Titre: Title:	Probabilistic Analysis of Operating Modes in Cache-Enabled Full-Duplex Device-to-Device Communications
Auteur: Author:	Mansour Naslcheraghi
Date:	2022
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Naslcheraghi, M. (2022). Probabilistic Analysis of Operating Modes in Cache-Enabled Full-Duplex Device-to-Device Communications [Master's thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/10276/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/10276/
Directeurs de recherche: Advisors:	Brunilde Sanso, & Jean-François Frigon
Programme: Program:	Génie électrique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Probabilistic analysis of operating modes in cache-enabled full-duplex device-to-device communications

MANSOUR NASLCHERAGHI
Département de génie électrique

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie électrique

Avril 2022

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Probabilistic analysis of operating modes in cache-enabled full-duplex device-to-device communications

présenté par **Mansour NASLCHERAGHI**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Chahé NERGUIZIAN, président
Brunilde SANSO, membre et directrice de recherche
Jean-François FRIGON, membre et codirecteur de recherche
Gunes KARABULUT KURT, membre

DEDICATION

To my parents, Salim and Shafigheh

ACKNOWLEDGEMENTS

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my masters degree. My deepest thanks go to my supervisor Prof. Brunilde Sansò and co-supervisor Prof. Jean-François Frigon, whose insights and guidance into the subject matter steered me through this research.

Special thanks to Ericsson Montreal for supporting this project from the beginning.

And thanks to my family for all the support they have shown me through the tough time, the culmination of three years of distance.

Finally, many thanks to Brigitte for her help and guidance in making final revisions.

RÉSUMÉ

Les communications Device-to-Device (D2D) activées par le cache, reconnues comme l'un des principaux catalyseurs du réseau cellulaire de cinquième génération (5G), sont une solution prometteuse pour alléger la lourde charge pesant sur les réseaux centraux mobiles et les liaisons terrestres. La mise en cache, cependant, impose un nouvel indice de performance clé (KPI) à l'utilisateur réseau qui est la probabilité de satisfaction de l'utilisateur. En d'autres termes, quelle est la probabilité que l'utilisateur obtienne les informations nécessaires du réseau? Cette probabilité dépend du type de transmission (c'est-à-dire semi-duplex ou duplex intégral) et de nombreux éléments liés au système de mise en cache, tels que la manière dont les informations sont mises en cache ou la popularité des informations mises en cache. L'analyse de ces éléments donne lieu à différents modes de fonctionnement. Afin d'évaluer le nouveau KPI, les probabilités de chaque mode de fonctionnement doivent être extraites des conditions de transmission et de mise en cache. Cette thèse présente une analyse approfondie de ces probabilités mettant en pertinence la relation entre les politiques de mise en cache, la popularité des contenus et les types de transmission. De telles relations permettent une évaluation fluide de la satisfaction des utilisateurs dans différentes conditions.

ABSTRACT

Cache-enabled D2D (Device-to-Device) communications, recognized as one of the key enablers of the fifth generation (5G) cellular network, is a promising solution to alleviate the great burden on mobile core networks and backhaul links. Caching, however, imposes a new networking user Key Performance Index (KPI) that is the probability of user satisfaction. In other words, how likely is the user to get the needed information from the network? Such probability depends on the type of transmission, (i.e. half duplex or full duplex) and on many elements related to the caching system, such as the way the information is cached, or the popularity of the cached information. The analysis of those elements gives rise to different modes of operation. In order to evaluate the new KPI, the probabilities of each mode of operation must be extracted from the transmission and caching conditions. This thesis presents a thorough analysis of those probabilities putting in relevance the relationship between caching policies, content popularity and transmission types. Such relationships allow the smooth evaluation of user satisfaction under different conditions.

