 5 runs of the experiment. We assume that the probability of requesting a content follows a Zipfian distribution with skewness parameter α. Nodes with the highest degree (i.e., number of connected links) are considered as custodians, and in case multiple nodes have the same degree, custodians are selected randomly among them.

For the multiple custodian case, content is uniformly distributed between the custodians and each content is permanently stored at only one of the custodians. Nodes with degree of 1 are selected as the users. Unless otherwise mentioned, all results are generated for the following simulation configuration; content universe $F = 10000$, cache size of each node is varied between $50 - 250$, and content popularity skewness (α) is varied between $0.6 - 1.1$.

Various performance metrics such as latency, hit rate, and average hop count are considered to demonstrate the superiority of *Greedy Caching*. In the results, performance improvement of using strategy *A* over strategy *B* is calculated by taking the percentage of the difference between the two strategies divided by the performance of strategy *B*. Our experiments demonstrate that *Greedy Caching* outperforms state-of-the-art strategies for a wide range of simulation parameters. Please note that results for Hash Routing (HR) are omitted from the figures due to aesthetic reasons, as in most scenarios its performance is poorer in comparison to the other strategies.

3.4.1 Latency Performance for GARR

In this section, the latency performance of *Greedy Caching* for single custodian and multiple custodian cases for GARR topology with variety of different settings is studied.

3.4.1.1 Impact of varying custodians

Fig. 3.3 demonstrates the latency results for the different policies for GARR for the single custodian case while Figs. 3.4 and 3.5 show the latency results for the 2-custodian and 5-custodian scenarios for different α values $(0.6, 0.8, 1.1)$. We observe from the figures that *Greedy Caching* outperforms state-of-the-art strategies by 7-22 % for a wide range of cache sizes. The primary reason behind the superior performance of *Greedy Caching* is better cache utilization, especially in the network core. An interesting point to note is that the overall latency increases for all strategies for the multiple custodian scenario. From simulations we observe that for the multiple custodian case, the custodians are further away from the users on average when compared to the single custodian case, with some shortest paths from users to a custodian traversing through another custodian. Internally in the simulator, links connected directly to a custodian have higher delay in comparison to the other links, thereby increasing the overall latency. We also

demonstrate in Section 3.4.4, that the overall hit rate reduces for multiple custodian network.

3.4.1.2 Impact of custodian placement

An important aspect dictating performance is custodian placement. Custodian placement determines the path traversed by user requests and thus directly influences the content cached within the network and the total cache utilization. Therefore, in this subsection, the impact of custodian placement on the performance of *Greedy Caching* is studied to demonstrate the widespread applicability of our results.

In a real world setting custodian placement is dependent on many factors (e.g., policy, cost), and thus we present performance results for a random custodian placement policy in Fig. 3.6 for α value of 0.8. A random custodian placement policy is implemented in Icarus simulator by randomly selecting the custodians from the set of nodes with degree greater than and equal to two. Different sets of custodians are selected using different seed values and the error bars in Figs. 3.6(a) - 3.6(c) are obtained for 5 different seed values. From the figures we observe that *Greedy Caching* outperforms other strategies by 7 - 29%. Another interesting observation from the figures is the increased variation in latency performance for the multiple custodian case. The primary reason behind this variation is the change in the average distance of the users from the custodians for different custodian placements.

3.4.1.3 Impact of content universe

We study the impact of varying content universe on the performance of *Greedy Caching* for GARR (Fig. 3.7). Results are generated for fixed content-to-cache ratio of 0.005, i.e., size of each cache is 0.5% of the content universe. Content universe is varied from 20000 to 100000. Simulation results suggest that *Greedy Caching* outperforms other strategies for large content universe as well. We observe from Fig. 3.7 that latency decreases as content universe increases. As the absolute cache size increases with content universe (the content-to-cache ratio is fixed), popular content is readily available in caches leading to increased cache utilization. As the primary purpose of Figure 3.7 is to demonstrate the scalability of our algorithm, the horizontal axis in Figure 3.7 is not plotted in linear scale to preserve the aesthetics of the figure. These results demonstrate that the *Greedy Caching* algorithm is efficient and can work with large catalogue sizes.

3.4.2 Impact of cache eviction policies

In this subsection, we study the performance of *Greedy Caching* for FIFO and random cache eviction policies in Figs. 3.8 and 3.9 respectively. Similar to the LRU cache eviction policy, we observe that *Greedy Caching* outperforms other caching strategies by 8 - 31%. An interesting observation is the superior performance of the Random cache insertion policy over ProbCache and CL4M. The possible reason behind the superior performance of the Random cache insertion policy is that both FIFO and random cache eviction policies do not take content popularity into account while evicting content, and thus ProbCache and CL4M that make caching decisions based on content popularity and for reducing content duplication exhibit poor performance.

(a) Single custodian

(b) Custodian = 2

(c) Custodian = 5
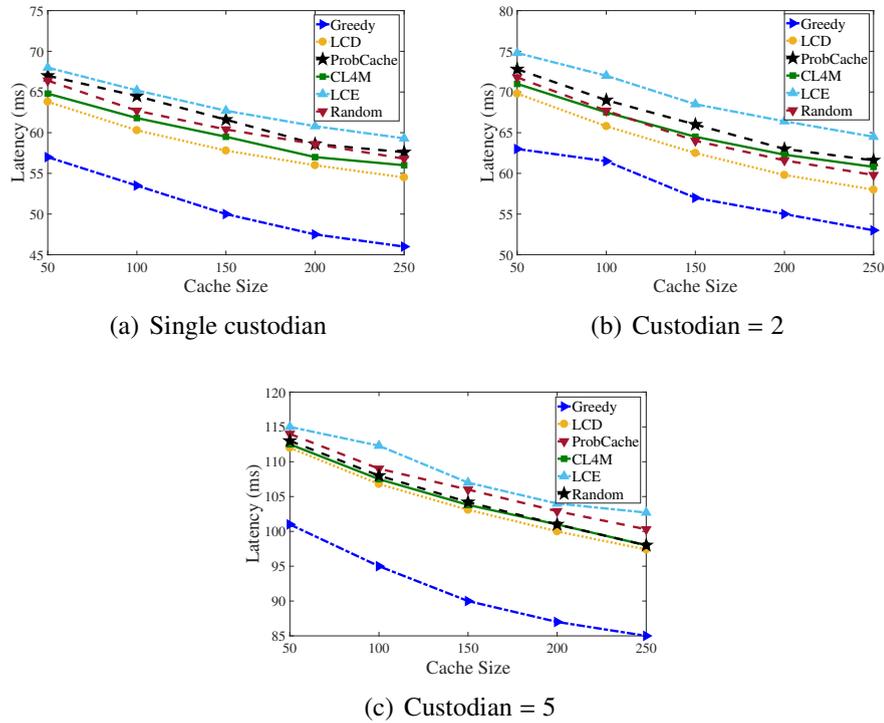
FIGURE 3.8: Latency for GARR for FIFO cache eviction

3.4.3 Performance for GEANT and WIDE topologies

Results for other topologies, e.g., GEANT and WIDE are shown in Figs. 3.10 and 3.11 respectively for $\alpha = 0.8$. From the results we conclude that *Greedy Caching* performs best among the existing strategies. For both GEANT and WIDE, we make observations similar to the GARR topology. In general, we observe a) an overall increase in latency

(a) Single custodian

(b) Custodian = 2
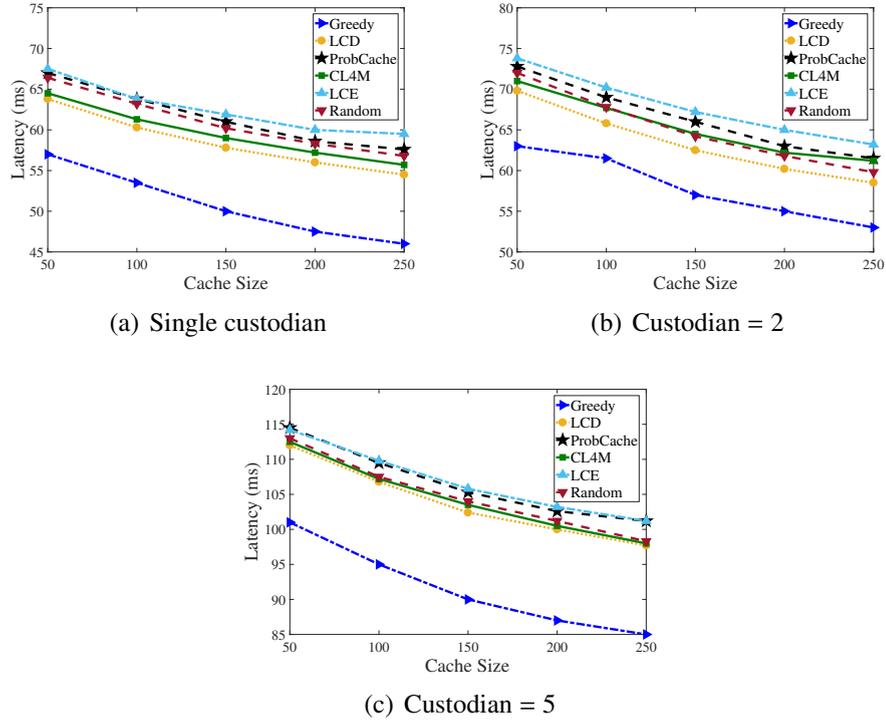
(c) Custodian = 5

FIGURE 3.9: Latency for GARR for Random cache eviction

with increasing number of custodians and b) decrease in latency for greater cache size and higher values of content popularity skewness.

From the simulations, we observe that the performance improvement for *Greedy Caching* varies between 9-25%, 10-28% and 5-20% for GARR, GEANT and WIDE respectively. We notice that *Greedy Caching* gives better performance for GARR and GEANT compared to WIDE. Primary reason behind this improved performance is that for WIDE, the edge to node ratio (1.1) is significantly lower in comparison to GEANT (1.46) and GARR (1.52). It means that multiple paths are not available to reach a node from a randomly selected user in the multiple custodian scenario. We observe from the simulations that in case of WIDE with multiple custodians, most paths to a custodian are via another custodian, as alternate paths are not available. This scenario reduces cache hit significantly and decreases the performance gains.

3.4.4 Discussion on cache hit rate

In addition to latency, hit rate is an important performance measure in ICN. If a request is served by a network cache, it is referred to as a hit and if it served by the custodian it is referred to as a miss. Therefore, higher hit rate suggests that more content requests
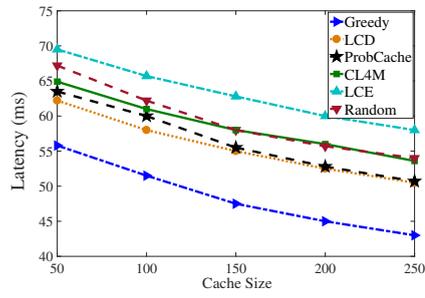
(a) Single custodian   (b) Custodian = 2
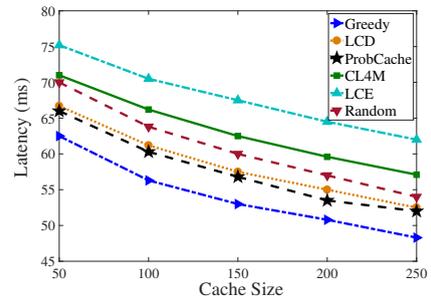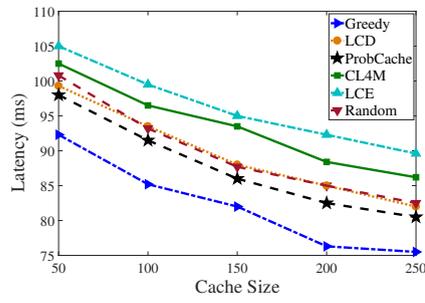
(c) Custodian = 5

FIGURE 3.10: Latency for GEANT
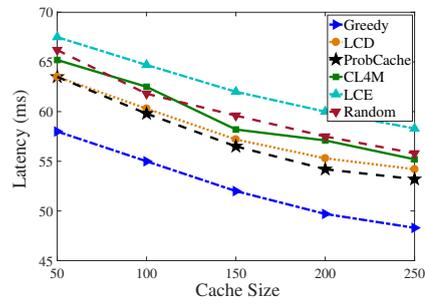
(a) Single custodian   (b) Custodian = 2
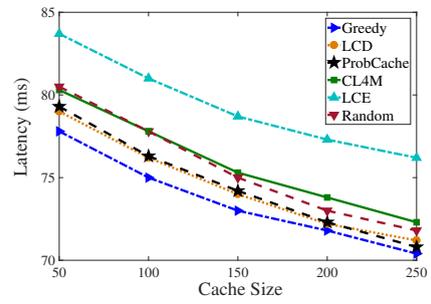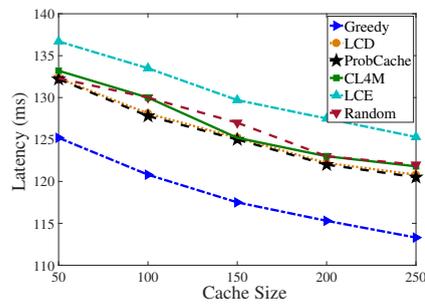
(c) Custodian = 5

FIGURE 3.11: Latency for WIDE

are served from network caches, thereby reducing load on the custodians. Hit rate can therefore be considered as a measure for both traffic offloading and cache utilization.

Figs. 3.12 and 3.13 show the hit rate for two different cache sizes 50 and 150 respectively for fixed $\alpha = 0.8$, for single and multiple custodian scenarios. Each sub-figure in Figs. 3.12 and 3.13 comprises of three topologies (GARR, GEANT, and WIDE). From the figures we observe that *Greedy Caching* improves hit rate by $15 - 50\%$. This can primarily be attributed to the intelligent content placement strategy adopted by *Greedy Caching* where every node caches content based on the total number of requests it receives from it downstream nodes.

We also observe that hit rate performance of *Greedy Caching* degrades with increasing number of custodians. This performance degradation can be attributed to the removal of edges by the feedback arc set algorithm to break cycles for DAG construction. Moreover, we also observe that hit rate increases with cache size and content popularity skewness. Greater cache size provides an increased opportunity to cache popular content, while higher value of the skewness parameter results in popular content being requested most of the time.

(a) Single custodian

(b) Custodian = 2

(c) Custodian = 5
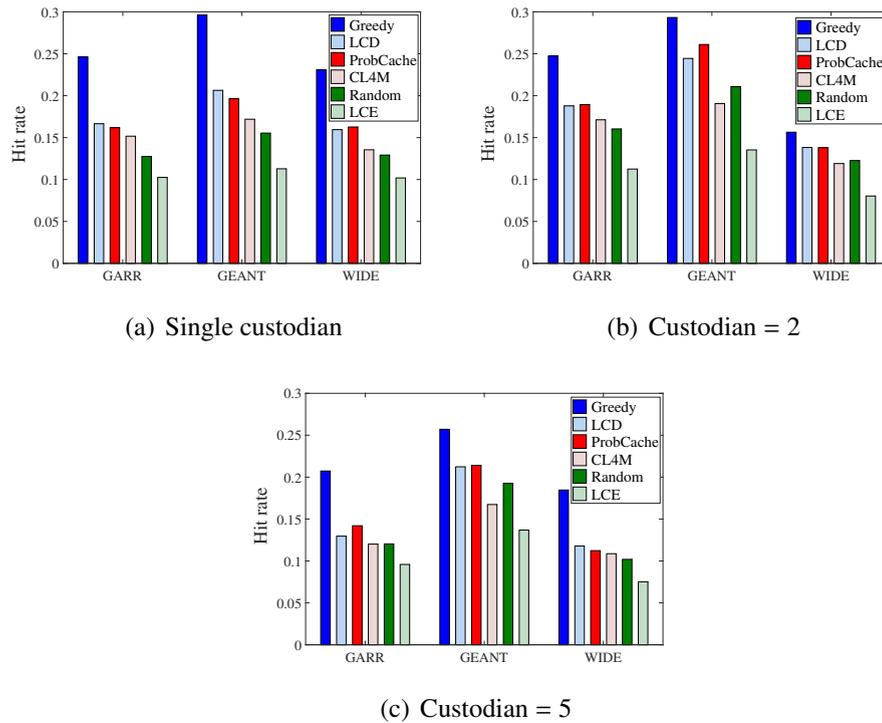
FIGURE 3.12: Hit rate for $\alpha = 0.8$, $C = 50$

(a) Single custodian

(b) Custodian = 2

(c) Custodian = 5
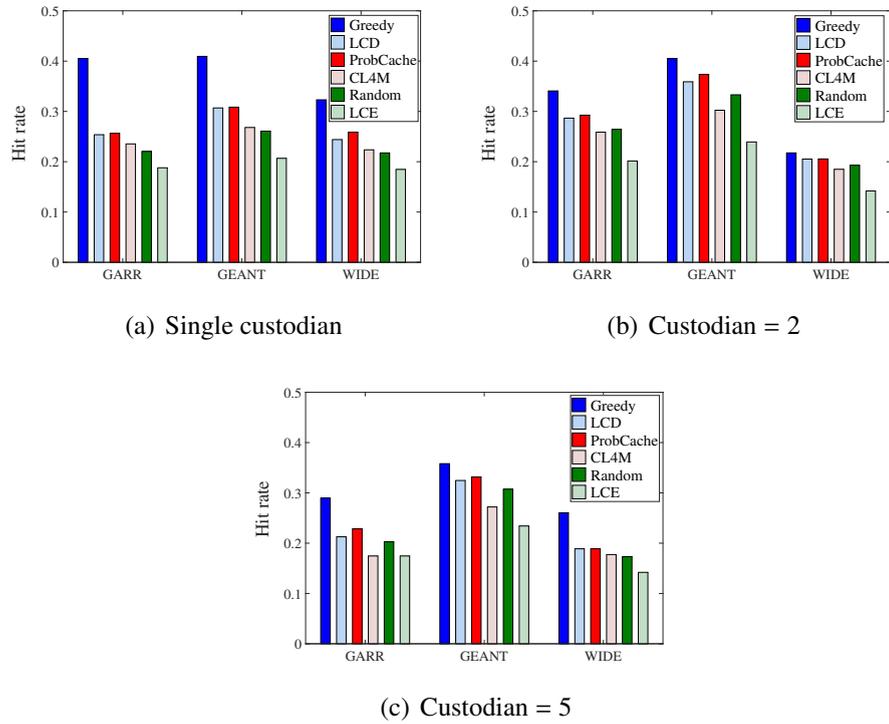
FIGURE 3.13: Hit rate for $\alpha = 0.8$, $C = 150$

3.4.5 Discussion on average hop count

Although hit rate provides an indication of the percentage of requests served from within the network, average hop count provides a measure of how far a request needs to travel to be satisfied. Figs. 3.14 and 3.15 show the average hop count results for single and multiple custodian cases. As expected, we observe that *Greedy Caching* has the least average hop count. As *Greedy Caching* focuses on the miss stream from downstream nodes to calculate the relative popularity of content at each cache, it makes superior caching decisions at both the network edge and the network core, thereby leading to better overall performance.

3.4.6 Discussion on average link-load

To investigate the impact of the different policies on network congestion, we study a metric called link-load. We calculate link-load of a link as a ratio of the total number of requests traversing the link to the total number of requests. Average link-load for the GARR topology with cache size of 50 and 250, and $\alpha = 0.8$ is tabulated in Table 3.1. Each entry in the table consists of mean and standard deviation of link load.
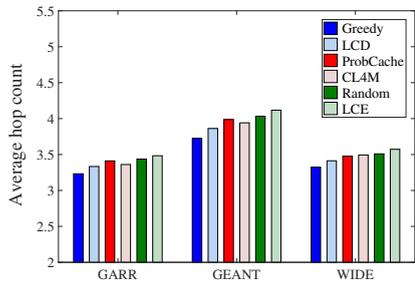
(a) Single custodian
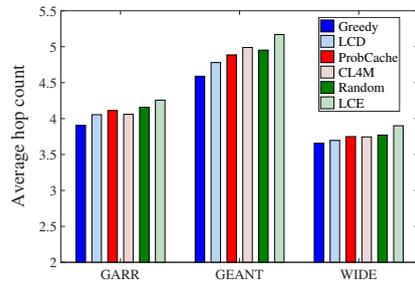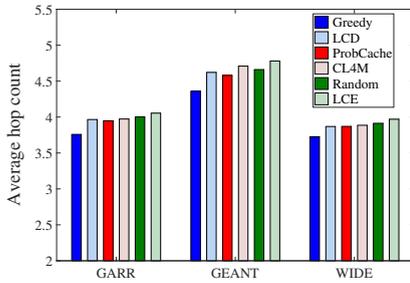
(b) Custodian = 2

(c) Custodian = 5

FIGURE 3.14: Average hop count for $\alpha = 0.8$, $C = 50$
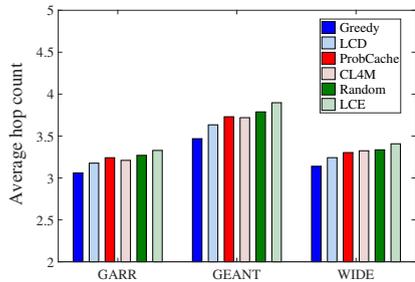
(a) Single custodian
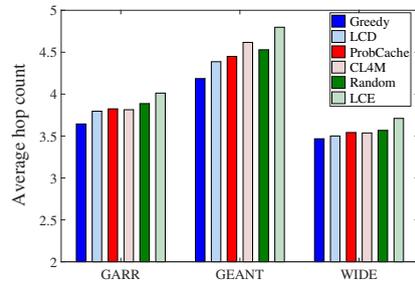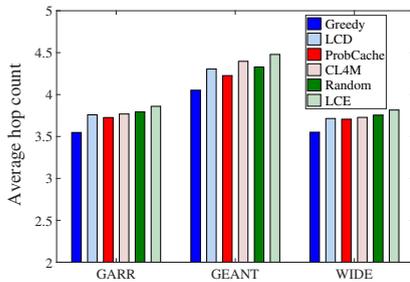
(b) Custodian = 2

(c) Custodian = 5

FIGURE 3.15: Average hop count for $\alpha = 0.8$, $C = 150$

We note that the average link-load depends on two parameters - the total number of links carrying traffic in the network and the average hop count. For example, as the

number of custodians increases, the total number of links carrying traffic between the users and the custodians increases as well. As the total traffic is distributed over a larger number of links in this case, average link-load is less for a larger number of custodians. We observe this as a general trend for all caching strategies.

We also observe a direct correlation between the average hop count and the average link-load. A smaller average hop count for a caching policy implies that a greater number of requests are served by caches located closer to the users and hence the number of requests traversing the network decreases. Note that as all caching strategies studied in the chapter follow shortest path routing, the total number of network links carrying traffic remains the same for all the caching strategies. Therefore, as the average hop count for *Greedy Caching* is smaller in comparison to the other strategies, its average link-load is smaller as well.

TABLE 3.1: Link load performance of *Greedy Caching* for GARR for $\alpha = 0.8$

Strategy	Custodian 2		Custodian 5	
	$C = 50$	$C = 250$	$C = 50$	$C = 250$
Greedy	0.058±0.027	0.05±0.016	0.048±0.013	0.043±0.009
LCD	0.061±0.03	0.053±0.019	0.052±0.017	0.046±0.011
ProbCache	0.062±0.032	0.053±0.018	0.051±0.016	0.046±0.01
CL4M	0.061±0.03	0.054±0.019	0.052±0.017	0.047±0.011
LCE	0.065±0.038	0.057±0.024	0.054±0.019	0.048±0.013
Random	0.063±0.034	0.055±0.02	0.053±0.017	0.047±0.011

3.4.7 Performance on Scale-free Network

To demonstrate the scalability of *Greedy Caching*, its performance on scale-free networks are investigated. Scale-free networks are generated using the Barabási-Albert preferential attachment model [99]. In the preferential attachment model, newly added nodes have a greater affinity to connect with nodes with a higher degree. In this subsection, results for scale-free networks comprising of 200 nodes are reported. Note that as network size increases, the length of the shortest path to reach the custodians also increases.

Figs. 3.16 and 3.17 show the latency for $\alpha = 0.8$ and varying cache size for the LRU and FIFO caching policies. It is observed that *Greedy Caching* continues to outperform state-of-the strategies (8 - 30%) for large networks as well. Next, hit rate results for scale free network is illustrated in Fig. 3.18 for $\alpha = 0.8$ and cache size of 50. We

(a) Single custodian

(b) Custodian = 2

(c) Custodian = 5
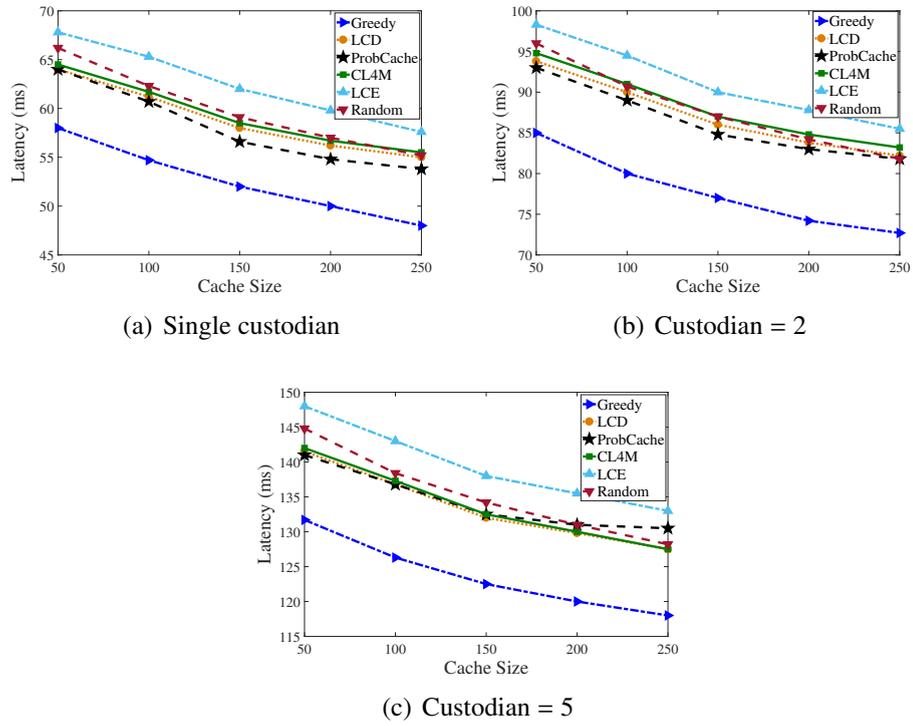
FIGURE 3.16: Latency for a 200 node scale-free network with with LRU cache eviction

(a) Single custodian

(b) Custodian = 2

(c) Custodian = 5
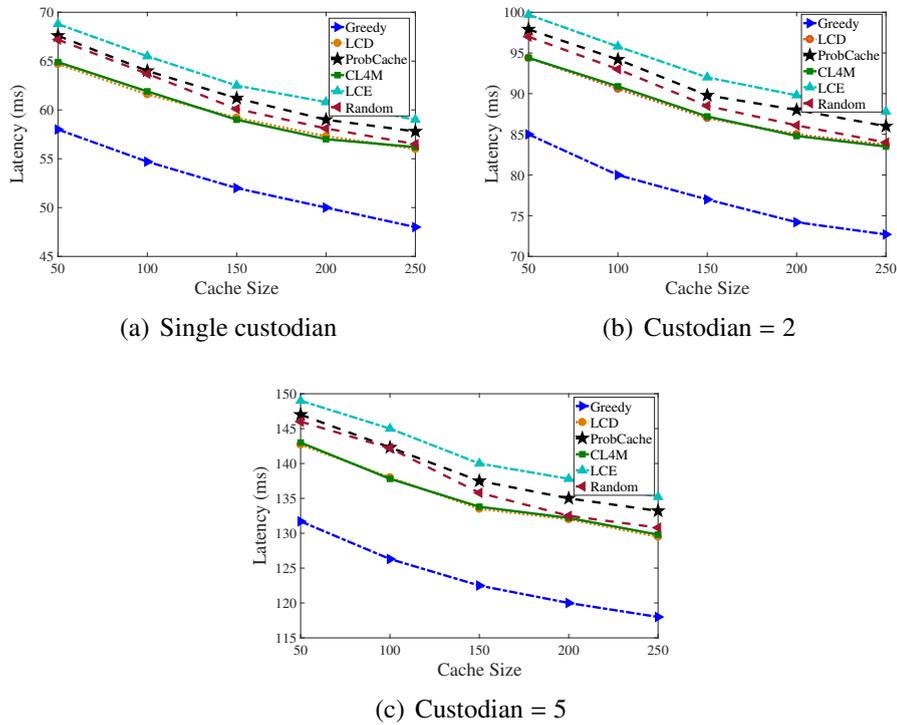
FIGURE 3.17: Latency for a 200 node scale-free network with FIFO cache eviction

observe that the hit rate is greater in comparison to smaller networks for the same set of network parameters (i.e., Fig. 3.12). The main reason is that as the length of the

shortest path to the custodian increases, the probability of obtaining the content at an en route router increases, and thus the hit rate increases.
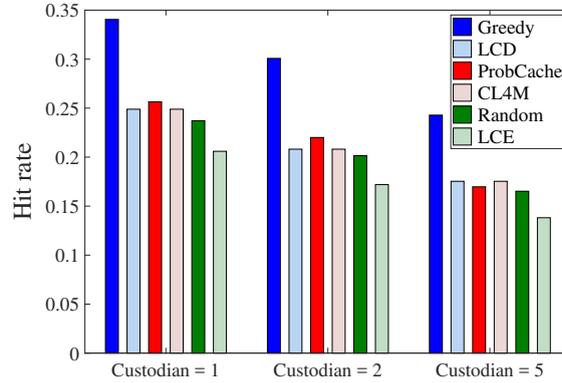
FIGURE 3.18: Hit rate for scale-free network

Results for the average link-load for cache sizes of 50 and 250 in Table 3.2 are discussed too. It is observed that the average link-load is lower in comparison to earlier results (i.e., Table 3.1). The total number of requests considered being the same, as the total traffic is distributed among a greater number of links in scale-free networks, the average link-load is smaller. Another interesting observation is the increased deviation in link-load in Table 3.2 in comparison to Table 3.1. As the number of links in scale-free networks is significantly higher than previously considered networks, traffic among the links vary significantly as well. This results in greater standard deviation in link load for scale-free networks.

TABLE 3.2: Link load for scale-free network for $\alpha = 0.8$

Strategy	Custodian 2		Custodian 5	
	$C = 50$	$C = 250$	$C = 50$	$C = 250$
Greedy	0.013±0.006	0.011±0.004	0.014±0.011	0.015±0.009
LCD	0.014±0.008	0.012±0.006	0.016±0.013	0.014±0.01
ProbCache	0.014±0.008	0.013±0.006	0.016±0.013	0.015±0.011
CL4M	0.014±0.008	0.012±0.006	0.016±0.013	0.014±0.01
LCE	0.014±0.009	0.013±0.006	0.017±0.015	0.015±0.011
Random	0.014±0.008	0.012±0.006	0.016±0.014	0.014±0.01

Simulations for different values of content request rate are also performed, and it is observed that request rate does not have a significant impact on latency. This result is expected and can be attributed to the design of the Icarus simulator. The simulator treats arrivals as an independent request stream and does not take into account increased network delays (due to congestion) because of increased arrival rates.

3.5 Conclusion

In this chapter, *Greedy Caching*, a caching policy for ICN that works with any underlying routing algorithm and determines the content to be cached at each network node is proposed. *Greedy Caching* adopts a greedy approach that considers the request miss stream from downstream caches to make caching decisions at upstream caches. Therefore, the algorithm attempts to maximize the hit rate at each individual cache. Via extensive simulations, it is shown that *Greedy Caching* significantly outperforms state-of-the-art caching and routing strategies.

Chapter 4

Mobility Analysis in Cache-enabled Small-Cell Network

4.1 Introduction

To alleviate increasing traffic, provide better coverage and user-level performance, a promising technology is small-cell densification, which refers to the deployment of large number of small-cell base stations (SBSs) over a region, thereby increasing spatial resource utilization [4]. Small-cell densification is an umbrella term for networks with femto cells, pico cells, and micro cells. The primary drawback of small-cell network is the bottleneck of backhaul link, particularly due to congestion in the link. Incorporating caching abilities in a small-cell network can alleviate the congestion problem. For example, assume that a popular video is downloaded by large number of users and it results in link congestion, however, caching this popular video at the SBSs and locally serving the requests for the video can significantly reduce the congestion. Authors in [100], developed a time-slot based analytical model to study optimal caching strategies for SBSs and user devices that support device-to-device (D2D) communication. Several application specific studies for cache-enabled small-cell and D2D networks are available in literature [101, 102]. This motivates us to select the network model for this chapter, and consider a small-cell network with caching abilities at SBSs.

Even though recent articles shed light on the performance of cache-enabled wireless networks, an important feature of most wireless systems, mobility of users remain unexplored. In terms of wireless content caching, we are interested in studying the

45

correlations between caching and mobility related parameters. For example, content downloading from a SBS can be greatly affected by user mobility. Due to its smaller coverage area, a user moving out of a SBS region may happen quickly. This move will result in a handover and incomplete data download. Thus, handover probabilities of cache-enabled systems in a mobile environment must be rigorously analyzed. To the best of our knowledge, the study of the effects of mobility over download delay in cache enabled wireless networks has not been reported. Hence, in this chapter we develop a mathematical model to explore the impact of user mobility on the performance of cache enabled wireless networks. We thus develop holistic relations between handover and transition length of the mobile device, cache size of SBS, size of content universe, and content popularity skewness, and study their cumulative effect on download delay.

4.1.1 Contribution

Specifically, the contributions of this chapter are listed as follows:

- A stochastic geometry based analytical framework for content-caching wireless networks is developed. A k-tier heterogeneous network model is considered with varying transmission power and caching ability at each tier. The caching ability of a SBS is assumed finite.

- Random waypoint mobility model is used to analyze the effects of user mobility. To this end, analytical expressions for different handover probabilities are derived, and also expressions for expected delay is derived as a function of mobility.

- Three handover management strategies based on network infrastructure are considered and expressions for expected delay for each of these strategies are derived.

- Extensive performance results for various caching and mobility parameters is provided. Effect of mobility on delay for wireless content-caching networks is analyzed.

Rest of the chapter is organized as follows.The main problem and corresponding analytical framework to derive handover probabilities is discussed in Section 4.3. Several handover management policies are analyzed and discussed is Section 4.4. Performance evaluation and discussion are given in Section 4.5. Finally in section 4.6, the chapter is concluded.

4.2 System Model

4.2.1 Problem Statement

Consider a cellular network with multiple macro BSs, SBSs, and users. Users are connected to a macro BS following nearest association rule. SBSs are connected to the macro BSs using backhaul links. Now SBSs may or may not have cache memory, and a content request can be directly served by the SBS if content is stored there. Content requests follow popularly model with power law distributions such as a Zipfian, Pareto or a Zeta distribution (we use Zipfian). Also assume that requests for content follow an independent reference model (IRM) [103], and the SBSs follow leave copy everywhere (LCE) [104] caching strategy and least recently used (LRU) cache eviction policy.

In this chapter the effects of mobility when a content request is served by the SBS is studied. We want to analyze handover schemes when the user is moving out of SBS's coverage area, and the effect of handover in expected download delay, and finally develop efficient handover management policies.

4.2.2 Network Model

Consider the customary cellular architecture where each cell consists of a single macro base station (macro BS), multiple SBSs, and users. Assume the macro BSs, SBSs, and users are distributed over a 2-dimensional space $\mathbb{R}^2$ following three independent homogeneous Poisson point processes (PPP's) with intensity λ_1, λ_2, and λ_3, respectively. We can also assume that $\lambda_2 > \lambda_1$, which reflects the typical deployment rates of the two types of serving nodes.
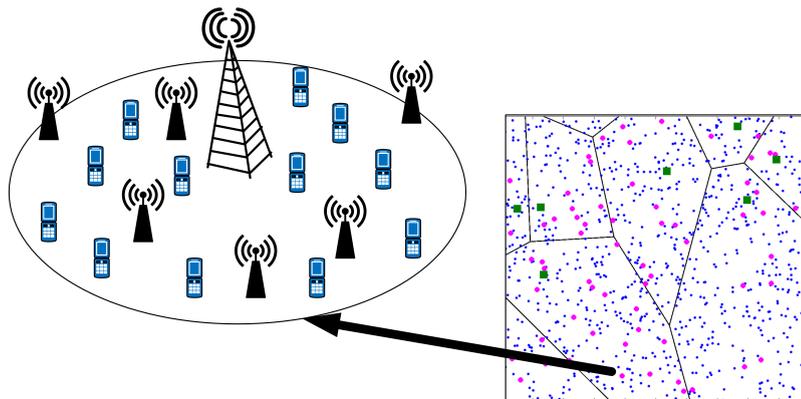
FIGURE 4.1: Network Architecture

Assume users are associated with the nearest macro BS, and therefore, the cellular architecture can be modeled using the Poisson-Voronoi tessellation, which is briefly described next. Consider the set of macro BSs to be $\Phi_1 = \{x_i\}$. Hence the Voronoi cell $C_{x_i}(\Phi_1)$ of point x_i is defined as [105],

$$C_{x_i}(\Phi_1) = \{y \in \mathbb{R}^2 : \| y - x_i \|_2 \leq \| y - x_j \|_2 , \forall x_j \in \Phi_1\} \tag{4.1}$$

Therefore, each macro BS x_i serves the mobile users within C_{x_i}, which follow the nearest association strategy. Similarly, the set of SBSs and users is given by Φ_2 and Φ_3 respectively.

4.2.3 Network Architecture

Basic network architecture is shown by Fig. 4.1. From the figure we observe that each cell consists of a macro BS (green nodes) and multiple SBSs (pink nodes) and users (blue nodes). Now each SBS can be identical in terms of cache size and transmission power or have different cache sizes and transmission power. To replicate this scenario we consider two different configurations for SBS network.

4.2.3.1 Two-Tier Model

Firstly, let us consider that there are two types of SBSs in a cell, one with finite cache storage and one without any cache. When user requests a content, it can be served by the macro BS or a SBS with cache storage, if the content is stored there. Thinning over the PPP Φ_2 is applied to determine the locations of SBSs with cache storage. For thinning parameter ζ, intensity of SBS with cache is $\lambda_4 = \zeta\lambda_2$. Understandably, greater value of ζ results in higher probability of finding a SBS source. SBS without cache acts as relay to enhance handover performance, which is discussed in Section 4.4.

4.2.3.2 K-Tier Model

A generalized analysis is developed by considering SBSs to be distributed over K-tiers. In this case, we consider that caching storage and transmission power varies for each tier. However, unlike two-tier architecture, all SBSs have a finite cache storage in this case. Locations of SBSs for i−th tier are obtained by employing independent thinning

48

parameter ζ_i. We can also state that, $\cup_{i=1}^{K} \zeta_i = 1$. In this article we consider that cache storage and transmission power is greater for SBSs in higher tier. As higher tier SBSs have greater transmission power and correspondingly greater coverage area, we consider their intensity to be less than lower tier SBSs. For K-tier model it is also important to select the serving node when multiple SBSs from different tiers are available. Strategies to select the serving node in presence of multiple sources is discussed in Section 4.3.

4.2.4 Mobility Model

Formally RWP can be described by an infinite sequence of tuple: $\{(X_{n-1}, X_n, V_n)\}_{n \in \mathbb{N}}$, where X_{n-1} denotes the starting position of a user at n-th movement period. User moves to X_n position with velocity V_n in n−th movement. We next describe the way of selecting waypoints in $\mathbb{R}^2$. Given the current waypoint X_{n-1}, the next waypoint X_n is chosen such that angle between the waypoints follow uniform distribution between $[0, 2\pi]$ and the transition length $\mathcal{L}_n = \parallel X_n - X_{n-1} \parallel$ follows Rayleigh distribution [56]. Therefore, probability distribution function (PDF) of transition length is given by

$$f_{\mathcal{L}}(l) = 2\pi\rho l e^{-\pi\rho l^2}, \quad l \geq 0 \tag{4.2}$$

Here ρ is the scaling parameter of Rayleigh distribution and its physical significance is, smaller scaling parameter suggests larger transition length, suitable for vehicles, and larger scaling parameter suggests smaller transition length, suitable for pedestrians. Let us also consider that selection of waypoints is independent for each movement period.

4.3 Handover Analysis

In this section several serving node selection strategies, expressions for different handover probabilities are discussed. Selection of serving node becomes relevant when multiple sources are available to serve the content request. Selection strategy has more impact on K-tier SBS network, because coverage area is different for SBSs of different tier, and judicious selection of the source can significantly reduce handover probability. Finally, expected delay is analyzed as a function of handover probabilities. The common symbols used throughout this chapter are tabulated in Table 4.1.