TABLE OF CONTENTS

DEDICATION iii

ACKNOWLEDGEMENTS iv

RÉSUMÉ v

ABSTRACT vi

TABLE OF CONTENTS vii

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF SYMBOLS AND ACRONYMS xi

CHAPTER 1 INTRODUCTION 1
- 1.1 Motivations 5
- 1.2 Objectives and Contributions 6
- 1.3 Thesis Organization 8

CHAPTER 2 LITERATURE REVIEW 9
- 2.1 Caching Strategies 9
  - 2.1.1 Deterministic Caching Policies 10
  - 2.1.2 Stochastic Caching Policy 12
- 2.2 Transmission Types and Operating Modes 13
  - 2.2.1 Half-Duplex 13
  - 2.2.2 Full-Duplex 17
- 2.3 Objective of the Thesis in Comparison with the Existing Works 21

CHAPTER 3 SYSTEM MODEL 25
- 3.1 System Overall Structure 25
- 3.2 Caching Model 25
  - 3.2.1 Deterministic Caching 27
  - 3.2.2 Stochastic Caching 27
- 3.3 Modeling User Operating Modes 28

CHAPTER 4 ANALYZING NODE OPERATING PROBABILITIES 30
4.1 Deterministic Caching 30
4.2 Stochastic Caching 32
4.3 Probability Mass Functions (PMFs) and Satisfied UEs 38
4.3.1 PMFs 38
4.3.2 Expected Number of Satisfied Users 39

CHAPTER 5 NUMERICAL RESULTS 40
5.1 Analysis of mode probability 40
5.1.1 By type of mode 40
5.1.2 By type of caching features 42
5.1.3 PMF analysis 44
5.1.4 Expected number of satisfied users 44

CHAPTER 6 CONCLUSION AND RECOMMENDATIONS 47
6.1 Some Insights on the Proposed Analysis 47
6.2 Potential Research Directions 49
6.2.1 Physical Layer Investigations 49
6.2.2 Scheduling Problem 49
6.2.3 Satisfying Cache Hit Outage Mode 50
6.2.4 Security Aspects 50

REFERENCES 51

LIST OF TABLES

Table 2.1 Objective of the thesis in comparison with the existing FD-D2D caching works 24
Table 3.1 Summary of notation being used throughout the thesis 26
Table 4.1 Table of Caching Permutations for N=3 and m=4 34
Table 5.1 Simulation Parameters in Fig. 5.4 42

LIST OF FIGURES

Figure 1.1 Edge caching: Bringing contents closer to end users by putting required contents in the network entities closer to the end-users 2
Figure 1.2 Demonstration of HD and FD operations. (a) HD unicast link, (b) HD multicast link, (c) bidirectional FD links, (d) three node FD operation, and (e) FD multicast operation 4
Figure 1.3 HD and FD connections for two users 4
Figure 1.4 FD-Enabled D2D caching network 5
Figure 2.1 Overall view of the existing literature in cache-enabled D2D networks 10
Figure 2.2 Most popular contents cached in subset of users 11
Figure 2.3 Contents are being cached in users randomly 12
Figure 2.4 Half-Duplex transmission mode: (a): single link transmission and (b): multicast transmission 14
Figure 2.5 Full-Duplex Transmission Mode: (a): Bidirectional FD operation, (b): Half-Duplex operation, and (c): Three Node FD operation 18
22figure.2.6
Figure 2.7 All possible outcome from the observations for a small network consisting 7 nodes 22
Figure 3.1 Demonstration of the centrally-controlled D2D caching network . . . 25
Figure 3.2 All possible operating modes for an arbitrary node 29
Figure 5.1 Deterministic caching with $M = 500$, $\gamma_r = 0.8$ 41
Figure 5.2 Limited results of the stochastic caching with $M = 7$, $\gamma_r = 0.8$, $\gamma_c = 1.6$ 41
Figure 5.3 Stochastic caching with $M = 500$, $\gamma_r = 0.8$, $\gamma_c = 1.6$ 42
Figure 5.4 Comparison of probabilities versus γ_r , γ_c, and M 43
Figure 5.5 Comparison of PMFs with respect to γ_c and M 45
Figure 5.6 Expected number of satisfied users for both deterministic and stochastic caching policies 46