TABLE 4.1: Notations and Descriptions

Notation	Description
K	Content universe
α	Skewness parameter of Zipfian distribution
η	Path-loss exponent
P_{tm}	Transmit power of macro BS
P_{ts}	Transmit power of SBS
r_m	Distance between user and macro BS
r_s	Distance between user and SBS
C	Cache size
$\mathcal{H}$	Hit-rate
P_{th}	Threshold signal strength for detection
$\mathcal{L}$	Transition length of the mobile device
$\mathcal{T}$	Transition time for the mobile device
$\mathcal{V}$	Velocity of the mobile device
P_h	Probability of handover
P_{hss}	Probability of SBS-SBS handover
P_{hsbs}	Probability of handover between SBS and same cell macro BS
P_{hsbd}	Probability of handover between SBS and different cell macro BS

First of all we need to determine the probability of accessing a content from a cache. The content universe is denoted by $\mathcal{F}$ with n number of unique content, i.e., $\mathcal{F} = \{f_1, f_2, \cdots, f_n\}$. Let us consider that the requested content is f_i and the cache of interest is l. To model probability of requesting a content we follow widely used Zipfian distribution with skewness parameter α [33]. Then the probability of requesting content f_i is,

$$p_{f_i} = \frac{i^{-\alpha}}{\sum_{i=1}^{n} i^{-\alpha}} \tag{4.3}$$

Let us assume the incoming request rate at SBS l is λ_l, therefore incoming request rate for content i is given by

$$\lambda_{li} = \lambda_l p_{f_i} \tag{4.4}$$

Now following hit-rate analysis in [106] we can calculate the probability of obtaining content f_i at SBS l,

$$\mathcal{H}_{li} = 1 - e^{-\lambda_{li}\tau_{li}}, \tag{4.5}$$

where τ_{li} is the characteristic time of content f_i at SBS l. Characteristic time for a content in a SBS indicates the amount of time in future a recently accessed content is likely to remain in that SBS. Now τ_{li} can be obtained by solving,

$$C_l = \sum_{i=1}^{n} 1 - e^{-\lambda_{li}\tau_{li}}, \tag{4.6}$$

where C_l is the cache size of lth SBS. Now we discuss serving node selection strategies for both two-tier and K-tier architecture. Selection strategies are developed based on received signal strength (RSS), and we assumed that a SBS acts as a serving node only when its RSS is greater than the RSS of macro BS. Therefore, first we need to determine the RSS from the macro BS. For sake of simplicity in calculation we ignore fading and shadowing and assume path-loss as the primary signal degrading parameter. Assuming path-loss exponent to be η, received power from a macro BS is given by

$$P_{macroBS} = \kappa P_{tm} r_m^{-\eta} \tag{4.7}$$

where κ is the proportionality constant and it is normalized to be 1 throughout the chapter, and r_m denotes the distance of user from macro BS. Now we discuss source selection strategies for two-tier and K-tier model.

4.3.1 Two-Tier Model

For two-tier model user selects a SBS as the source if its RSS is greater than the RSS from macro BS. Considering same path-loss exponent, received power from SBS is given by

$$P_{SBS} = \kappa P_{ts} r_s^{-\eta}, \tag{4.8}$$

where r_s denotes the distance of user from SBS. Per selection strategy, user downloads the content from SBS if the RSS is greater than the macro BS. Therefore, to determine the probability of downloading content from a SBS, firstly, we determine the maximum allowable distance between the user and SBS,

$$r_{max} = \left(\frac{P_{ts}}{P_{tm}}\right)^{1/\eta} r_m \tag{4.9}$$

Now probability of finding m numbers of SBS with cache storage within r_{max} distance is,

$$\xi_m = e^{-\pi r_{max}^2 \zeta \lambda_2} \frac{(\pi r_{max}^2 \zeta \lambda_2)^m}{m!} \tag{4.10}$$

Now assuming homogeneous traffic distribution among SBSs, the probability of downloading content i from a SBS source is given by

$$P_{succ} = \sum_{m=1}^{\infty} \sum_{n=1}^{m} \xi_m \binom{m}{n} \mathcal{H}_i^n (1 - \mathcal{H}_i)^{m-n} \tag{4.11}$$

However, the assumption of homogeneous traffic among all the SBSs is invalid as cache size and coverage area is different for each tier. Therefore, we take a different approach to determine the probability of downloading a content from a SBS.

4.3.2 K-Tier Model

In this model all the SBSs are distributed over K-tiers, we can assume that only the nearest SBS of each tier might have greater RSS than the macro BS. Now, distance between the user and nearest j−th tier SBS is a random variable r_{s_j}, which follows Rayleigh distribution with mean $(2\sqrt{\zeta_j \lambda_2})^{-1}$. Similarly, distance between nearest macro BS and user follows Rayleigh distribution with mean $(2\sqrt{\lambda_1})^{-1}$ [105]. Similar to 2-tier model, user is associated with a SBS only if its RSS is greater than the RSS from macro BS. So first we define a set of candidate SBSs (Ω) which can be associated with the user. Probability that j−th tier SBS is an element of Ω is derived as follows,

$$\begin{aligned} \Upsilon_j &= P\left(\mathcal{H}_{ij} P_{ts_j} r_{s_j}^{-\eta} \geq P_{tm} r_m^{-\eta}\right) \\ &= P\left(r_m \geq \left(\frac{1}{\mathcal{H}_{ij}}\right)^{1/\eta} \left(\frac{P_{tm}}{P_{ts_j}}\right)^{1/\eta} r_{s_j}\right) \\ &= \frac{\zeta_j \lambda_2}{\zeta_j \lambda_2 + \lambda_1 \Psi^2}, \end{aligned} \tag{4.12}$$

where $\Psi = \left(\frac{P_{tm}}{\mathcal{H}_{ij} P_{ts_j}}\right)^{1/\eta}$. Associating with one of the elements in Ω depends on the association strategy. In this chapter we consider two association strategies, maximum RSS association and highest tier association.

4.3.2.1 Maximum RSS association

In this strategy, among the SBSs in Ω, user is associated with the SBS with maximum RSS. Assuming RSS from $j-$th tier SBS is a random variable ψ_j, we can express the association strategy using stochastic ordering. User is associated with $j-$th tier SBS when,

$$\psi_j > \psi_{i,i \in \Omega \backslash j} \tag{4.13}$$

From the properties of stochastic ordering we can write $\mathbb{E}[\psi_j] > \mathbb{E}[\psi_i]_{i \in \Omega \backslash j}$.

Therefore, probability of downloading the content from one of the SBS can be expressed as a sum of weighted probabilities of $\mathbb{E}[\psi]$,

$$
\begin{aligned}
P_{succ} &= \bigcup_{j=1}^{K} P_{succ_j} \\
&= \sum_{j=1}^{K} \frac{\mathbb{E}[\psi_j]}{\sum_{i=1}^{K} \mathbb{E}[\psi_i]} \Upsilon_j
\end{aligned}
\tag{4.14}
$$

4.3.2.2 Highest tier association

In this strategy, among the available SBSs in Ω, user is associated with highest tier SBS. This intuitive approach leverages the greater coverage area of higher tier SBSs to reduce the probability of handover. Combining probability Υ_j for each tier, downloading the content from one of the SBS is given by

$$P_{succ} = \bigcup_{j=1}^{K} P_{succ_j} = \Upsilon_k + \sum_{j=1}^{K-1} \prod_{i=j+1}^{K} \Upsilon_j (1 - \Upsilon_i) \tag{4.15}$$

Using these probabilities of downloading a content from a SBS for each considered scenario, given by (4.11), (4.14), and (4.15), we derive the probability of handover from a SBS. Traditionally, probability of handover is defined as the probability of moving out of the coverage area of associated BS. However, for delay analysis we need to redefine the probability of handover to incorporate incomplete content download.

Definition 2. *Probability of handover is defined as the probability of moving out of the coverage area of BS prior to completion of content download.*

Using this definition the probability of handover while downloading the content from a SBS is derived. Let us assume that user moves out of the coverage area when RSS from SBS drops below a threshold power. To study the effect of mobility over cache enabled network, only the handover between SBS and user, and seamless mobility among macro BSs is considered. Using this approach we can observe the impact of mobility over small-cell caching.

First the generic expression of handover probability is derived for a K-tier model and thereafter, handover probability for two-tier model is derived as a special case. Probability of handover in a K-tier network is given by

$$
\begin{aligned}
P_h &= \bigcup_{j=1}^{K} P_{succ_j} \, P_{hs_j} \\
&= \bigcup_{j=1}^{K} P_{succ_j} \, P\left(\mathcal{L} > \left(\frac{P_{ts_j}}{P_{th}} \right)^{1/\eta} \middle| \mathcal{T} < T_c \right),
\end{aligned}
\tag{4.16}
$$

where P_{hs_j} represents the handover probability between $j-$th tier SBS and user. Two random variables, $\mathcal{L}$ and $\mathcal{T}$ represent transition length and transition time, and are dependent on the mobility model. We also know that $\mathcal{T} = \mathcal{L}/\mathcal{V}$, where $\mathcal{V}$ can be a positive constant or a positive random variable. Therefore, we can further simplify (4.16) as

$$
P_h = \bigcup_{j=1}^{K} P_{succ_j} \frac{P\left(\left(\frac{P_{ts_j}}{P_{th}} \right)^{1/\eta} < \mathcal{L} < \mathcal{V}T_c \right)}{P(\mathcal{L} < \mathcal{V}T_c)}
\tag{4.17}
$$

Let us consider two cases, a) user is moving at a constant velocity, i.e., $\mathcal{V} \equiv v$, and b) velocity of user is uniformly distributed on $[v_{\min}, v_{\max}]$, i.e., $\mathcal{V} = \mathrm{Uni}(v_{\min}, v_{\max})$. $\mathrm{Uni}(a, b)$ denotes uniform distribution on $[a, b]$.

$$P_h = \begin{cases} \sum_{j=1}^{K} \dfrac{\mathbb{E}[\psi_j]}{\sum_{i=1}^{K} \mathbb{E}[\psi_i]} \Upsilon_j \left(1 - \dfrac{1 - \exp\left(-\rho\pi\left(\dfrac{P_{ts_j}}{P_{th}}\right)^{2/\eta}\right)}{1 - \exp\left(-\rho\pi\left(vT_c\right)^2\right)} \right), & \text{for } \textit{Maximum RSS Association} \\[6mm] \Upsilon_k \left(1 - \dfrac{1 - \exp\left(-\rho\pi\left(\dfrac{P_{ts_k}}{P_{th}}\right)^{2/\eta}\right)}{1 - \exp\left(-\rho\pi\left(vT_c\right)^2\right)} \right) + \sum_{j=1}^{K-1} \prod_{i=j+1}^{K} \Upsilon_j (1 - \Upsilon_i) \\[6mm] \times \left(1 - \dfrac{1 - \exp\left(-\rho\pi\left(\dfrac{P_{ts_k}}{P_{th}}\right)^{2/\eta}\right)}{1 - \exp\left(-\rho\pi\left(vT_c\right)^2\right)} \right), & \text{for } \textit{Highest Tier Association} \end{cases}$$

$$(4.18)$$

Theorem 1. *If $\mathcal{V} \equiv v$, then handover probability is given by (4.18),*

Proof. See Appendix A.1 $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 2. *If $\mathcal{V}$ is a uniform random variable distributed between $[v_{\min}, v_{\max}]$, i.e., $\mathcal{V} = Uni(v_{min}, v_{max})$, then handover probability is given by (4.19),*

$$P_h = \begin{cases} \sum_{j=1}^{K} \dfrac{\mathbb{E}[\psi_j]}{\sum_{i=1}^{K} \mathbb{E}[\psi_i]} \Upsilon_j \left(1 - \dfrac{T_c(v_{max} - v_{min})\left(1 - \exp\left(-\rho\pi\left(\dfrac{P_{ts}}{P_{th}}\right)^{2/\eta}\right)\right)}{T_c(v_{max} - v_{min}) + Q(\sqrt{2\rho\pi}v_{max}) - Q(\sqrt{2\rho\pi}v_{min})} \right), \\[2mm] \hspace{6cm} \text{for } \textit{Maximum RSS Association} \\[6mm] \Upsilon_k \left(1 - \dfrac{T_c(v_{max} - v_{min})\left(1 - \exp\left(-\rho\pi\left(\dfrac{P_{ts}}{P_{th}}\right)^{2/\eta}\right)\right)}{T_c(v_{max} - v_{min}) + Q(\sqrt{2\rho\pi}v_{max}) - Q(\sqrt{2\rho\pi}v_{min})} \right) + \sum_{j=1}^{K-1} \prod_{i=j+1}^{K} \Upsilon_j (1 - \Upsilon_i) \\[6mm] \times \left(1 - \dfrac{T_c(v_{max} - v_{min})\left(1 - \exp\left(-\rho\pi\left(\dfrac{P_{ts}}{P_{th}}\right)^{2/\eta}\right)\right)}{T_c(v_{max} - v_{min}) + Q(\sqrt{2\rho\pi}v_{max}) - Q(\sqrt{2\rho\pi}v_{min})} \right), \\[2mm] \hspace{6cm} \text{for } \textit{Highest Tier Association} \end{cases}$$

$$(4.19)$$

Proof. See Appendix A.2 □

Corollary 1. *Similar to Theorem 1 and 2, handover probability for a two-tier SBS model is given by (4.20),*

$$
P_h = \begin{cases}
\sum_{m=1}^{\infty} \sum_{n=1}^{m} \xi_m \binom{m}{n} \mathcal{H}_i^n (1 - \mathcal{H}_i)^{m-n} \left(1 - \dfrac{1 - \exp\left(-\rho\pi \left(\dfrac{P_{ts}}{P_{th}}\right)^{2/\eta}\right)}{1 - \exp\left(-\rho\pi \left(vT_c\right)^2\right)} \right), & \text{for } \mathcal{V} \equiv v \\[4ex]
\sum_{m=1}^{\infty} \sum_{n=1}^{m} \xi_m \binom{m}{n} \mathcal{H}_i^n (1 - \mathcal{H}_i)^{m-n} \left(1 - \dfrac{T_c(v_{max} - v_{min}) \left(1 - \exp\left(-\rho\pi \left(\dfrac{P_{ts}}{P_{th}}\right)^{2/\eta}\right)\right)}{T_c(v_{max} - v_{min}) + Q(\sqrt{2\rho\pi}v_{max}) - Q(\sqrt{2\rho\pi}v_{min})} \right), \\
& \text{for } \mathcal{V} \equiv \text{Uni}(v_{min}, v_{max})
\end{cases}
\tag{4.20}
$$

Proof. Similar to (4.17), handover probability for two-tier model can be expressed as,

$$
P_h = P_{succ} \, P_{hs}
$$
$$
= P_{succ} \, \frac{P\left(\left(\dfrac{P_{ts}}{P_{th}}\right)^{1/\eta} < \mathcal{L} < \mathcal{V}T_c \right)}{P(\mathcal{L} < \mathcal{V}T_c)}
\tag{4.21}
$$

Replacing P_{succ} with (4.11) and P_{hf} with corresponding expressions in Theorem 1 and 2, we obtain the handover probability for two-tier SBS model. □

Theorem 1, 2 are extremely important as they represent the overall probability of handover while downloading the content from a SBS for considered network architecture and mobility model. In the next section several handover management policies are discussed, where user connects with another SBS or macro BS to download the rest of the content.

4.4 Handover Management

In this section we discuss several handover management strategies and derive the probabilities of specific handovers, such as SBS-SBS handover, SBS-macro BS handover, and

finally, we express delay as a function of these handovers. Three handover management policies are considered, a) *associate with macro BS*, b) *associate with another source*, and c) *use relay to reconnect*. Detailed description of each policy and their impact on delay is provided below. Independent of handover management policies, expected delay can be expressed as,

$$\mathbb{E}[d] = \mathbb{E}[d_s] + \mathbb{E}[d_m] + \mathbb{E}[d_h], \tag{4.22}$$

where $\mathbb{E}[\cdot]$ is the expectation operator, and d_s, d_m, and d_h are download delay from SBS, macro BS, and download delay in case of handover, respectively. Understandably, calculation of delay in case of handover ($\mathbb{E}[d_h]$) requires the information about expected handover time. Therefore, we derive the the expression for expected handover time or the expected time of staying within a SBS's coverage area. Assuming the SBS cell to be circular we can write,

$$\overline{T}_h = \mathbb{E}[T] \int_0^{2\pi} \int_0^{r_{max}} f(r, \theta) r \, dr \, d\theta, \tag{4.23}$$

where $f(r, \theta)$ is the spatial node density and $\mathbb{E}[T]$ is the expected transition time that depends on the velocity of the mobile node. Lin *et al.* gave the expressions for $\mathbb{E}[T]$ in [56],

$$\begin{aligned}
\mathbb{E}[T] &= \frac{1}{2v\sqrt{\rho}}, \text{ for constant velocity } \mathcal{V} \equiv v \\
&= \frac{\ln v_{max} - \ln v_{min}}{2\sqrt{\rho}(v_{max} - v_{min})}, \text{ for r.v. } \mathcal{V} = \text{Uni}(v_{min}, v_{max})
\end{aligned} \tag{4.24}$$

where $\text{Uni}(a, b)$ denotes the uniform distribution on $[a, b]$.

Using the spatial distribution given in [56], the expected handover time is given as,

$$\begin{aligned}
\overline{T}_h &= \mathbb{E}[T] \int_0^{2\pi} \int_0^{r_{max}} \frac{\sqrt{\rho}}{\pi r} e^{-\rho \pi r^2} r \, dr \, d\theta \\
&= \mathbb{E}[T] \left(1 - 2Q\left(\sqrt{2\rho\pi} r_{max}\right)\right)
\end{aligned} \tag{4.25}$$

Now, depending on the handover management policy several expressions for $\mathbb{E}[d_h]$ are derived and correspondingly obtain the expression for delay.

4.4.1 Associate with macro BS

This is the simplest handover management policy, where user connects with a macro BS to complete the content download. Primary motif of this policy is to avoid handover from another SBS. Fig. 4.2 illustrates this handover management policy, where user connects with the nearest macro BS when it moves out of the coverage area of a SBS. In the figure, user is initially connected to SBS1 and downloading the requested content. When the RSS from SBS1 drops below the threshold value, handover takes place, and according to the handover management policy user connects with the nearest macro BS.
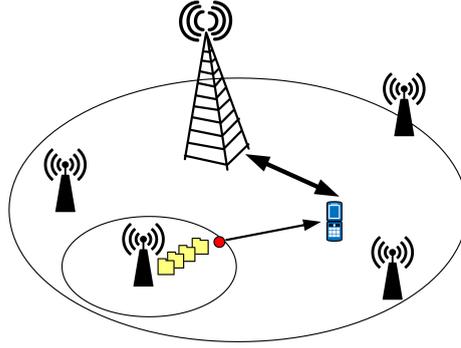
FIGURE 4.2: Associate with macro BS handover management

In this case delay can be expressed as

$$
\mathbb{E}[d] = \frac{M}{S_s} P_{succ}(1 - P_h) + \left(\frac{M}{S_m} + d_q\right)(1 - P_{succ})
$$
$$
+ \left(\overline{T}_h + \frac{M - S_s \overline{T}_h}{S_m} + d_q + d_h\right) P_{succ} P_h \tag{4.26}
$$

First term in (4.26) represents download delay from SBS. Similarly, second term represents download delay from macro BS, and finally, the third term represents download delay in case of handover and thereafter, downloading the rest of the content from a macro BS.

4.4.2 Associate with another source

In this handover management policy user connects with a new source to download the rest of the content. New source can be another SBS or a macro BS, and source is selected depending on the used source selection strategy. Therefore, in this policy following handovers are possible, *a) SBS-SBS handover, b) SBS-same cell macro BS handover,*

c) SBS-different cell macro BS handover. We depict this scenario in Fig. 4.3, where user can connect with macro BS1, macro BS2, or SBS2. Among these possible sources user connects to a new source based on the native source selection strategy.
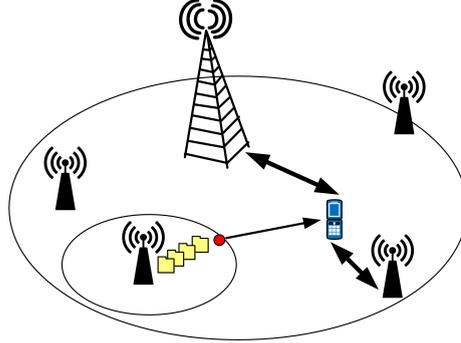
FIGURE 4.3: Associate with another source handover management

Now we derive the probability for each of these handover. From the overall handover probability we can derive the probability of SBS-SBS handover

$$P_{hss} = P_h P_{succ} \tag{4.27}$$

To determine the probability of a SBS to macro BS handover, we need to determine the probability of a node remaining in the same cell during the data transfer. However, it cannot be derived similar to (4.16), rather we need the probability of a node moving within a certain cell, or a certain measurable set $\mathcal{A}$. To derive this probability, conditioning over the random location of an user complicates the mathematical analysis. However, it can be assumed that the user is placed at the origin of the cell, and using Slivnyak's theorem, Which suggests that conditioning with respect to a certain point does not affect system behavior at other points [105]. It follows from the independence property of Poisson point processes. Even with the user placement on the origin, randomness of Poisson-Voronoi tessellation cell area makes it extremely difficult to derive a closed form solution.

Let us consider that linear contact distribution is given by

$$H_l(r) = 1 - \exp(-\lambda^{(2)}r), \tag{4.28}$$

where $H_l(r)$ gives probability of making contact with the cell boundary after traversing r length, and $\lambda^{(2)}$ is the intensity of cells in $\mathbb{R}^2$. We use linear contact distribution to determine if the node remains in the same cell. Leveraging contact distribution we develop an upper-bound for the probability of SBS to same cell macro BS handover.

Let us assume cells to be Poisson polygons to select the $\lambda^{(2)}$, however, it is impossible to determine the exact r at the time of handover, as it depends on the direction switch rate. Therefore, we provide a bound for the probabilities corresponding to SBS-macro BS handover, considering both same and different cell scenarios.

For Poisson polygons, $\lambda^{(2)} = \frac{4\lambda_1}{\pi\sigma}$ [105]. A bound for SBS-macro BS handover probability can be obtained by considering $r = vT_c$, i.e, considering transition in a single direction. Therefore, the bounds for SBS-macro BS handover for same and different cell are given by

$$P_{hsbs} \geq P_h(1 - P_{succ})(1 - H_l(vT_c)) \tag{4.29}$$

$$P_{hsbd} \leq P_h(1 - P_{succ})H_l(vT_c) \tag{4.30}$$

As assumed earlier, there are two possible scenarios of v and each results in a different bound. When v is a random variable, the contact distribution is given by

$$
\begin{aligned}
H_l(vT_c) &= \frac{1}{T_c(v_{max} - v_{min})} \int_{v_{min}}^{v_{max}} 1 - \exp\left(-\frac{4\lambda_1}{\pi\sigma}vT_c\right)dv \\
&= \frac{1}{T_c} + \frac{\pi\sigma}{4T_c^2\lambda_1(v_{max} - v_{min})}\left(\exp\left(-\frac{4\lambda_1}{\pi\sigma}v_{max}T_c\right) \right. \\
&\left. - \exp\left(-\frac{4\lambda_1}{\pi\sigma}v_{min}T_c\right)\right)
\end{aligned}
\tag{4.31}
$$

Similarly for a constant velocity v, the linear contact distribution is given by

$$H_l(vT_c) = 1 - \exp\left(-\frac{4\lambda_1}{\pi\sigma}vT_c\right). \tag{4.32}$$

Depending on the velocity of the mobile user we replace (4.31) or (4.32) in (4.29) and (4.30) to obtain the bounds for SBS-macro BS handover.

Finally, using (4.8) and (4.25) expected delay can be expressed as,

$$
\begin{aligned}
\mathbb{E}[d] = {} & \frac{M}{S_s} P_{succ}(1 - P_h) + \left(\frac{M}{S_m} + d_q\right)(1 - P_{succ}) \\
& + \left(\overline{T}_h + \frac{M - S_s \overline{T}_h}{S_m} + d_q + d_h\right) P_{succ}(1 - P_{hss}) \\
& + \left(\frac{M}{S_s} + d_h\right) P_{succ} P_{hss}
\end{aligned}
\tag{4.33}
$$

Third and fourth term of (4.33) represent the delay in case of SBS-macro BS handover and SBS-SBS handover, respectively.

4.4.3 Use relay to reconnect

Finally we propose an intuitive handover management policy, where source is selected depending on the amount of content downloaded prior to handover. If user has aready downloaded a significant amount of content, it might be judicious to use relay to enhance the RSS and download the rest of the content from the same source. In formal words, if the downloaded portion of the content is greater than a threshold value, then using another SBS as relay, signal strength is amplified (e.g., amplify-and-forward relaying strategy) and user continues to download the content from the initial source. However, if amount of download is less than the threshold, then rest of the content is downloaded from another source, following second handover management policy.

Therefore, in this case important parameter is probability of using relay. It can be derived as follows,

$$
\begin{aligned}
P_r &= P\left(\mathcal{T} > \gamma T_c\right) \\
&= P\left(\mathcal{L} > \gamma \mathcal{V} T_c\right) \\
&= 1 - P\left(\mathcal{L} \leq \gamma \mathcal{V} T_c\right),
\end{aligned}
\tag{4.34}
$$

here P_r is the probability of using relay. Depending on the velocity distribution final expressions of P_r is given by

$$
P_r = \begin{cases}
\exp\left(-\rho\pi(\gamma v T_c)^2\right), \text{ for } \mathcal{V} \equiv v \\
\\
\dfrac{Q(\sqrt{2\rho\pi}\gamma T_c v_{min}) - Q(\sqrt{2\rho\pi}\gamma T_c v_{max})}{T_c(v_{max} - v_{min})}, \\
\qquad\qquad\qquad \text{for } \mathcal{V} \equiv \text{Uni}(v_{min}, v_{max})
\end{cases}
\tag{4.35}
$$
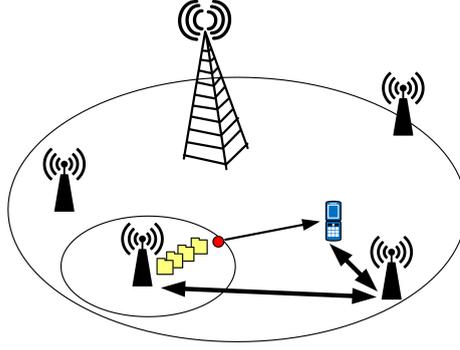
FIGURE 4.4: Relay to reconnect handover management

Assuming the processing time to use another SBS as relay has negligible impact on delay, we can express expected delay similar to (4.33),

$$
\begin{aligned}
\mathbb{E}[d] = {} & \frac{M}{S_s}P_{succ}(1 - P_h + P_h P_r) + \left(\frac{M}{S_m} + d_q\right)(1 - P_{succ}) \\
& + \left(\overline{T}_h + \frac{M - S_s\overline{T}_h}{S_m} + d_q + d_h\right)P_{succ}(1 - P_r) \\
& \times (1 - P_{hss}) + \left(\frac{M}{S_s} + d_h\right)P_{succ}(1 - P_r)P_{hss}
\end{aligned}
\tag{4.36}
$$

It can be realized that setting the threshold γ to a low value increases the probability of further handover, whereas setting γ to a higher value reduces the usefulness of using a relay. However, studying optimality of γ is out of the scope of this thesis and we would like to explore it in our future work. For the numerical results a fixed value of γ is considered to study if this handover management policy is really useful.

4.5 Numerical Results

In this section, we study the performance of the considered small-cell cache network for several network parameters. We primarily observe whether a k-tier network approach is better for handover and delay performance, and which is the best handover management policy among the discussed ones. Now we would like to mention the considered assumptions in generating the numerical results. Firstly, we assume macro BSs, SBSs, and mobile users are scattered based on independent homogeneous PPPs with intensities of $\{\lambda_1, \lambda_2, \lambda_3\} = \{\frac{1}{\pi 500^2}, \frac{20}{\pi 500^2}, \frac{300}{\pi 500^2}\}$. Secondly, we assume fixed size of content universe, transmission power of a macro BS, queuing delay, and handover delay. Thirdly, to keep uniformity in comparing the 2-tier network and the K-tier network, we consider same network cache size. Finally, we assume that higher tier SBSs have greater cache size, and higher transmission power.

We consider the thinning parameter in the 2-tier network to be $\zeta = 0.85$, i.e., 85% of the SBSs are equipped with cache storage and cache size at each SBS is varied between $50 - 250$. Transmission powers are considered to be $\{P_1, P_2\} = \{25, 15\}$ dBm. For K-tier results, we consider a 4-tier SBS network, and locations of SBSs are determined using thinning, $\zeta = \{0.3, 0.25, 0.25, 0.2\}$. As mentioned earlier, cache size and transmission power is different for SBS of each tier. To maintain uniformity, we consider that cache size of SBS in 4th tier is varied between $50 - 250$ and cache size for tier $1, 2,$ and 3 are $0.75, 0.8,$ and 0.9 times of the 4th tier, respectively. Similarly, transmission powers of tier $1 - 4$ are $\{12, 14, 16, 18\}$ dBm, respectively. To generate the handover and delay results we consider fixed path loss exponent value of $\eta = 4$. Rest of the parameters are mentioned in Table 4.2.

TABLE 4.2: Numerical Results Parameter

Parameter	Description	Value
P_{th}	Threshold signal strength for detection	10 dBm
K	Content universe	1000 content
M	Content size	50 Mb
α	Content popularity skewness	$0.5 - 0.9$
γ	Threshold for relay to reconnect	0.5
σ	Poisson polygon intensity	0.1×10^{-3}
S_s	SBS data rate	10 Mbps
S_m	macro BS data rate	5 Mbps
d_q	Queuing delay	2 sec
d_h	Handover delay	0.5 sec

(a) $\alpha = 0.5$

(b) $\alpha = 0.7$

(c) $\alpha = 0.9$
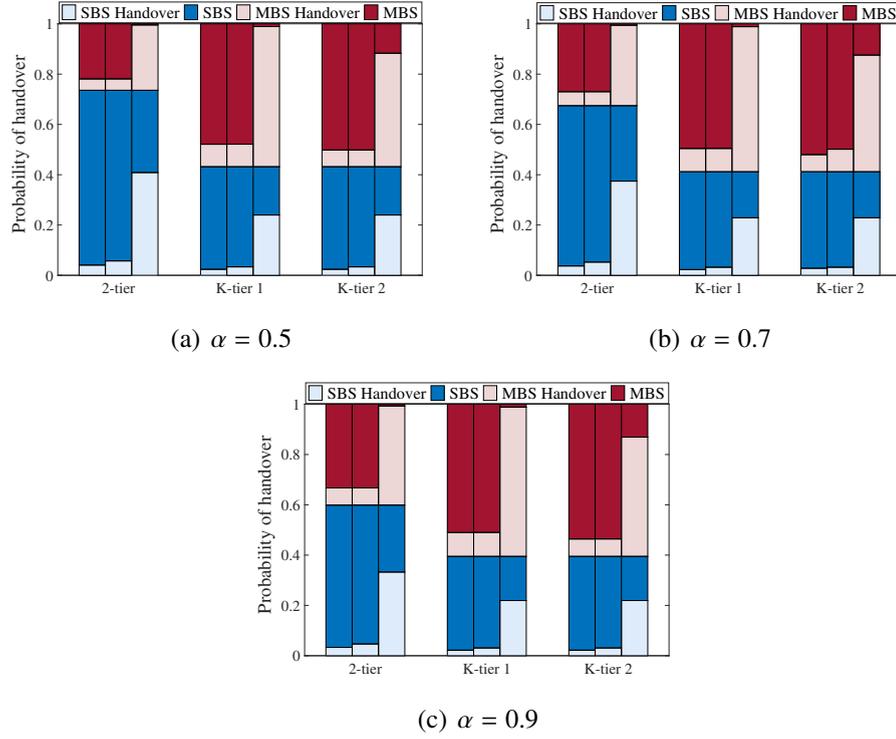
FIGURE 4.5: Probability of Handover for cache size = 150

First we observe the handover probabilities for several content popularity skewness values, $\alpha = 0.5, 0.7$, and 0.9, and cache size of 150 in Fig. 4.5. Handover results are illustrated in a grouped stacked bar format. Each figure has three group of bars, 2-tier model, K-tier model with maximum RSS association strategy, and K-tier model with highest RSS selection. Each group of bar consists of results for different velocities. We consider three different velocities, fixed velocity of 1 m/s, uniform velocity between [3, 5] m/s, and fixed velocity of 10 m/s. Finally, each stacked bar depicts four results, probability of downloading content from a SBS, probability of handover while the downloading the content from a SBS, probability of downloading the content from a macro BS, and probability of handover while downloading the content from a macro BS. From the figure we observe that probability of downloading a content from a SBS is higher for K-tier models. It happens because the transmission power of higher tier SBS (tier 3 and 4) is greater than the transmission power of a SBS in 2-tier model. Similarly, handover probability increases for K-tier model as the transmission power of lower tier SBS (tier 1 and 2) is low, and accordingly their coverage area is low as well. Now, for K-tier model with maximum RSS association strategy (K-tier 1 in the figure), more often user starts downloading the content from a low tier SBS and therefore, SBS handover probability is high in this case. However, for K-tier model with highest tier association strategy (K-tier 2 in the figure), probability of handover is reduced. We also

(a) Associate with macro BS   (b) Associate with another source

(c) Relay to reconnect
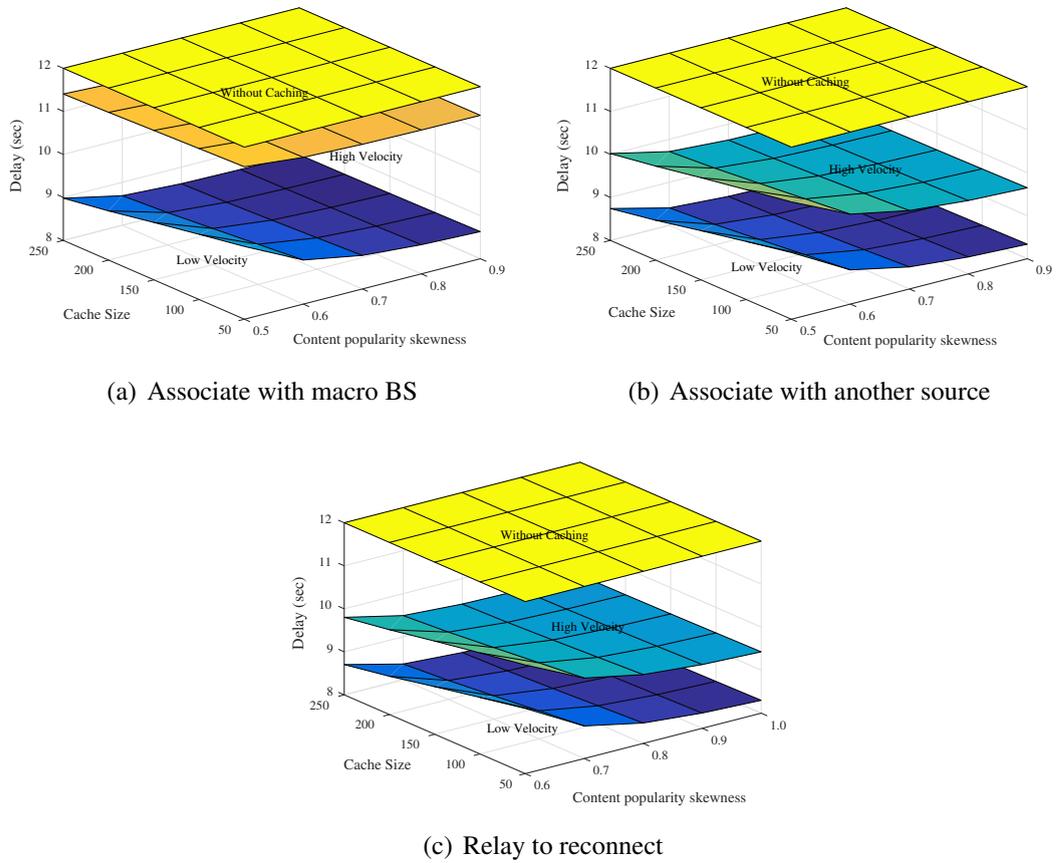
FIGURE 4.6: Expected download delay for 2-tier network

observe that probability of downloading a content from a SBS increases with content popularity skewness.

Now, we study the delay performance for varying content popularity skewness between 0.5 – 0.9, and varying cache size between 50 – 250. We provide results for each handover management policy and two different velocities, low velocity (1 m/s) and high velocity (10 m/s). We consider transition length to be 0.1 and 0.001 for low velocity and high velocity, respectively. From the results we observe that delay performances for low velocity are almost identical for each handover management policy, however, handover management policy plays an important role for high velocity cases. Results depict that relay to reconnect handover management policy performs best and connect to macro BS policy performs worst. As relay to reconnect policy alleviates the handover delay, we observe better delay performance. Whereas, connect to macro BS performs worst due to the additional queuing delay. We would like to mention that, even though connect to macro BS performs worst and provides negligible advantage over the existing communication, it is extremely simple and easy to implement. Therefore, for a real

(a) Associate with macro BS

(b) Associate with another source

(c) Relay to reconnect
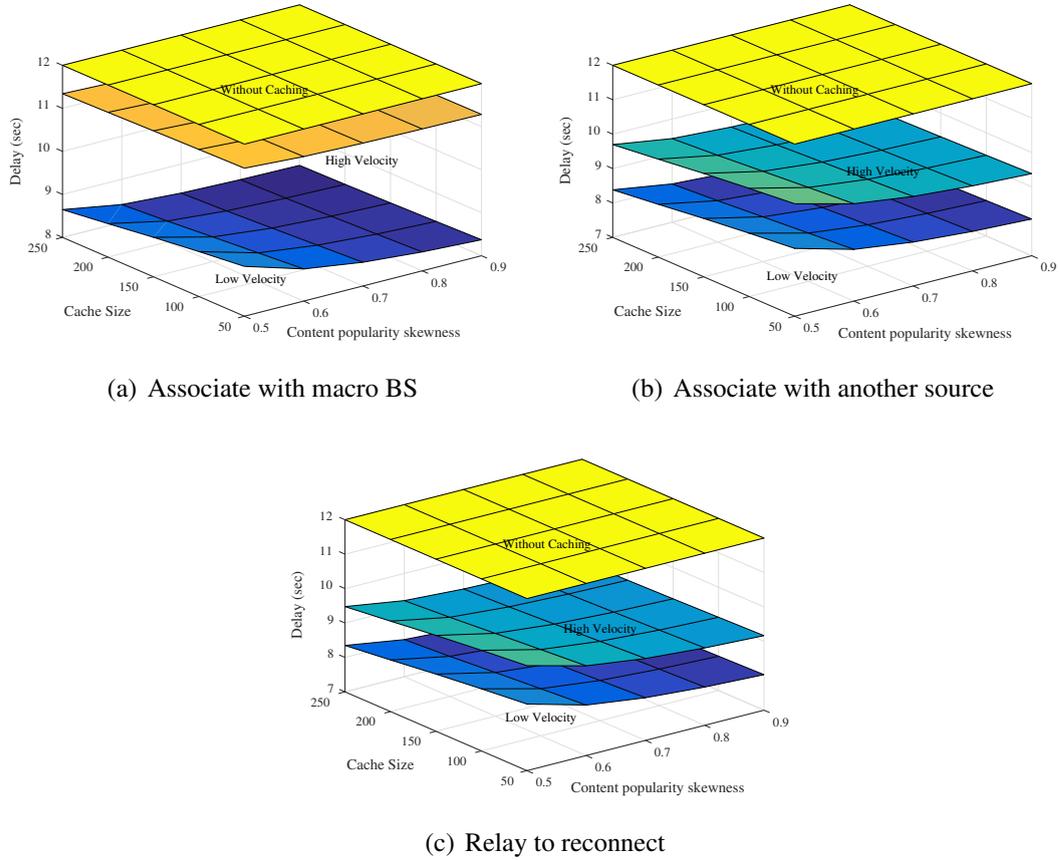
FIGURE 4.7: Expected download delay for K-tier network with Maximum RSS selection

life scenario with low velocity users, connect to macro BS would give satisfactory performance. Overall, from the delay results it is evident that the K-tier model with highest tier selection strategy performs best, and the primary reason behind this result is higher probability of downloading the content from a SBS and lower probability of handover while downloading the content from a SBS. Using the K-tier model with highest tier selection strategy and relay to reconnect handover management policy, small-cell caching gives approximately 42% and 36% delay performance improvement over a network without content caching.