LIST OF SYMBOLS AND ACRONYMS

6G	Sixth Generation
5G	Fifth Generation
LTE	Long Term Evolution
D2D	Device-to-Device
HD	Half-Duplex
FD	Full-Duplex
SI	Self-Interference
SIC	Self-Interference Cancellation
SBS	Small Base Station
BS	Macro Base Station
HetNet	Heterogeneous Networks
MEC	Mobile Edge Caching
3GPP	Third Generation Partnership Project
QoS	Quality of Service
QoE	Quality of Experience
V2X	Vehicle-to-Everything
C-V2X	Cellular Vehicle-to-Everything
TX	Transmitter
RX	Receiver
SR	Self-Request
SR-HDTX	Self-Request and Half-Duplex Transmitter
FDTR	Full-Duplex Transceiver
HO	Hitting Outage
HDRX	Half-Duplex Receiver
HDTX	Half-Duplex Transmitter
SG	Stochastic Geometry
BS	Base Station
FD-D2D	Full-Duplex Device-to-Device
KPI	Key Performance Index
PMF	Probability Mass Function
UAV	Unmanned Aerial Vehicle
NOMA	None Orthogonal Multiple Access
i.i.d	independent and identically distributed

CHAPTER 1 INTRODUCTION

Nowadays, users are more interested to get their desired information through multimedia contents such as video, voice, and podcasts rather than relying on traditional textual information. This type of data has a large size and it requires a high data rate and more capacity. For instance, in a dense hotspot network such as public transport stations, stadiums, and social events, users are more interested to get firsthand knowledge about the contents by relying on multimedia files. This compels a huge burden in the cellular network infrastructure and this is rapidly increasing year by year. It is predicted by Cisco that almost 20% of the contents account for 80% of the total data traffic, especially multimedia contents [1].

With the advent of modern cellular networks, namely fifth generation (5G) evolution and beyond, the challenge of delivering highly demanding multimedia contents with high data rate to the end users became one of the key aspects in planning modern cellular network infrastructure. To rapidly alleviate this increasing burden on the cellular networks, a variety of technologies are being introduced by the third generation partnership project (3GPP) standardization community to address these demands and currently some of these technologies gained required prototypes.

Among the technologies proposed to address the aforementioned challenges, the mobile edge-caching (MEC) technology has attracted widespread attention because of its potential to improve system performance and enhance user experience [2]. Technically, the key idea behind this technology is to bring the contents closer to the end users by utilizing caching capability of the edge devices. Thanks to this technology, the end users are able to fetch their desired contents from the edge devices rather than retrieving it from the core network, which correspondingly can alleviate the huge burden compelled on the backhaul links. However, the term “edge devices” is not limited to only small base stations (e.g. femtocells and picocells) with caching capability, in fact, the end user devices (mobile devices) by themselves can participate in distributed content caching and delivering mechanisms. Thanks to the rapidly growing advancements in designing mobile devices, nowadays many of the mobile device manufacturers provide sufficient local storage capability to users, which paves the road to enable these user devices to store some highly demanding contents and deliver them asynchronously to nearby devices through device-to-device (D2D) communications, which is one of the key enablers in 5G, beyond 5G, and sixth generation (6G) networks.

The idea of distributed caching mechanism through D2D communications is proposed by [3, 4] in 2012. Since then, this paradigm opened a variety of research directions and the follow-up research put this idea in different applications and use cases from offloading mobile traffic to safety-related content delivery over connected vehicles [5–10]. It is shown that this technology can offload traffic using direct links [5], reduce retrieve time [6, 7], improve bandwidth efficiency and enhance data service performance [8], and improve overall service quality and delivery rate by proactively placing highly demanding data in users' storage [9,10]. In summary, this technology brings the following advantages:

- Brings contents closer to user terminals
- Avoids traffic congestion
- Increases Quality of Service (QoS)
- Increases resource reuse
- Easy to deploy: it is shown by the pioneers [3] that this technology does not require major modifications on the existing infrastructure and it can be deployed on the current cellular network with minor modifications, thanks to advancements in smart devices operating in current cellular networks.