We also study the effect of varying path loss exponent and intensity of cache enabled SBS over the network. As it is easier to study the latter one in a 2-tier model, we consider a 2-tier network with relay to reconnect handover management policy. In Figs. 4.9 and 4.10 we study the effect of varying the thinning parameter ζ between $0.1 - 0.9$ and consider two different path loss exponent values ($\eta = 2, 4$) for both low and high velocity users. From the figures we observe an important phenomena, increasing the intensity of cache enabled SBS can sometime degrade the performance, for example
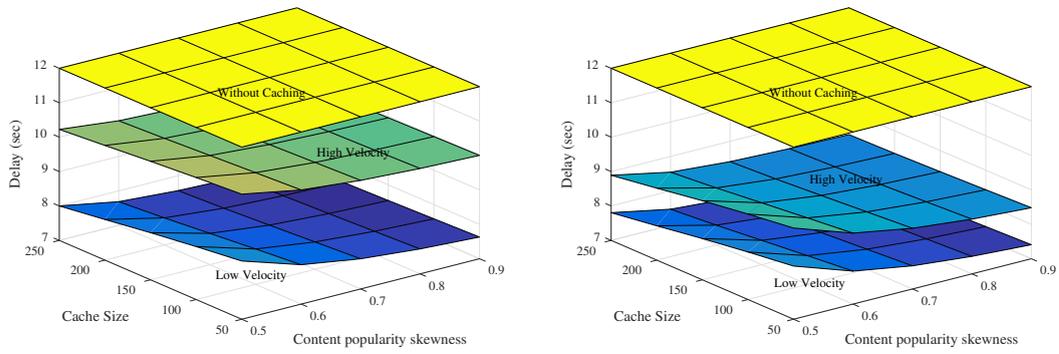
(a) Associate with macro BS   (b) Associate with another source
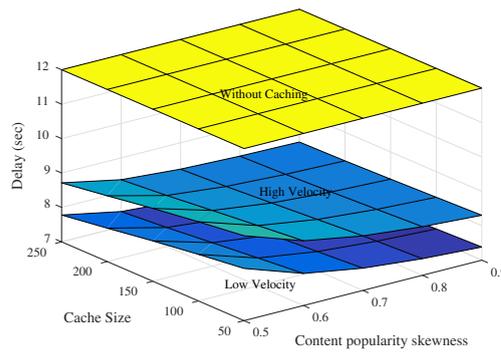
(c) Relay to reconnect

FIGURE 4.8: Expected download delay for K-tier network with Highest tier selection
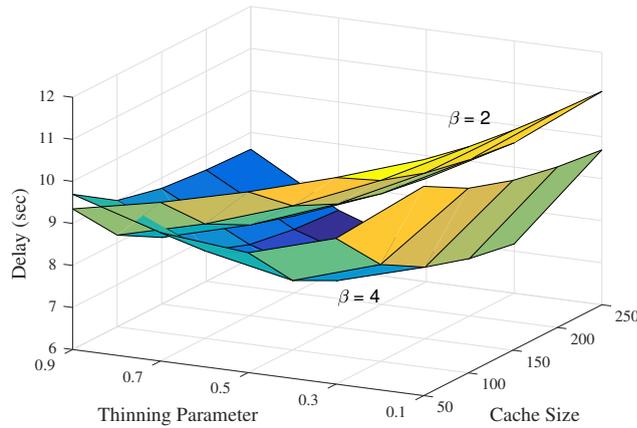
FIGURE 4.9: Effect of varying thinning parameter for low velocity

Figure 4.9 illustrates that best performance is achieved for cache enabled SBS intensity of 0.5 at $\eta = 4$. Higher intensity of cache-enabled SBS offloads the traffic and increases the probability of serving a content from a SBS, but it also increases the probability of handover and therefore, degrade the performance. We can intuitively conclude that there
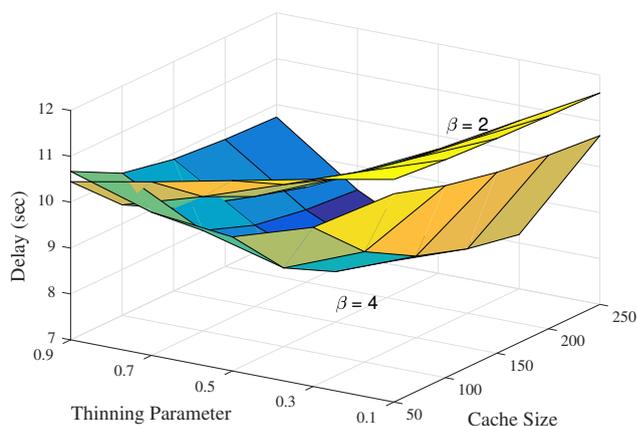
FIGURE 4.10: Effect of varying thinning parameter for high velocity

exists an optimal intensity of cache enabled SBS to achieve best delay performance.

4.6 Conclusion

Wireless content-caching, where popular multimedia content is stored at local access points, is a promising emerging concept. Previously in literature, performance of caching in a wireless is investigated without considering user mobility. Therefore, in this chapter we have analyzed the effect of mobility. By using stochastic geometry and the random waypoint model, we derived the handover probabilities and the expected delay. We also explored several handover management policies. The numerical results explored the effect of several caching related parameters, user velocity, and cache intensity on expected delay and handover probabilities.

Chapter 5

Asymptotic Analysis of Generalized Fading Channels

5.1 Introduction

Popular generalized fading models obey the property of logarithmic singularity at zero, and they are classified as LS wireless channels. Due to the logarithmic singularity at zero, Taylor series and other Taylor series based asymptotic measures produce significant amount of error in high SNR analysis. While the wireless performance under the GG distribution has been extensively analyzed, the derived expressions of the performance metrics tend to be complicated. Moreover, the existing analysis ignores the presence of the logarithmic singularity. Taking all these into consideration, we develop a new asymptotic approach for LS wireless channels. The proposed approach is extremely useful for wireless communication researches who are dealing with interference analysis of a cache-enabled network, small-cell network. However, the proposed approach is not limited to this scenario, and applicable for all the scenarios using GG and related channel models.

Contributions of this chapter are summarized below.

- A new asymptotic performance measure for wireless channels is proposed. It includes the logarithmic term, which leads to a generalization of the classical diversity gain and coding gain relationship (e.g., (2)).

- New asymptotic expressions for BER and outage probability are derived. Closed-form solutions of BER is derived for different modulation schemes, Binary Phase Shift Keying (BPSK), Differential Binary Phase Shift Keying (DBPSK), and M-ary Phase Shift Keying (PSK).

- Different antenna diversity models, selection combining and maximal ratio combining are considered, and the resulting PDF to obtain numerical results are derived.

- Closeness of the proposed asymptotic model is compared with the existing ones. Kolmogorov-Smirnov test is used to measure the closeness of the proposed asymptotic approach too.

The rest of the chapter is organized as follows. The details of the proposed asymptotic approach is provided in Section 5.2. Closed-form solutions for BER and outage probability for different scenarios are given in Section 5.3. Numerical results for different statistical measures with various diversity schemes and performance comparison with existing asymptotic approaches are given in Section 5.4. A brief overview and conclude of the chapter is given in Section 5.5.

5.2   Main Results

In this section we discuss the details of the new asymptotic performance measure. For asymptotic performance measure of LS channels the PDF is considered to be,

$$f(\beta) = a\beta^t + b\beta^\mu \log(\beta) + R(\beta), \quad \text{for } \beta \to 0^+ \tag{5.1}$$

where a, b, t and μ are constants to be determined from the distribution of the fading channel, log is natural logarithm throughout the chapter, and $R(\beta)$ is the vanishing remainder term. The singularity is determined by the term $b\beta^\mu \log(\beta)$.

Theorem 3. *Now averaging over the two terms in (2), we get a generalized version of asymptotic performance measure,*

$$P_e(\overline{\gamma}) = (G_c\overline{\gamma})^{-G_d} \exp\left(\frac{c' \log(\overline{\gamma})}{\overline{\gamma}^\alpha}\right) + R(\overline{\gamma}), \quad as \quad \overline{\gamma} \to \infty \tag{5.2}$$

where G_c, G_d, and $R(\overline{\gamma})$ are coding and diversity gain, and remainder term, respectively.

Proof. Using (5.1) an expression for error probability similar to one obtained using Taylor series can be derived,

$$P_e(\overline{\gamma}) = \frac{c_1}{\overline{\gamma}^{t+1}} + \frac{c_2}{\overline{\gamma}^{\mu+1}} + \frac{c_3}{\overline{\gamma}^{\mu+1}} \log(\overline{\gamma}), \tag{5.3}$$

where c_1, c_2, and c_3 are constants depending on the modulation scheme and t, μ are the constants from (5.1). As our primary interest is asymptotic solution, only the dominant terms are considered. Therefore, for $\overline{\gamma} \to \infty$ the second term in (5.3) can be ignored and reduced to,

$$P_e(\overline{\gamma}) = \frac{c_1}{\overline{\gamma}^{t+1}} + \frac{c_3}{\overline{\gamma}^{\mu+1}} \log(\overline{\gamma}) \tag{5.4}$$

Applying elementary series approximations over (5.4) the generalized form of error probability given by (5.2)

$$\begin{aligned} P_e(\overline{\gamma}) &= \frac{c_1}{\overline{\gamma}^{t+1}} \left(1 + \frac{c_3/c_1}{\overline{\gamma}^{\mu-t}} \log(\overline{\gamma})\right) \\ &= \frac{c_1}{\overline{\gamma}^{t+1}} \exp\left(\frac{c' \log(\overline{\gamma})}{\overline{\gamma}^{\mu-t}}\right), \end{aligned} \tag{5.5}$$

where $c' = c_3/c_1$ is another constant. Understandably, (5.5) can be written in the form of (5.2) with $G_c = c_1^{-1/t+1}$, $G_d = t + 1$, and $\alpha = \mu - t$. □

Proposition 1. *However, the asymptote in (5.1) works only in the high SNR regime. To overcome this shortfall and extend the SNR range we add one more higher order term with (5.1) and develop a new asymptote by approximating first two terms of the series expansion as an exponential function, similar to [88],*

$$\begin{aligned} f(\beta) &= a\beta^t + a_1\beta^{t+1} + b\beta^\mu \log(\beta) + b_1\beta^{\mu+1} \log(\beta) + R(\beta) \\ &= a\beta^t \exp\left(-\frac{a_1}{a}\beta\right) + b\beta^\mu \log(\beta) \exp\left(-\frac{b_1}{b}\beta\right) + R(\beta) \end{aligned} \tag{5.6}$$

Similar to (5.1), $R(\cdot)$ is the remainder term.

The asymptote in (5.6) significantly improves the performance in the SNR range of 5-15 dB. Considering few more higher order terms and developing an asymptote similar to (5.6) can be a simple solution to improve the performance.

Comparing (2.11) and (5.2), we observe an additional exponential term, namely correction term. This correction term can be found in (5.22) as the second and third term

on the right hand side. We can consider this new expression, given by (5.2), as a generalized solution that works for fading models both with and without the logarithmic singularity. In absence of logarithmic singularity, the $\log(\bar{\gamma})$ term vanishes and the expression reduces to (2.11). Over the rest of the chapter we emphasize the effect of this correction term in asymptotic performance measure of LS channels.
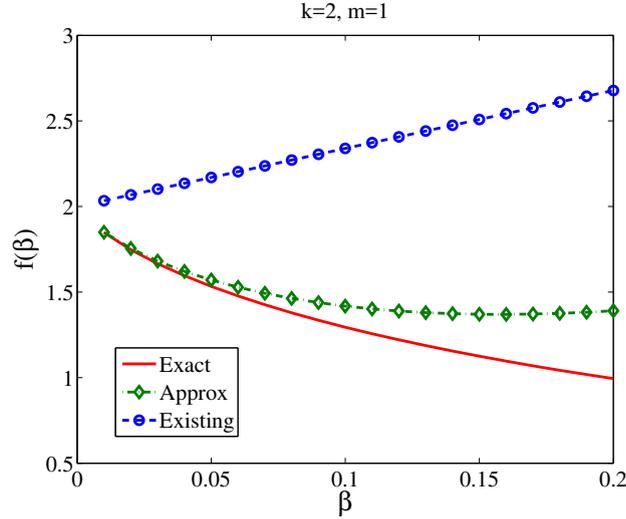
FIGURE 5.1: PDF near origin

Earlier in the thesis it was mentioned Taylor series approach to be ineffective for fading channels with logarithmic singularity. We generate a plot with the exact PDF, asymptote used by Wang and Giannakis, and proposed asymptotic approach in (5.1), near the origin to demonstrate our point. From Fig. 5.1, it can be observed that the existing approach is not at all close to the exact PDF and their slopes are completely opposite. To delve into the details behind this disparity, we recall the theory of asymptotic of integrals. The intuition behind the existing Taylor series approach is based on a well-known asymptotic theory, namely, Watson's lemma [107]. Using Watson's lemma an integral can be approximated as an infinite sum,

$$\int_0^{\vartheta} g(x)f(x)dx \approx \sum_{n=0}^{\infty} \frac{g^{(n)}(0)\Gamma(\nu+n+1)}{n!x^{\nu n+1}} \text{ , for } x \to 0 \tag{5.7}$$

This series converges to the integral *iff* $\mid x \mid < \epsilon$. For GG distribution, as there is a logarithmic singularity this condition is not satisfied and therefore, we observe the divergence between exact PDF and existing asymptotic approach in Fig. 5.1. Therefore, direct application of Wang-Giannakis asymptotic approach [80] via the Taylor series to derive the closed form solutions for BER and outage probability can be inaccurate for LS channels.

Now that the basics of asymptotic measures and underneath motivation behind the analysis of LS channels are discussed, (5.1) is applied over fading channels. As GG fading channel is the most popular generalized distribution with logarithmic singularity property, we apply mentioned asymptotic approaches over GG channel and develop the closed-form solutions and numerical results.

5.2.1 Application on Fading channels

The expressions of the variables and constants mentioned in (5.1). Each of the variable and constant can be derived from the PDF of the GG distribution, and the PDF is given by

$$f(\beta) = \frac{2\beta^{(m+k)/2-1}}{\Gamma(m)\Gamma(k)} \left(\frac{km}{\theta_0}\right)^{(m+k)/2} K_{k-m}\left(2\sqrt{\frac{km\beta}{\theta_0}}\right), \tag{5.8}$$

where $K_n(\cdot)$ is the modified Bessel function of second kind with order of $k - m$, m and θ_0 are the shaping parameter and the average local mean of β, respectively [108]. $K_n(x)$ is one of the solutions of the modified Bessel differential equation,

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2 + n^2)y = 0 \tag{5.9}$$

The order of the Bessel function determines the decaying rate of noise and m determines the shape of the GG density curve, and therefore, determines the affects the probability of successfully decoding the received signal. Average local mean θ_0 is a measure of the scattering, where greater value indicates greater scattering.

Since the asymptote of PDF $f(\beta)$ depends on the polynomial terms of β, we derive a series expansion of the GG PDF. A sum formula for the modified Bessel function with order of n is given by Abramowitz and Stegun [[109], 9.6.11],

$$\begin{aligned}K_n(x) = \frac{1}{2}\left(\frac{x}{2}\right)^{-n}\sum_{k=0}^{n-1}\frac{(n-k-1)!}{k!}\left(-\frac{x^2}{4}\right)^k + (-1)^{n+1}\log\frac{x}{2}\\ \times I_n(x) + (-1)^n\frac{1}{2}\left(\frac{x}{2}\right)^n\sum_{k=0}^{\infty}(\psi(k+1) + \psi(n+k+1))\\ \times\frac{(x^2/4)^k}{k!(n+k)!},\end{aligned} \tag{5.10}$$

73

where $\psi(\cdot)$ is the logarithmic derivative of the gamma function, commonly known as the digamma function and $I_n(\cdot)$ is the modified Bessel function of first kind. Now, we use the asymptotic approximation of the Bessel function of second kind, given in (5.10) to derive the expressions for a, b, t, and μ of (5.1). Replacing n with $k - m$ and x with $\left(2\sqrt{\frac{km}{\theta_0}}\right)$ in (5.10) and thereafter considering the first term in the series we obtain,

$$
\begin{aligned}
K_{k-m}\left(2\sqrt{\frac{km\beta}{\theta_0}}\right) = {} & \frac{(k-m-1)!}{2}\left(\frac{km}{\theta_0}\right)^{(m-k)/2} + \frac{(-1)^{k-m}}{2(k-m)!}\beta^{(k-m)/2} \\
& \times \left(\frac{km}{\theta_0}\right)^{(k-m)/2}\left(\psi(1) + \psi(k-m+1) - \log\left(\frac{km}{\theta_0}\right)\right) \\
& + \frac{(-1)^{k-m+1}}{2(k-m)!}\left(\frac{km}{\theta_0}\right)^{(k-m)/2}\beta^{(k-m)/2}\log(\beta)
\end{aligned}
\tag{5.11}
$$

We would like to mention that $(\psi(1) = -\gamma)$ is a constant, where γ is the Euler-Mascheroni constant.

Now replacing $K_{k-m}(\cdot)$ from (5.11) in (5.8) we obtain,

$$
\begin{aligned}
f(\beta) = {} & \frac{(k-m-1)!}{\Gamma(m)\Gamma(k)}\left(\frac{km}{\theta_0}\right)^m \beta^{(3m-k)/2-1} + \frac{(-1)^{k-m+1}}{(k-m)!\Gamma(m)\Gamma(k)} \\
& \times \left(\frac{km}{\theta_0}\right)^k \beta^{k-1}\left(\gamma + \log\left(\frac{km}{\theta_0}\right) - \psi(k-m+1)\right) \\
& + \frac{(-1)^{k-m+1}}{(k-m)!\Gamma(m)\Gamma(k)}\left(\frac{km}{\theta_0}\right)^k \beta^{k-1}\log(\beta) + R(\beta)
\end{aligned}
\tag{5.12}
$$

As mentioned earlier, only dominant terms are considered in asymptotic measures. Therefore, the second term in (5.12) can be ignored, and consequently the parameters of interest, a, b, t, and μ are determined.

$$
\begin{aligned}
t &= \left(\frac{3m-k}{2}\right) - 1 \\
\mu &= k - 1 \\
a &= \frac{(k-m-1)!}{\Gamma(m)\Gamma(k)}\left(\frac{km}{\theta_0}\right)^m \\
b &= \frac{(-1)^{k-m+1}}{(k-m)!\Gamma(m)\Gamma(k)}\left(\frac{km}{\theta_0}\right)^k
\end{aligned}
\tag{5.13}
$$

It is already mentioned that this approach can be used for generalized-K distribution too. In various literature generalized-K is same as the GG distribution. However, in optics literature generalized-K stands for a more general distribution. Barakat in his seminal

paper came up with this distribution to analyze weak-scattering radiation [110],

$$f(\beta) = \frac{2\vartheta}{\Gamma(\vartheta)\eta^{\vartheta+1}}\left(\frac{\vartheta}{1 + \frac{\xi^2}{4\vartheta}}\right)^{(\vartheta-1)/2}\beta^{(\vartheta-1)/2}I_0\left(\frac{\xi}{\eta}\beta^{1/2}\right)$$
$$\times K_{\vartheta-1}\left(\frac{2}{\eta}\left[\left(1 + \frac{\xi^2}{4\vartheta}\right)\vartheta\beta\right]^{1/2}\right), \tag{5.14}$$

where $I_0(\cdot)$ is the modified Bessel function of first kind with order of zero, and it can be expressed in the form of following series [[109], 9.6.12],

$$I_0(x) = \sum_{k=0}^{\infty}\frac{\left(\frac{1}{4}x^2\right)^k}{(k!)^2} \tag{5.15}$$

Using (5.10) and (5.15) we can derive the asymptotic performance measures for generalized-K distribution too.

5.3   Closed-form solutions

In this section the closed-form solutions of important performance measures, BER and outage probability are derived. BER expressions for BPSK, DBPSK and M-ary PSK is calculated. Asymptotes of density functions for different antenna diversity schemes are explored too.

5.3.1   BER

In fading channels, the average BER is taken over the possible ranges of signal strength due to fading [111]. It is calculated by integrating the conditional error probability (CEP) $P_b(\overline{\gamma})$ of a modulation scheme over the PDF of the instantaneous SNR $f(\beta)$,

$$P_e(\overline{\gamma}) = \int_0^{\infty} P_b(\overline{\gamma})f(\beta)d\beta \tag{5.16}$$

Various expressions for the binary modulations are available in the literature, mainly using the Q-function, however, a general form of CEP for binary modulations that works for BPSK, DBPSK, and BFSK (binary frequency shift keying) is given by [112],

$$P_b(\overline{\gamma}) = \frac{\Gamma(\lambda, \nu\overline{\gamma}\beta)}{2\Gamma(\lambda)} \tag{5.17}$$

where $\Gamma(\lambda, \nu x)$ denote the upper incomplete gamma function, and λ, ν are the modulation dependent parameters. $\Gamma(\cdot, \cdot)$ is defined by,

$$\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt \tag{5.18}$$

For BPSK, DBPSK, and BFSK modulation, (ν, λ) are (1,0.5), (1,1), and (0.5,0.5), respectively [112]. BER for binary modulations can be evaluated by substituting (5.17) in (5.16),

$$P_e(\overline{\gamma}) = \int_0^\infty \frac{\Gamma(\lambda, \nu\overline{\gamma}\beta)}{2\Gamma(\lambda)} \left(a\beta^t + b\beta^\mu \log(\beta)\right) d\beta \tag{5.19}$$

Now following properties of incomplete gamma function [[113], 8.352.4, 8.359.3] are used to derive the closed form solutions for BER,

$$\Gamma(0.5, x^2) = \sqrt{\pi}\left(1 - \Phi(x)\right)$$
$$\Gamma(n, x) = (n-1)! e^{-x} \sum_{m=0}^{n-1} \frac{x^m}{m!} \qquad \text{for } n \in \mathbb{I} \tag{5.20}$$

Applying the first property in (5.20), BER for BPSK modulation using (5.1) is given by

$$P_e(\overline{\gamma}) = \frac{\sqrt{\pi}}{2\Gamma(\lambda)} \int_0^\infty \left(1 - \Phi\left(\sqrt{\frac{\nu\overline{\gamma}\beta}{2}}\right)\right)(a\beta^t + b\beta^\mu \log(\beta)) d\beta \tag{5.21}$$

Theorem 4. *Closed-form solution to evaluate BER for BPSK modulation is given by*

$$P_e(\overline{\gamma}) = \frac{a\Gamma(t + 3/2)}{(2t + 2)(\nu\overline{\gamma}/2)^{2t}\Gamma(\lambda)} + \frac{b\Gamma(\mu + 1/2)}{(\nu\overline{\gamma}/2)^{2\mu}\Gamma(\lambda)}\left(\frac{1}{2\mu + 2} + \frac{2\mu + 1}{2}\right.$$
$$\left. \times \psi\left(\frac{2\mu + 1}{2}\right)\right) - \frac{b\Gamma(\mu + 3/2)}{(2\mu + 2)(\nu\overline{\gamma}/2)^{2\mu}\Gamma(\lambda)} \log(\nu\overline{\gamma}/2) \tag{5.22}$$

Details of the derivation is given in Appendix B.1.

From the closed-form solution in (5.22) we can easily identify the terms corresponding to c_1, c_2, and c_3 in (5.3). Closed-form solutions using proposition 1 are omitted for

brevity.

Theorem 5. *Applying the second property in (5.20), closed-form expression to evaluate BER for DBPSK is given by*

$$P_e(\overline{\gamma}) = \frac{(\lambda - 1)!}{2\Gamma(\lambda)} \sum_{m=0}^{\lambda-1} \frac{(v\overline{\gamma})^m}{m!} \left(\frac{a\Gamma(t + m + 1)}{(v\overline{\gamma})^{t+m+1}} + \frac{b\Gamma(\mu + m + 1)}{(v\overline{\gamma})^{\mu+m+1}} \right.$$
$$\left. \times \Big(\psi(\mu + m + 1) - \log(v\overline{\gamma}) \Big) \right) \tag{5.23}$$

Details of the derivation is given in Appendix B.2.

Another important family of modulation is M-ary PSK. For M=2, PSK revert to binary modulation, and corresponding BER is already derived, however, for $M > 2$, CEP can not be expressed in closed form. Generic SEP for M-ary PSK is given by

$$P_b(\overline{\gamma}) = \frac{1}{\pi} \int_0^{\frac{(M-1)\pi}{M}} \exp\left(- \frac{\beta\overline{\gamma} \sin^2(\pi/M)}{\sin^2(\theta)} \right) d\theta \tag{5.24}$$

Theorem 6. *Using the CEP given in (5.24), closed-form solution for BER for M-ary PSK is given by (5.25). Details of the derivation is discussed in Appendix B.3.*

$$P_e(\overline{\gamma}) = \frac{a\Gamma(t + 1)}{\pi\overline{\gamma}^{t+1} \sin^{2t+2}(\pi/M)} \left(\frac{(t + 1)(M - 1)\pi}{2M} - \frac{(t + 1)\sin(2(M - 1)\pi/M)}{4} \right)$$
$$+ \frac{b\big(\psi(\mu + 1) - 2\Gamma(m + 1)\log(\sin(\pi/M))\big)}{\pi\overline{\gamma}^{\mu+1} \sin^{2\mu+2}(\pi/M)} \times \left(\frac{(\mu + 1)(M - 1)\pi}{2M} - \frac{(\mu + 1)\sin(2(M - 1)\pi/M)}{4} \right)$$
$$+ \left[\frac{(\sin^{-1}(\frac{(M-1)\pi}{M}))^{1+2\mu}}{(1 + 2k)^2} \left({}_3F_2\left[-\frac{1}{2}, \frac{1}{2} + \mu, \frac{1}{2} + \mu, \frac{3}{2} + \mu, \frac{3}{2} + \mu, \Big(\sin^{-1}\Big(\frac{(M - 1)\pi}{M}\Big)\Big)^2 \right] \right.$$
$$\left. - {}_3F_2\left[\frac{1}{2}, \frac{1}{2} + \mu, \frac{1}{2} + \mu, \frac{3}{2} + \mu, \frac{3}{2} + \mu, \Big(\sin^{-1}\Big(\frac{(M - 1)\pi}{M}\Big)\Big)^2 \right] - (1 + 2\mu)\log\Big(\sin^{-1}\Big(\frac{(M - 1)\pi}{M}\Big)\Big) \right.$$
$$\times \left({}_2F_1\left[-\frac{1}{2}, \frac{1}{2} + \mu, \frac{3}{2} + \mu, \Big(\sin^{-1}\Big(\frac{(M - 1)\pi}{M}\Big)\Big)^2 \right] \right.$$
$$\left. - {}_2F_1\left[\frac{1}{2}, \frac{1}{2} + \mu, \frac{3}{2} + \mu, \Big(\sin^{-1}\Big(\frac{(M - 1)\pi}{M}\Big)\Big)^2 \right] \right) \right) \bigg] \frac{2b\Gamma(\mu + 1)}{\pi\overline{\gamma}^{\mu+1} \sin^{2\mu+2}(\pi/M)} \tag{5.25}$$

It can be observed that replacing M by 2 simplifies the modulation scheme to binary modulation and therefore, the BER in (5.25) reduces to a similar expression given in (5.22). However, for $M > 2$, there is no such simple closed-form solution.

5.3.2 Outage Probability

In addition to the BER, outage probability is another important performance measure for fading channels. There are various definitions of outage probability, mainly based on symbol error rate and SNR. For this chapter let us consider the definition given by Stuber [71], outage probability is the probability of instantaneous SNR γ dropping below a certain threshold γ_{th}. Therefore, it can be evaluated using,

$$P_{out}(\gamma_{th}, \overline{\gamma}) = P[\gamma \leq \gamma_{th}] = P[\beta\overline{\gamma} \leq \gamma_{th}] \tag{5.26}$$

As SNR is always non-negative, therefore, the outage probability is simply the cdf of β,

$$P_{out}(\gamma_{th}, \overline{\gamma}) = \int_0^{\gamma_{th}/\overline{\gamma}} f(\beta)d\beta = F_\beta\left(\frac{\gamma_{th}}{\overline{\gamma}}\right) \tag{5.27}$$

where F_β is the cdf of β. Now exact cdf of GG distribution is not computation friendly and therefore, it is important to use asymptotes of the PDF to obtain a computation friendly expression. Integrating the asymptote in (5.1) we obtain,

$$P_{out}(\gamma_{th}, \overline{\gamma}) = \frac{\gamma_{th}}{\overline{\gamma}}\left(\frac{a(\gamma_{th}/\overline{\gamma})^t}{1+t} + \frac{b(\gamma_{th}/\overline{\gamma})^\mu((1+\mu)\log(\gamma_{th}/\overline{\gamma}) - 1)}{(1+\mu)^2}\right) \tag{5.28}$$

Wang and Giannakis also observed that outage probability can be expressed as a function of outage diversity and coding gain, and similar to the error probability in (2.11), they expressed outage probability as [80],

$$P_{out} = (O_c\overline{\gamma})^{-O_d} + R(\overline{\gamma}), \tag{5.29}$$

where O_d and O_c are outage diversity and coding gain, respectively.

For LS wireless channels (5.29) does not hold, we express (5.28) in a more generalized form. Similar to (5.2), we derive a generalized form of outage probability using outage diversity and coding gain that works for channels with and without logarithmic singularity. Using a derivation similar to (5.5) outage probability can be expressed as

$$P_{out} = (O_c\overline{\gamma})^{-O_d} \exp\left(\frac{c'}{\overline{\gamma}^\alpha}\left(\eta \log(\gamma_{th}/\overline{\gamma}) - 1\right)\right), \tag{5.30}$$

where $O_c = \dfrac{1}{\gamma_{th}}\left(\dfrac{a}{t+1}\right)^{-1/(t+1)}$, $O_d = 1+t$, $c' = \dfrac{b\gamma_{th}^{\mu-t}(t+1)}{a(1+\mu)^2}$, $\alpha = \mu - t$, and $\eta = 1 + \mu$.

5.3.3 Antenna Diversity

Antenna diversity is a technique to mitigate the effects of short-term fading. Underlying motive behind the antenna diversity techniques is the high probability of receiving unaffected transmitted signal, provided the receiving antennas are efficiently spaced. Several combining methods have been proposed to leverage the signal statistics using the antenna diversity. However, calculation of exact BER and outage probability for different antenna diversity models is challenging for GG distribution, as the resulting PDF for greater number of antenna ($L > 2$) is extremely difficult to calculate.

Bithas *et al.* on their seminal paper derived the results of diversity models for Generalized-K distribution using Pade approximation [114]. Wang and Giannakis proposed an alternative method to calculate the BER, they extrapolated the simple coding and diversity gain to obtain the coding and diversity gain for different combing schemes [80]. However, using these methods over GG model lead to extremely large equations. Therefore, we derive the asymptotic PDFs for the considered antenna diversity schemes. Expressions for outage probability, BER, and other performance measures can be derived using these PDFs. In this chapter selection combining (SC) and maximal ratio combining (MRC) are considered as the diversity schemes.

5.3.3.1 Selection Combining

Simplest of the combining schemes, SC, measures the SNR at each branch and selects the branch with the highest SNR. Therefore, for a L-branch system, resulting β is given by

$$\beta_{SC} = \max_{1 \le l \le L} \beta_l \tag{5.31}$$

Now assuming the β to be iid over all the branches and using the proposed asymptote in (5.1), we derive the cdf for SC scheme,

$$F_{SC}(\beta) = \left(\frac{a\beta^{t+1}}{t+1} + \frac{b\beta^{\mu+1}(\mu+1)\log(\beta)-1)}{(\mu+1)^2}\right)^L \tag{5.32}$$

79

Finally, differentiating the right-hand-side (rhs) of (5.32) w.r.t β gives the PDF of interest,

$$f_{SC}(\beta) = L\left(\frac{a\beta^{t+1}}{1+t} + \frac{b\beta^{\mu+1}\log(\beta)}{1+\mu} - \frac{b\beta^{\mu+1}}{(1+\mu)^2}\right)^{L-1}\left(a\beta^t + b\beta^\mu\right.$$
$$\left.\times\left(\frac{1+\log(\beta)}{1+\mu} - \frac{1}{(1+\mu)^2}\right) + b\mu\beta^\mu\left(\frac{\log(\beta)}{1+\mu} - \frac{1}{(1+\mu)^2}\right)\right) \quad (5.33)$$

5.3.3.2 Maximal Ratio Combining

Among the linear diversity combiners, MRC performs best in a multipath fading environment [115]. Unlike SC, where signal from only one branch is considered, signals are weighted according to their SNR and summed up in MRC. Therefore, the resulting signal power is sum of the signals from all the antennas. Hence the resulting β for MRC is given by

$$\beta_{MRC} = \sum_{l=1}^{L} \beta_l \quad (5.34)$$

Similar to SC, we consider β to be iid over all the branches and obtain the cdf for MRC,

$$F_{MRC}(\beta) = \frac{a(\beta/L)^{t+1}}{t+1} + \frac{b(\beta/L)^{\mu+1}(\mu+1)\log(\beta/L) - 1)}{(\mu+1)^2} \quad (5.35)$$

Now differentiating the rhs of (5.35) w.r.t β we obtain the PDF for L-branch MRC,

$$f_{MRC}(\beta) = \frac{a\beta^t}{L^{t+1}} + \frac{b\beta^\mu}{(1+\mu)L^{1+\mu}} + \frac{b\beta^\mu\left((1+\mu)(\log(\beta/L)) - 1\right)}{(1+\mu)L^{1+\mu}} \quad (5.36)$$

One important advantage of the proposed asymptote is the availability of concise closed-form solutions. Therefore, asymptotic approach in (5.1) is useful to solve resource allocation related optimization problems. Comparing the derived PDFs in (5.33) and (5.36) we observe that effect of L is different over the asymptotic measures for MRC and SC and as we will discuss in the next section, L plays an important role in the divergence of asymptotic measures.

5.4 Numerical Results

Numerical results are generated for BER and outage probability. For BER we consider both single antenna and antenna diversity schemes. Throughout this section we will use following legends, *Exact, Approx1, Approx2, Existing1,* and *Existing 2* to represent exact solution, solution using the asymptote in (5.1), using the asymptote in (5.6), Taylor series based approach used by Wang and Giannakis, and asymptotic measure given by Dhungana and Tellambura in [88], respectively.

Outage probability is calculated for two different cases, ($k = 2$, $m = 1$, $\theta = 1$) and ($k = 3$, $m = 2$, $\theta = 1$), and γ_{th} is assumed to be 0 *dB*. Performance of the proposed asymptotic measures are compared with two existing asymptotic approaches for SNR varying from $5 - 20$ *dB*. From Fig. 5.2 we observe that proposed approximations give considerable improvement over the existing ones. As mentioned earlier, Approx2 gives better performance than Approx1 in the SNR range of $5 - 10$ dB. It is also observed that approximation by Dhungana and Tellambura gives better performance than Wang-Giannakis model but worse than Approx1 even after using two terms from the series expansion.
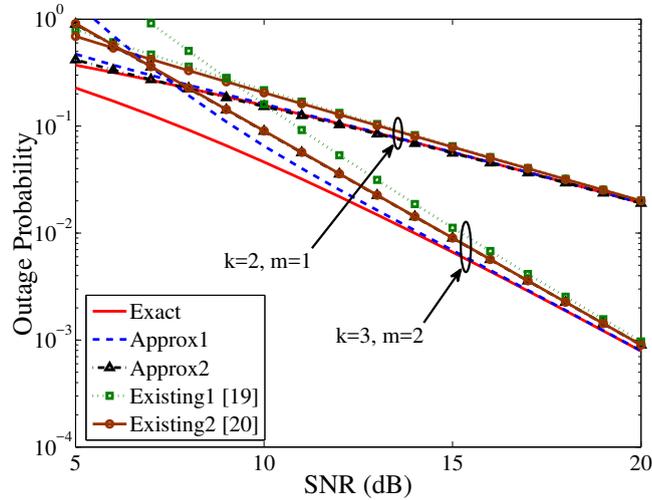
FIGURE 5.2: Outage probability for varying SNR.

The relative error for all the approximations are plotted as well, where relative error is calculated using (5.37),

$$\hat{e} = \frac{|\,approx - exact\,|}{exact} \tag{5.37}$$

81
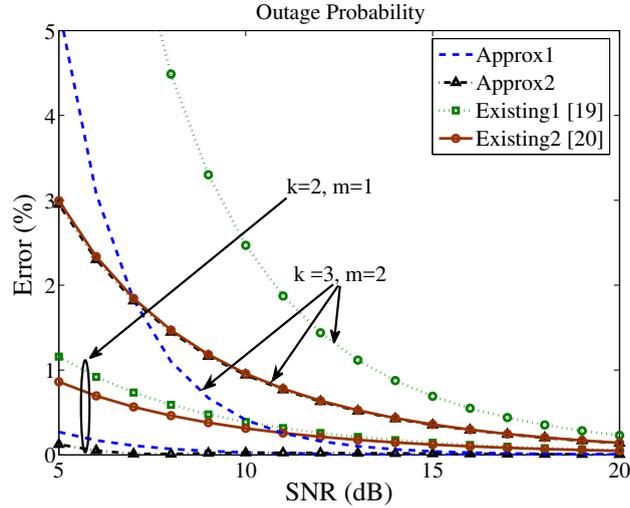
FIGURE 5.3: Approximation error of outage probability.

Errors from the proposed and existing asymptotic measures are compared in Fig. 5.3. We observe significant improvement using the proposed asymptotic approach in the range of $8 - 20$ dB.
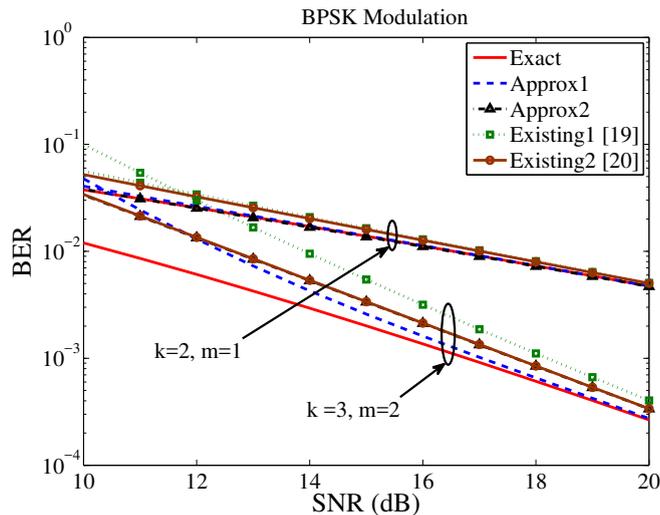
FIGURE 5.4: BER for BPSK modulation with varying SNR.

Now let us compare the asymptotic measures for BPSK, DBPSK and M-ary PSK modulation and observe the performance by varying SNR from $5 - 20$ dB. From Figs. 5.4-5.6 we observe Approx1 performs better for higher SNR, this is the primary shortfall of asymptotic approaches, which converges to the exact one in infinity. It can also be observed that Approx2 performs better than other asymptotes in the region of $10 - 14$ dB, and it is the effect of including the additional exponential function which reduces the divergence to a great extent in the lower SNR regime $(0 - 10)$ dB. Please keep in mind
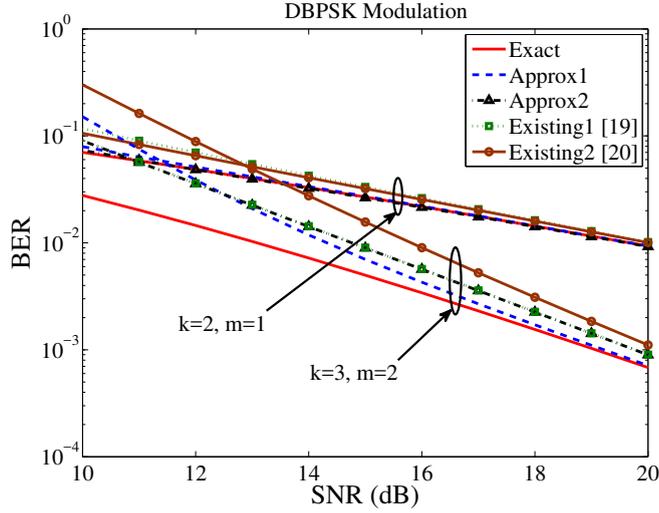
FIGURE 5.5: BER for DBPSK modulation with varying SNR.
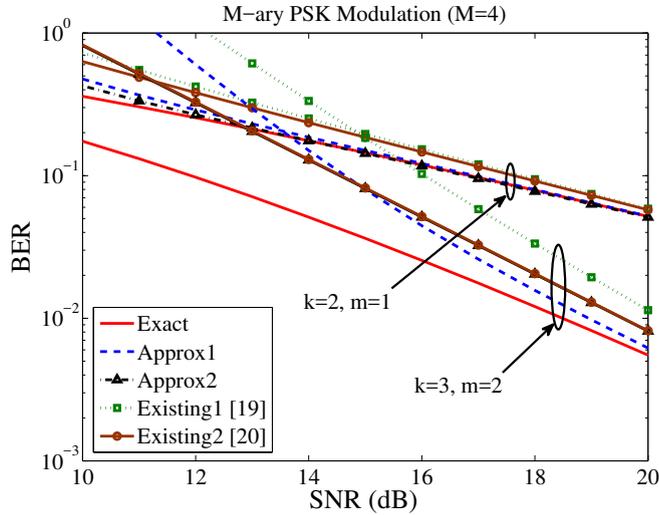
FIGURE 5.6: BER for M-ary PSK modulation with varying SNR.

that only a single term from the series is used in Approx1, and therefore, inclusion of more terms will certainly enhance the performance.