So, the idea of caching as a system performance essential tool has been clearly stated in the literature.

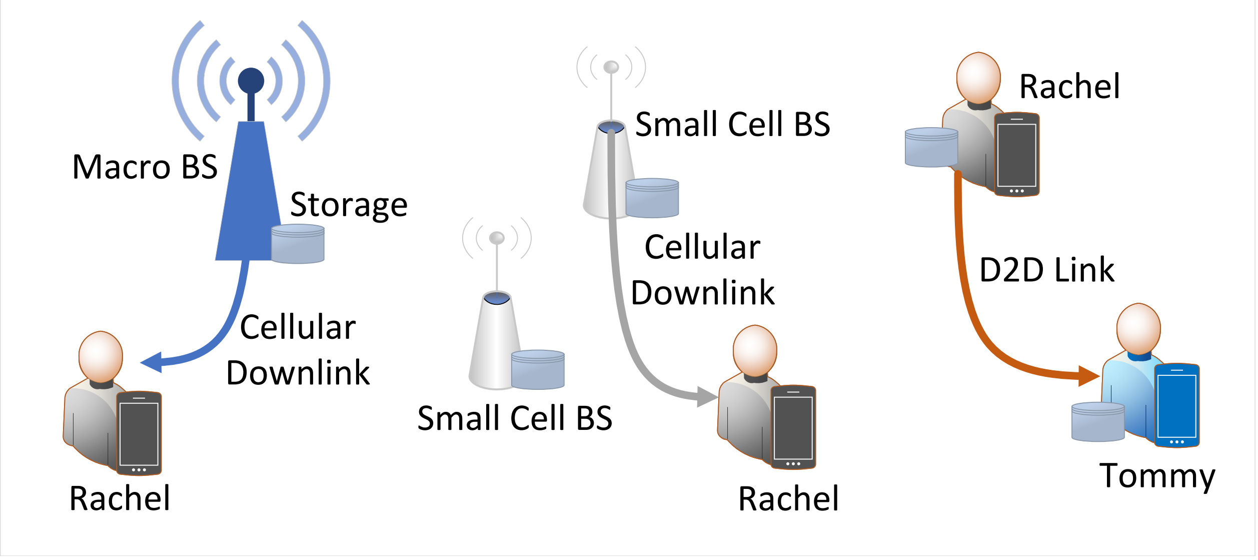

Figure 1.1 Edge caching: Bringing contents closer to end users by putting required contents in the network entities closer to the end-users

On the other hand, full-duplex (FD) communication technology is getting interest from the mobile operators and it is proposed by 3GPP as one of the key enablers for 5G because of its capability to more efficient use of the spectrum. In fact, when operators aim to increase the efficiency of spectrum usage, any technology that offers this efficiency is worth a closer inspection. This technology promises to enable wireless links to simultaneously transmit and receive over a single spectrum channel at the same time. However, FD technology comes with some challenges, mainly the potential for interference between the reception and transmission signals. Companies such as Lextrum, Kumu Networks, and GenXComm that make full duplex technology say that the reason it has not caught on before now is because operators did not have the same urgency when it comes to conserving their spectrum assets as they do with 5G. Plus, full duplex has finally matured to the point where there is less concern about interference, thanks to emerging more efficient techniques in mitigating self-interference (SI) [11, 12]. This technology is broadly investigated in cellular communications [13–15] and C-V2X networks [16–21].

Let us first define the HD and FD operations in a cache-enabled D2D network. Fig. 1.2 delineates the HD and FD operations with different possibilities. In the HD operation, one node can either transmit data to a single receiver (e.g (a)) or transmit data to multiple receivers via a single multicast transmission (e.g. (b)). However, when nodes are FD-capable, one node can simultaneously transmit to some nearby nodes and simultaneously, can receive data from another node. This is possible through three different possibilities: two nodes can exchange data simultaneously through bidirectional FD link (e.g. (c)), one node can receive data from a nearby transmitter and simultaneously transmit data to another node (e.g. (d)), and it is possible that the FD node can transmit data to multiple receivers via a multicast transmission (e.g., (e)). In this thesis we consider the logical link capabilities of the nodes (i.e. HD or FD capable) and neglect the physical layer impairments associated with FD transmissions.