Performance for antenna diversity schemes is given in Figs. 5.7-5.10 for varying SNR. We use two different ranges of SNR for SC and MRC to clearly illustrate the difference between the asymptotic measures. For visual clarity we omitted the Existing2 approximation, however, we observed it's performance to be similar to the earlier results. One important observation from Figs. 5.7-5.10 is significantly better performance using the asymptotic approaches for MRC. These results can be explained by variable L's opposing influence on the PDFs of SC and MRC. From the PDF of SC we observe that L is exponentially proportional to $f_{SC}(\beta)$, therefore, at SNR range of $5-10$ dB asymptotes
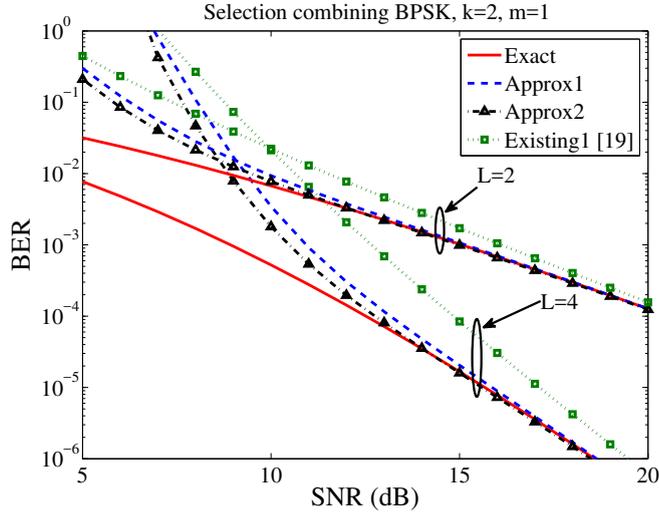
FIGURE 5.7: BER for selection combining and BPSK with varying SNR.
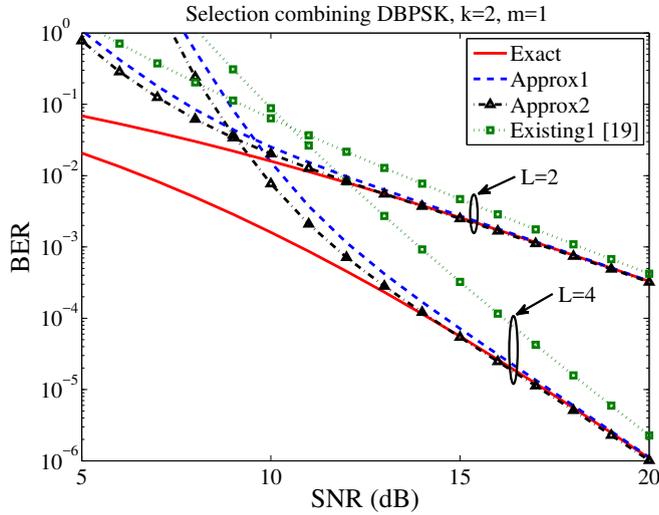
FIGURE 5.8: BER for selection combining and DBPSK with varying SNR.

deviate exponentially from the exact PDF and we obtain the results in Figs. 5.7 and 5.8. However, in case of MRC, variable L is inversely proportional to $f_{MRC}(\beta)$, therefore, even for SNR range of $0 - 10$ dB, Approx1 closely follows the Exact result.

In a nutshell, from the figures we observe the proposed asymptotic measures closely follow exact result. However, we should keep in mind that we are considering only one term from the series for Approx1. Including more terms in the asymptote understandably reduces the error. Now the effects of including additional terms is studied using a well-known statistical test.
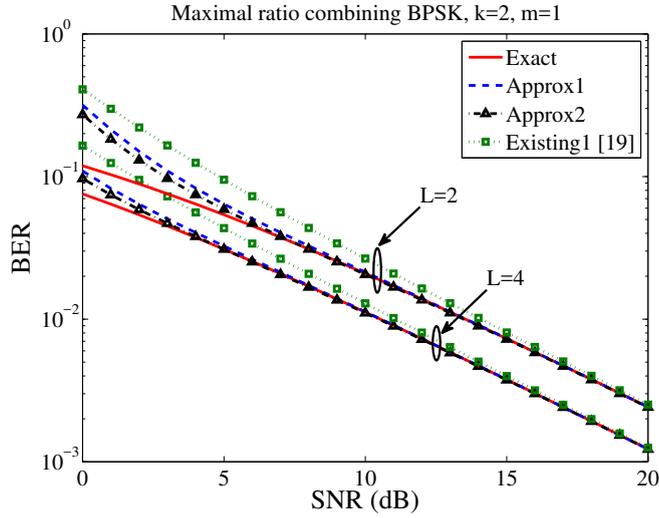
FIGURE 5.9: BER for maximal ratio combining and BPSK with varying SNR.
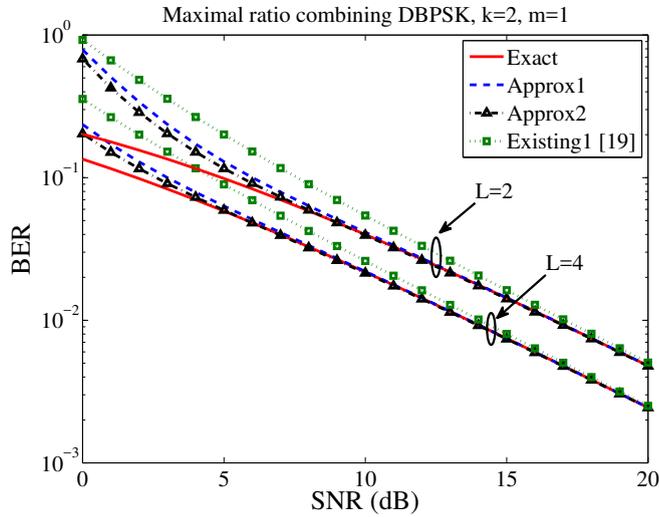
FIGURE 5.10: BER for maximal ratio combining and DBPSK with varying SNR.

5.4.1 KS Test

In this section the closeness of the proposed asymptotic measures, and effect of considering additional terms is measured using the Kolmogorov-Smirnov (KS) goodness-of-fit test [116]. KS test is a null hypothesis test where T-statistic is evaluated from the absolute difference between empirical cumulative distribution function (CDF) and approximated CDF. KS test statistic is defined as,

$$T \triangleq \max | F_\beta(z) - \hat{F}_\beta(z) | \tag{5.38}$$

85
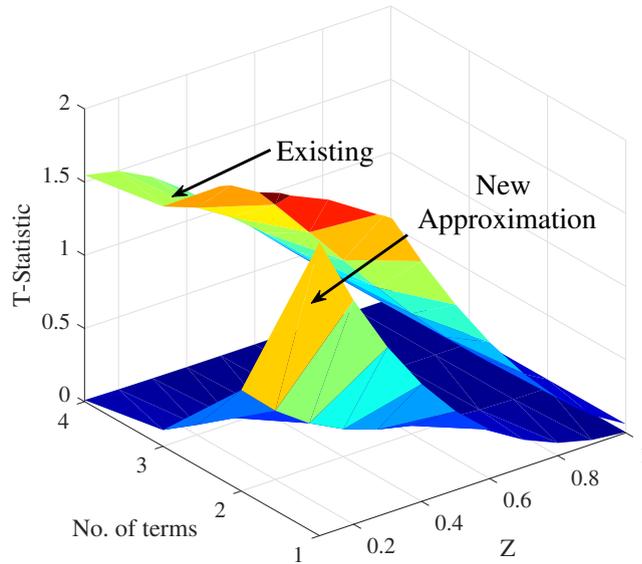
FIGURE 5.11: KS test statistic.

Till now, all the results are generated by considering a single term from the series. Now, let us explore the effect of including more terms in new asymptote, given by (5.1) and asymptotic approach by Wang and Giannakis. We compute the T-statistic for varying number of terms $(1 - 4)$ and varying z, and define a null hypothesis to determine the significance of both approaches.

- *Definition:* We define H_0 as the null hypothesis, that accepts a measurement with significance of 95% when corresponding T is less than a threshold (T_{max}), and similarly rejects a measurement with 5% significance when T is greater than than T_{max}.

The standard value of $T_{max} = 0.05$ is used to compare the performance. Test statistics for distribution parameter $k = 2$, $m = 1$, varying z from $0 - 1$ and number of terms from $1 - 4$, is given in Fig. 5.11. From the results it is evident that the performance of the new asymptotic approach is significantly better than the existing approach, the proposed asymptote with one term outperforms the Taylor series based asymptote, even with four terms. It is observed that the hypothesis H_0 does not accept any of the results using the existing asymptotic approach with 95% significance, whereas all the results using the proposed asymptote with four terms are accepted with 95% significance.

Even though all the observations generated using the new asymptote with a single term from the series are classified as unacceptable according to the hypothesis H_0, it is evident from the results that the new asymptotic approach gives significant improvement over the existing one for GG channels. Moreover, most of the results using the proposed asymptotic approach with two terms are considered acceptable by the hypothesis H_0.

5.5   Conclusion

As mentioned before, an important property of fading models, logarithmic singularity, has been overlooked in earlier works. There are few subtle differences between the channels with and without the logarithmic singularity, such as, Taylor series based asymptotic approaches are not applicable, and the classical coding and diversity gain model does not hold for LS channels. To circumvent these shortfalls we have proposed new asymptotic measures and developed a generalized coding and diversity gain model that is applicable for fading models with and without the logarithmic singularity. Moreover, generalized fading models are observed to follow this logarithmic singularity property and therefore, it is extremely important to develop high SNR asymptotic analysis. We also know that the Gamma-Gamma (GG) distribution is a versatile model for the propagation modes scattering, reflection, and diffraction and optical, navigation and relay channels [108]. Moreover, it is useful for analysis of future communication systems such as FSO MIMO, HetNets with co-located interference too, and therefore, requires serious attention. However, as GG PDF contains logarithmic singularity, existing asymptotic approaches generate significant amount of error. Therefore, an asymptotic analysis based on the proposed asymptote is developed, and also several asymptotic measures for error and outage probability are derived. Numerical results are provided for different modulation and antenna diversity schemes and parameters. Results suggest significant improved accuracy in the SNR range of 10-25 dB. For this range, the proposed asymptotic model achieves much better accuracy than the conventional, existing model. To increase the SNR range we another asymptotic approach is proposed, which shows significant improvement in the range of 5-15 dB.

Chapter 6

Conclusion and Future Research Directions

6.1 Conclusions

This thesis explored several aspects of a wireless cache enabled network: a) developing a greedy algorithm based caching strategy to reduce latency, b) investigating the effects of mobility, and c) establishing a new asymptotic measure for generalized fading channels which are suitable for interference characterization of cache networks and many other scenarios.

Chapter 2 provided the required mathematical and theoretical background for the thesis. First, state-of-the-art caching strategies for wired backbone networks are discussed, as are several works on wireless cache enabled networks. Second, statistical tools to model the spatial and temporal distribution of the nodes in a cellular network using point processes and popular mobility models are studied. Finally, various aspects of wireless channel modeling, usefulness of asymptotic analysis of a fading channel are briefly discussed.

In Chapter 3 a greedy algorithm based caching policy is studied. A greedy algorithm is applied over an optimization problem to maximize cache hit rate. Greedy caching is employed over several real-life networks and scale-free networks.

Chapter 4 investigated the effect of user mobility on delay performance for a cache network. In this chapter caching abilities at edge caches or wireless base stations are

considered. Thereafter, an expression for handover probabilities and delay is derived, and performance for different handover management policies are discussed. Also the effect of cache size and content popularity skewness over delay performance is studied.

Chapter 5 developed a new asymptotic measure for wireless channels that includes channel models with logarithmic singularity. Different modulation schemes and antenna diversity models are considered and the performance of the new asymptotic measure is compared with the existing ones.

6.2 Future Research Directions

Several improvements and extensions are proposed for future research problems.

- **Development of an optimal caching strategy**
  In this thesis an optimized caching strategy using greedy algorithm is proposed. However, as discussed in chapter 2 it is not an optimal solution. So developing an optimal caching policy to maximize hit rate is still an open problem. Moreover, a study related to the hardness of the optimization problem, whether it is np-complete, is highly relevant.

- **Optimal mobility-aware caching**
  Development of an optimal mobility-aware caching strategy is a highly challenging open problem. Existing works consider that cell sojourn time is available at the BS, however, it is an unrealistic assumption due to the dynamic nature of user mobility. Therefore, an optimal algorithm with dynamic sojourn time calculation is a research problem worth solving.

- **Analyzing group mobility models**
  In this thesis a general RWP mobility model and basic caching strategy is discussed. A study of group mobility models intertwined with sophisticated caching strategies and their impact on user-level performance is an important research problem.

- **Interference characterization of small-cell cache network**
  Using the asymptotic analysis of GG channels, interference analysis of a cache network, especially incorporating cache size and corresponding hit rate effects, is an important research problem.

Bibliography

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2016-2021," Tech. Rep., 2017.

[2] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selén, and J. Sköld, "5G wireless access: requirements and realization," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 42–47, Dec. 2014.

[3] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, Nov. 2014.

[4] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-cell Networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[5] L. Saino, I. Psaras, and G. Pavlou, "Icarus: a caching simulator for information centric networking (ICN)," in *ACM SimuTools*, Mar. 2014, pp. 66–75.

[6] W. Hoiles, O. N. Gharehshiran, V. Krishnamurthy, and N. Dào, "Adaptive caching in the YouTube content distribution network: A revealed preference game-theoretic learning approach," *IEEE Transactions on Cognitive Communications and Networking*, vol. 1, no. 1, pp. 71 – 85, Oct. 2015.

[7] "Netflix open connect." [Online]. Available: https://openconnect.netflix.com/en/

[8] M. Zhang, H. Luo, and H. Zhang, "A survey of caching mechanisms in information-centric networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1473 – 1499, Thirdquarter 2015.

[9] A. Dabirmoghaddam, M. M. Barijough, and J. Garcia-Luna-Aceves, "Understanding optimal caching and opportunistic caching at "the edge" of information-centric networks," in *ACM ICN*, Sept. 2014, pp. 47–56.

[10] N. Laoutaris, H. Che, and I. Stavrakakis, "The LCD interconnection of LRU caches and its analysis," *Elsevier Performance Evaluation*, vol. 63, no. 7, pp. 609–634, July 2006.

[11] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache less for more in information-centric networks," in *IFIP Networking*, May 2012, pp. 27–40.

[12] Y. Li, T. Lin, H. Tang, and P. Sun, "A chunk caching location and searching scheme in content centric networking," in *IEEE ICC*, June 2012, pp. 1550 – 3607.

[13] I. Psaras, W. K. Chai, and G. Pavlou, "In-network cache management and resource allocation for information-centric networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 11, pp. 2920 – 2931, Nov. 2014.

[14] K. Suksomboon, S. Tarnoi, Y. Ji, M. Koibuchi, K. Fukuda, S. Abe, N. Motonori, M. Aoki, S. Urushidani, and S. Yamada, "PopCache: Cache more or less based on content popularity for information-centric networking," in *IEEE LCN*, Oct. 2014, pp. 236–243.

[15] A. Ioannou and S. Weber, "Towards on-path caching alternatives in information-centric networks," in *IEEE LCN*, Sept. 2014, pp. 362 – 365.

[16] C. Bernardini, T. Silverston, and O. Festor, "MPC: Popularity based caching strategy for content centric networks," in *IEEE ICC*, June 2013, pp. 3619 – 3623.

[17] J. M. Wang and B. Bensaou, "Progressive caching in CCN," in *IEEE GLOBECOM*, Dec. 2012, pp. 2727 – 2732.

[18] W. Wang, Y. Sun, Y. Guo, D. Kaafar, J. Jin, J. Li, and Z. Li, "CRCache: Exploiting the correlation between content popularity and network topology information for ICN caching," in *IEEE ICC*, June 2014, pp. 3191 – 3196.

[19] Y. Wang, M. Xu, and Z. Feng, "Hop-based probabilistic caching for information-centric networks," in *IEEE GLOBECOM*, Dec. 2013, pp. 2102 – 2107.

[20] S. Saha, A. Lukyanenko, and A. Ylä-Jääski, "Cooperative caching through routing control in information-centric networks," in *IEEE INFOCOM*, Apr. 2013, pp. 100 – 104.

[21] L. Saino, I. Psaras, and G. Pavlou, "Hash-routing schemes for information centric networking," in *ACM ICN*, Aug. 2013, pp. 27–32.

[22] S. Wang, J. Bi, J. Wu, and A. V. Vasilakos, "CPHR: In-network caching for information-centric networking with partitioning and hash-routing," *IEEE/ACM Transactions on Networking*, vol. 5, no. 24, pp. 2742–2755, Oct. 2016.

[23] V. Sourlas, L. Gkatzikis, P. Flegkas, and L. Tassiulas, "Distributed cache management in information-centric networks," *IEEE Transactions on Network and Service Management*, vol. 10, no. 3, pp. 286–299, Sept. 2013.

[24] B. Banerjee, A. Seetharam, A. Mukherjee, and M. Naskar, "Characteristic time routing in information centric networks," *Elsevier Computer Networks*, vol. 113, pp. 148 – 158, Feb. 2017.

[25] S. Zaman and D. Grosu, "A distributed algorithm for the replica placement problem," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1455 – 1468, Sept. 2011.

[26] L. Qiu, V. Padmanabhan, and G. Voelker, "On the placement of web server replicas," in *IEEE INFOCOM*, vol. 3, Apr. 2001, pp. 1587 – 1596.

[27] J. Kangasharju, J. Roberts, and K. Ross, "Object replication strategies in content distribution networks," *Elsevier Computer Communications*, vol. 25, no. 4, pp. 376 – 383, Mar. 2002.

[28] A. Ghodsi, S. Shenker, T. Koponen, A. Singla, B. Raghavan, and J. Wilcox, "Information-centric networking: Seeing the forest for the trees," in *ACM HotNets-X*, Nov. 2011, pp. 1–6.

[29] L. Zhang *et al.*, "Named data networking," in *ACM SIGCOMM*, vol. 44, no. 3, July 2014, pp. 66–73.

[30] S. Eido and A. Gravey, *How Much LTE Traffic Can Be Offloaded?* Springer International Publishing, 2014, pp. 48–58.

[31] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *IEEE INFOCOM*, Mar. 2012, pp. 1107–1115.

[32] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[33] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.

[34] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 131–145, Jan. 2016.

[35] E. Baştuğ, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," in *IEEE International Symposium on Wireless Communications Systems (ISWCS)*, Aug. 2014, pp. 649–653.

[36] S. Krishnan and H. S. Dhillon, "Distributed caching in device-to-device networks: A stochastic geometry perspective," in *49th Asilomar Conference on Signals, Systems and Computers*, Nov. 2015, pp. 1280–1284.

[37] G. S. Paschos, S. Gitzenis, and L. Tassiulas, "The effect of caching in sustainability of large wireless networks," in *IEEE WiOpt*, May 2012, pp. 355–360.

[38] L. Qiu and G. Cao, "Cache increases the capacity of wireless networks," in *IEEE INFOCOM*, Apr. 2016, pp. 1–9.

[39] A. Liu and V. K. N. Lau, "How much cache is needed to achieve linear capacity scaling in backhaul-limited dense wireless networks?" *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 179–188, Feb. 2017.

[40] G. Zhang, J. Liu, J. Ren, L. Wang, and J. Zhang, "Capacity of content-centric hybrid wireless networks," *IEEE Access*, vol. 5, pp. 1449–1459, Mar. 2017.

[41] E. Baştuğ, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[42] M. R. Bhatnagar and Z. Ghassemlooy, "Performance evaluation of FSO MIMO links in Gamma-Gamma fading with pointing errors," in *IEEE ICC*, June 2015, pp. 5084–5090.

[43] D. Jiang and Y. Cui, "Caching and multicasting in large-scale cache-enabled heterogeneous wireless networks," in *IEEE GLOBECOM*, Dec. 2016, pp. 1–7.

[44] D. Malak, M. Al-Shalash, and J. G. Andrews, "Optimizing content caching to maximize the density of successful receptions in device-to-device networking,"

IEEE Transactions on Communications, vol. 64, no. 10, pp. 4365–4380, Oct. 2016.

[45] J. Song, H. Song, and W. Choi, "Optimal content placement for wireless femto-caching network," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4433–4444, July 2017.

[46] Z. Chen, N. Pappas, and M. Kountouris, "Probabilistic caching in wireless D2D networks: cache hit optimal versus throughput optimal," *IEEE Communications Letters*, vol. 21, no. 3, pp. 584–587, Mar. 2017.

[47] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 156–163, Nov. 2010.

[48] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks, Volume I - Theory*, ser. Foundations and Trends in Networking Vol. 3: No 3-4, pp 249-449.   NoW Publishers, 2009.

[49] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 996–1019, Thirdquarter 2013.

[50] H. ElSawy, A. Sultan-Salem, M. S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 167–203, Firstquarter 2017.

[51] M. Haenggi, *Stochastic Geometry for Wireless Networks*, 1st ed.   Cambridge University Press, 2012.

[52] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 630–643, June 2011.

[53] A. Jardosh, E. Belding-Royer, K. Almeroth, and S. Sur, "Towards realistic mobility models for mobile ad hoc networks," in *ACM Mobicom*, 2003, pp. 217–229.

[54] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *IEEE INFOCOM*, vol. 3, 2005, pp. 1897–1907.

[55] C. Song, T. Koren, P. Wang, and A. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, pp. 818–823, Sept. 2010.

[56] X. Lin, R. K. Ganti, P. J. Flemming, and J. G. Andrews, "Towards understanding the fundamentals of mobility in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1686–1698, Apr. 2013.

[57] R. R. Roy, *Handbook of Mobile Ad Hoc Networks for Mobility Models*, 1st ed. Springer-Verlag New York, 2010.

[58] S. Krishnan and H. S. Dhillon, "Effect of user mobility on the performance of Device-to-Device networks with distributed caching," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 194–197, Apr. 2017.

[59] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: modeling and methodology," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 77–83, Aug. 2016.

[60] R. Wang, J. Zhang, S. H. Song, and K. B. Letaief, "Mobility-aware caching in D2D networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5001–5015, Aug. 2017.

[61] H. Wang, S. Chen, M. Ai, and H. Xu, "Localized mobility management for 5G ultra dense network," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8535–8552, 2017.

[62] F. Giust, L. Cominardi, and C. J. Bernardos, "Distributed mobility management for future 5G networks: overview and analysis of existing approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 142–149, Jan. 2015.

[63] A. F. Molisch, *Wireless Communications*, 2nd ed. Wiley-IEEE Press, Nov. 2010.

[64] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[65] *Urban Transmission Loss Models for Mobile Radio in the 900 and 1,800 MHz Bands*, 1991.

[66] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Review of the Electrical Communication Laboratory*, vol. 16, no. 9-10, pp. 825–873, 1968.

[67] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, Aug. 1980.

[68] M. Nakagami, "The m-distribution—a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Pergamon, 1960, pp. 3 – 36.

[69] S. Atapattu, C. Tellambura, and H. Jiang, "A mixture Gamma distribution to model the SNR of wireless channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4193–4203, Oct. 2011.

[70] M. C. Teich and P. Diament, "Multiply stochastic representations for K distributions and their Poisson transforms," *Optical Engineering*, vol. 6, no. 1, pp. 80–91, Jan. 1989.

[71] G. L. Stuber, *Principles of Mobile Communication*. Springer, 2012.

[72] D. J. Lewinsky, "Nonstationary probabilistic target and clutter scattering models," *IEEE Transactions on Antennas and Propagation*, vol. 31, no. 3, pp. 490–498, May 1983.

[73] P. S. Bithas, N. C. Sagias, and P. T. Mathiopoulos, "The bivariate generalized-K (KG) distribution and its application to diversity receivers," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2655–2662, Sept. 2009.

[74] I. Kostic, "Analytical approach to performance analysis for channel subject to shadowing and fading," *IEE Proceedings - Communications*, vol. 152, no. 6, pp. 821–827, Dec. 2005.

[75] I. Trigui, A. Laourine, S. Affes, and A. Stephenne, "Performance analysis of mobile radio systems over composite fading/shadowing channels with co-located interference," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3448–3453, July 2009.

[76] N. Chatzidiamantis and G. Karagiannidis, "On the distribution of the sum of Gamma-Gamma variates and applications in RF and optical wireless communications," *IEEE Transactions on Communications*, vol. 59, no. 5, pp. 1298–1308, May 2011.

[77] P. Shankar, "Maximal ratio combining (MRC) in shadowed fading channels in presence of shadowed fading cochannel interference (CCI)," *Wireless Personal Communications*, vol. 68, pp. 15–25, Jan. 2013.

[78] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behaviour," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[79] M. Hassan, X. Song, and J. Cheng, "Subcarrier intensity modulated wireless optical communications with rectangular QAM," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 4, no. 6, pp. 522–532, June 2012.

[80] Z. Wang and G. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Transactions on Communication*, vol. 51, no. 8, pp. 1389–1398, Aug. 2003.

[81] C. Tellambura, A. J. Mueller, and V. K. Bhargawa, "Analysis of M-ary phase-shift keying with diversity reception for land-mobile satellite channels," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 4, pp. 910–922, Nov. 1997.

[82] C. Tellambura, A. Annamalai, and V. K. Bhargava, "Closed form and infinite series solutions for the MGF of a dual-diversity selection combiner output in bivariate Nakagami fading," *IEEE Transactions on Communications*, vol. 51, no. 4, pp. 539–542, Apr. 2003.

[83] C. Tellambura, "Evaluation of the exact union bound for trellis-coded modulations over fading channels," *IEEE Transactions on Communications*, vol. 44, no. 12, pp. 1693–1699, Dec. 1996.

[84] S. Kusaladharma and C. Tellambura, "Aggregate interference analysis for underlay cognitive radio networks," *IEEE Wireless Communications Letters*, vol. 1, no. 6, pp. 641–644, Dec. 2012.

[85] ——, "On approximating the cognitive radio aggregate interference," *IEEE Wireless Communications Letters*, vol. 2, no. 1, pp. 58–61, Feb. 2013.

[86] S. Kusaladharma, P. Herath, and C. Tellambura, "Secondary user interference characterization for underlay networks," in *IEEE VTC Fall*, Sept. 2015, pp. 1–5.

[87] S. Kusaladharma and C. Tellambura, "Massive mimo based underlay networks with power control," in *IEEE ICC*, May 2016, pp. 1–6.

[88] Y. Dhungana and C. Tellambura, "New simple approximations for error probability and outage in fading," *IEEE Communications Letters*, vol. 16, no. 11, pp. 1760–1763, Oct. 2012.

[89] ——, "Uniform approximations for wireless performance in fading channels," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4768–4779, Nov. 2013.

[90] V. R. S. Banjade, C. Tellambura, and H. Jiang, "New asymptotics for performance of energy detector," in *IEEE GLOBECOM*, Dec. 2014, pp. 4020–4024.

[91] ——, "Asymptotic performance of energy detector in fading and diversity reception," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2031–2043, June 2015.

[92] A. Annamalai, C. Tellambura, and V. K. Bhargava, "A general method for calculating error probabilities over fading channels," *IEEE Transactions on Communications*, vol. 53, no. 5, pp. 841–852, May 2005.

[93] N. Laoutaris, S. Syntila, and I. Stavrakakis, "Meta algorithms for hierarchical web caches," in *IEEE International Conference on Performance, Computing, and Communications*, Apr. 2003, pp. 445 – 452.

[94] I. Psaras, W. K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," in *ACM ICN*, Aug. 2012, pp. 55–60.

[95] K. Cho *et al.*, "Wave: Popularity-based and collaborative in-network caching for content-oriented networks," in *IEEE INFOCOM (Workshop)*, Mar. 2012.

[96] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon, "Providing public intradomain traffic matrices to the research community," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, pp. 83–86, Jan. 2006.

[97] V. Ramachandran, "Finding a minimum feedback arc set in reducible flow graphs," *Elsevier Journal of Algoriths*, vol. 9, no. 3, pp. 299 – 313, Sept. 1988.

[98] P. Eades, X. Lin, and W. F. Smyth, "A fast and effective heuristic for the feedback arc set problem," *Elsevier Information Processing Letters*, vol. 47, no. 6, pp. 319 – 323, Oct. 1993.

[99] A.-L. Barabási and R. Albert, "Emergence of scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509 – 512, Oct. 1999.

[100] M. Gregori, J. Gómez-Vilardebó, J. Matamoros, and D. Gündüz, "Wireless content caching for small cell and D2D networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1222–1234, May 2016.

[101] J. Li, J. Sun, Y. Qian, F. Shu, M. Xiao, and W. Xiang, "A commercial video-caching system for small-cell cellular networks using game theory," *IEEE Access*, vol. 4, pp. 7519–7531, June 2016.

[102] S. Müller, O. Atan, M. van der Schaar, and A. Klein, "Context-aware proactive content caching with service differentiation in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 1024–1036, Feb. 2017.

[103] J. Kurose, "Information-centric networking: The evolution from circuits to packets to content," *Elsevier Computer Networks*, vol. 66, pp. 112–120, June 2014.

[104] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *ACM CoNEXT*, 2009, pp. 1–12.

[105] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 3rd ed.   Wiley, 2013.

[106] H. Che, Y. Tung, and Z. Wang, "Hierarchical web caching systems: modeling, design and experimental results," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 7, pp. 1305–1314, Sept. 2002.

[107] G. N. Watson, "The harmonic functions associated with the parabolic cylinder," *Proc. London Math. Soc.*, vol. s2-17, no. 1, pp. 116–148, 1918.

[108] S. Al-Ahmadi, "The Gamma-Gamma signal fading model: A survey," *IEEE Antennas and Propagation Magazine*, vol. 56, no. 5, pp. 245–260, Oct. 2014.

[109] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*.   Dover, 1972, no. 10.

[110] R. Barakat, "Weak-scatterer generalization of the K-density function with application to laser scattering in atmospheric turbulence," *Journal of the Optical Society of America A*, vol. 3, no. 4, pp. 401–409, 1986.

[111] T. Rappaport, *Wireless Communications: Principles and Practice*.   Prentice Hall, Dec. 2001.

[112] M. K. Simon and M. S. Alouni, *Digital Communication over Fading Channels*. John Wiley & Sons, 2000.

[113] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 8th ed. Elsevier, Sept. 2014.

[114] P. S. Bithas, P. T. Mathiopoulos, and S. A. Kotsopoulos, "Diversity reception over Generalized-K (k_g) fading channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4238–4243, Dec. 2007.

[115] W. C. Jakes, *Microwave Mobile Communications*. Wiley, 1974.

[116] A. Papoulis, *Probability, Random Variables, and Stochastic Procceses*, 3rd ed. McGraw-Hill, 1991.

[117] E. W. Ng and M. Geller, "A table of integrals of the error functions," *Journal of Research of the National Bureau of Standards*, vol. 73B, Oct. 1969.

Appendix A

Appendices for Chapter 4

A.1 Proof of Theorem 1

We express the P_{hs_j} term in (4.16) as a function of cdf,

$$P_{hs_j} = \frac{F_{\mathcal{L}}(vT_c) - F_{\mathcal{L}}\left(\left(\frac{P_{ts_j}}{P_{th}}\right)^{1/\beta}\right)}{F_{\mathcal{L}}(vT_c)}, \tag{A.1}$$

where $F_{\mathcal{L}}(\cdot)$ stands for the cdf of the random variable for transition length. Replacing $F_{\mathcal{L}}(x)$ with $1 - e^{-\rho \pi x^2}$, cdf of the Rayleigh distribution. Thereafter, we replace P_{succ_j} with (4.14) and (4.15) to obtain Theorem 1.

A.2 Proof of Theorem 2

When $\mathcal{V}$ is a random variable, $P(\mathcal{L} < vT_c)$ cannot be expressed as the cdf of $\mathcal{L}$. Hence, in this case we express P_{hs_j} as,

$$P_{hs_j} = \frac{P(\mathcal{L} < vT_c) - F_{\mathcal{L}}\left(\left(\frac{P_{ts}}{P_{th}}\right)^{1/\beta}\right)}{P(\mathcal{L} < vT_c)} \tag{A.2}$$

Without losing the sense of generality we can assume $\mathcal{L}$ and $\mathcal{V}$ to be independent. Therefore, we can write,

$$P(\mathcal{L} < vT_c) = \frac{1}{T_c} \int_{T_c v_{min}}^{T_c v_{max}} \int_0^v f_{\mathcal{L}}(l) f_{\mathcal{V}}(v) dl dv, \tag{A.3}$$

where $f_{\mathcal{L}}(\cdot)$ and $f_{\mathcal{V}}$ are the pdf corresponding to the random variable for transition length and and velocity respectively. We already know $\mathcal{L}, \mathcal{V}$ follow Rayleigh and Uniform distribution respectively. Therefore, using corresponding $f_{\mathcal{L}}$ and $f_{\mathcal{V}}$ we obtain the resulting expression for (A.3),

$$P(\mathcal{L} < vT_c) = 1 + \frac{Q(\sqrt{2\rho\pi}T_c v_{max}) - Q(\sqrt{2\rho\pi}T_c v_{min})}{T_c(v_{max} - v_{min})} \tag{A.4}$$

Finally, replacing (A.4) in (A.2) we obtain the expression of handover probability between $j-$th tier SBS and user, and thereafter, replacing the term P_{succ_j} with (4.14) and (4.15) we obtain Theorem 2.

Appendix B

Appendices for Chapter 5

B.1 Proof of Theorem 4

Replacing Φ with error function and applying its odd function property in (5.21) gives,

$$P_e(\overline{\gamma}) = \frac{\sqrt{\pi}}{2\Gamma(\lambda)} \int_0^\infty \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{\nu\overline{\gamma}\beta}{2}}\right)(a\beta^t + b\beta^\mu \log(\beta))d\beta \tag{B.1}$$

Replacing $\nu\overline{\gamma}\beta$ with $2\tau^2$ in (B.1) gives,

$$P_e(\overline{\gamma}) = \frac{\sqrt{\pi}}{\Gamma(\lambda)} \int_0^\infty \mathrm{erfc}(\tau)\left(\frac{a\tau^{2t+1}}{(\nu\overline{\gamma}/2)^{2t}} + \frac{b\tau^{2\mu+1}}{(\nu\overline{\gamma}/2)^{2\mu}}\right.$$
$$\left. \times (2\log(\tau) - \log(\nu\overline{\gamma}/2))\right)d\tau \tag{B.2}$$

Now using [[117], 4.1.11,4.6.9] a closed-form solution of the integral in (B.2) is obtained,

$$P_e(\overline{\gamma}) = \frac{a\sqrt{\pi}}{(\nu\overline{\gamma}/2)^{2t}\Gamma(\lambda)} \frac{\Gamma(t+3/2)}{\sqrt{\pi}(2t+2)} + \frac{2b\sqrt{\pi}}{(\nu\overline{\gamma}/2)^{2\mu}\Gamma(\lambda)} \frac{\Gamma(\mu+1/2)}{2\sqrt{\pi}}$$
$$\times \left(\frac{1}{2\mu+2} + \frac{2\mu+1}{2}\psi\left(\frac{2\mu+1}{2}\right)\right) - \frac{b\sqrt{\pi}}{(\nu\overline{\gamma}/2)^{2\mu}\Gamma(\lambda)}$$
$$\times \frac{\Gamma(\mu+3/2)}{\sqrt{\pi}(2\mu+2)} \log(\nu\overline{\gamma}/2) \tag{B.3}$$

After some rearrangement of (B.3) leads to the closed-form solution using given in (5.22) of Theorem 4.

B.2 Proof of Theorem 5

Applying the second property in (5.20) over (5.19) we can write,

$$P_e(\overline{\gamma}) = \frac{1}{2\Gamma(\lambda)} \int_0^\infty \left(a\beta^t + b\beta^\mu \log(\beta)\right)(\lambda - 1)! e^{-\nu\overline{\gamma}\beta} \sum_{m=0}^{\lambda-1} \frac{(\nu\overline{\gamma}\beta)^m}{m!} d\beta$$

$$= \frac{a(\lambda - 1)!}{2\Gamma(\lambda)} \sum_{m=0}^{\lambda-1} \frac{(\nu\overline{\gamma})^m}{m!} \int_0^\infty \beta^{t+m} e^{-\nu\overline{\gamma}\beta} d\beta + \frac{b(\lambda - 1)!}{2\Gamma(\lambda)}$$

$$\times \sum_{m=0}^{\lambda-1} \frac{(\nu\overline{\gamma})^m}{m!} \int_0^\infty \beta^{\mu+m} e^{-\nu\overline{\gamma}\beta} \log(\beta) d\beta \tag{B.4}$$

Using [[113], 3.351.3, 4.352.1], (B.4) can be rewritten as,

$$P_e(\overline{\gamma}) = \frac{a(\lambda - 1)!}{2\Gamma(\lambda)} \sum_{m=0}^{\lambda-1} \frac{(\nu\overline{\gamma})^m}{m!} \Gamma(t + m + 1)(\nu\overline{\gamma})^{-t-m-1} + \frac{b(\lambda - 1)!}{2\Gamma(\lambda)}$$

$$\times \sum_{m=0}^{\lambda-1} \frac{(\nu\overline{\gamma})^m}{m!} \frac{\Gamma(\mu + m + 1)}{(\nu\overline{\gamma})^{\mu+m+1}} \left(\psi(\mu + m + 1) - \log(\nu\overline{\gamma})\right) \tag{B.5}$$

Rearranging (B.5) gives the closed-form solution given in Theorem 5.

B.3 Proof of Theorem 6

Using the SEP for M-ary PSK, given in (5.24), BER can be calculated as following,

$$P_e(\overline{\gamma}) = \frac{a}{\pi} \int_0^\infty \int_0^{\frac{(M-1)\pi}{M}} \exp\left(-\frac{\beta\overline{\gamma}\sin^2(\pi/M)}{\sin^2(\theta)}\right)\beta^t d\theta d\beta$$

$$+ \frac{b}{\pi} \int_0^\infty \int_0^{\frac{(M-1)\pi}{M}} \exp\left(-\frac{\beta\overline{\gamma}\sin^2(\pi/M)}{\sin^2(\theta)}\right) \log(\beta)\beta^\mu d\theta d\beta$$

$$= A_1 + A_2 \tag{B.6}$$

where, A_1 and A_2 are the first and second integral, respectively. We solve the integrals, A_1 and A_2 separately for a reader friendly representation,

$$
\begin{aligned}
A_1 &= \frac{a\Gamma(t+1)}{\pi\overline{\gamma}^{t+1}\sin^{2t+2}(\pi/M)} \int_0^{\frac{(M-1)\pi}{M}} \sin^{2t+2}(\theta)d\theta \\
&= \frac{a\Gamma(t+1)}{\pi\overline{\gamma}^{t+1}\sin^{2t+2}(\pi/M)} \Bigg(\frac{(t+1)(M-1)\pi}{2M} \\
&\quad - \frac{(t+1)\sin(2(M-1)\pi/M)}{4} \Bigg)
\end{aligned}
\tag{B.7}
$$

Similarly,

$$
\begin{aligned}
A_2 &= \frac{b\big(\psi(\mu+1) - 2\Gamma(\mu+1)\log(\sin(\pi/M))\big)}{\pi\overline{\gamma}^{\mu+1}\sin^{2\mu+2}(\pi/M)} \\
&\quad \times \int_0^{\frac{(M-1)\pi}{M}} \sin^{2\mu+2}(\theta)d\theta + \frac{2b\Gamma(\mu+1)}{\pi\overline{\gamma}^{\mu+1}\sin^{2m+2}(\pi/M)} \\
&\quad \times \int_0^{\sin^{-1}(\frac{(M-1)\pi}{M})} \frac{\log(x)}{\sqrt{1-x^2}}x^{2\mu+2}dx
\end{aligned}
\tag{B.8}
$$

As μ is positive, final expression of A_2 is given by (B.9), where $_pF_q$ is the hypergeometric function and ψ is the digamma function.

$$
\begin{aligned}
A_2 &= \frac{a_2\big(\psi(\mu+1) - 2\Gamma(m+1)\log(\sin(\pi/M))\big)}{\pi\overline{\gamma}^{\mu+1}\sin^{2\mu+2}(\pi/M)} \Bigg(\frac{(\mu+1)(M-1)\pi}{2M} \\
&\quad - \frac{(\mu+1)\sin(2(M-1)\pi/M)}{4} \Bigg) + \frac{2\alpha_2\Gamma(\mu+1)}{\pi\overline{\gamma}^{\mu+1}\sin^{2\mu+2}(\pi/M)} \Bigg[\frac{x^{1+2\mu}}{(1+2k)^2} \\
&\quad \times \left({}_3F_2\left[-\frac{1}{2}, \frac{1}{2}+\mu, \frac{1}{2}+\mu, \frac{3}{2}+\mu, \frac{3}{2}+\mu, x^2\right] - (1+2\mu)\log(x) \right. \\
&\quad \times \left({}_2F_1\left[-\frac{1}{2}, \frac{1}{2}+\mu, \frac{3}{2}+\mu, x^2\right] - {}_2F_1\left[\frac{1}{2}, \frac{1}{2}+\mu, \frac{3}{2}+\mu, x^2\right] \right) \\
&\quad - {}_3F_2\left[\frac{1}{2}, \frac{1}{2}+\mu, \frac{1}{2}+\mu, \frac{3}{2}+\mu, \frac{3}{2}+\mu, x^2\right] \Bigg) \Bigg]_0^{\sin^{-1}(\frac{(M-1)\pi}{M})}
\end{aligned}
\tag{B.9}
$$

Adding (B.7) and (B.9) we obtain the expression given by (5.25) in Theorem 6.