Now, consider two users (see Fig. 1.3), namely “Rachel” and “Tommy” (see Fig. 1.3), and assuming that they both are interested to retrieve contents cached in their user devices, with a conventional HD communication, either Rachel or Tommy would be able to fetch the desired content at a given time. However, by employing FD capability in this small network, both users can be satisfied. There are many advantages in using FD capability in a cache-enabled D2D network and recent studies show the potential benefits when FD technology is integrated to the cache-enabled D2D network.

As can be seen, in a simple network shown in Fig. 1.3, FD capability can potentially provide more opportunity for content deliveries and users satisfaction, in a given time slot. However, in a dense network, the outcomes is not as simple as shown in Fig. 1.3 and we might see a complicated random graph depending on the users’ demands and cached contents in the

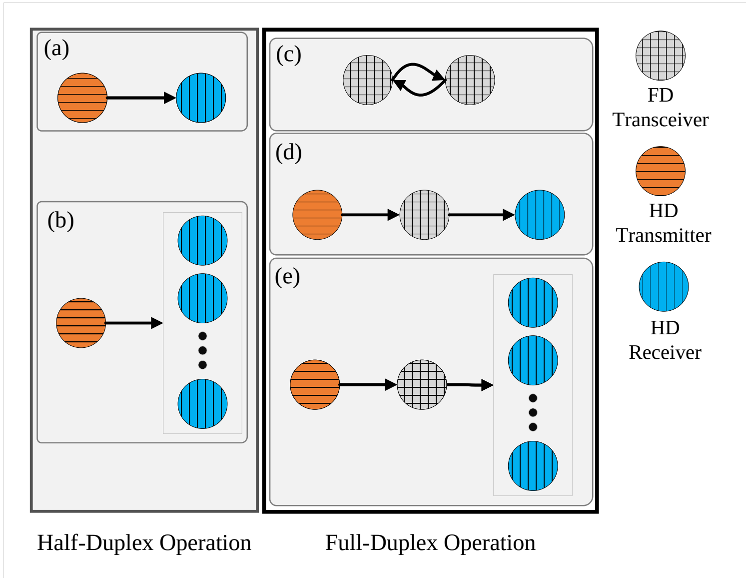

Figure 1.2 Demonstration of HD and FD operations. (a) HD unicast link, (b) HD multicast link, (c) bidirectional FD links, (d) three node FD operation, and (e) FD multicast operation

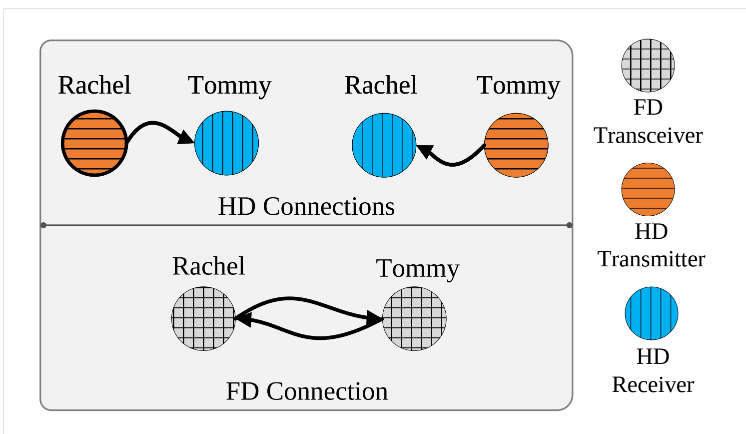

Figure 1.3 HD and FD connections for two users

storage of user devices. The accurate analysis of the expected number of satisfied users becomes more challenging as it depends on different factors such as the number of users in network and the policy that is being used to store contents in user devices. This forms the basis of the motivations and objectives in this thesis.

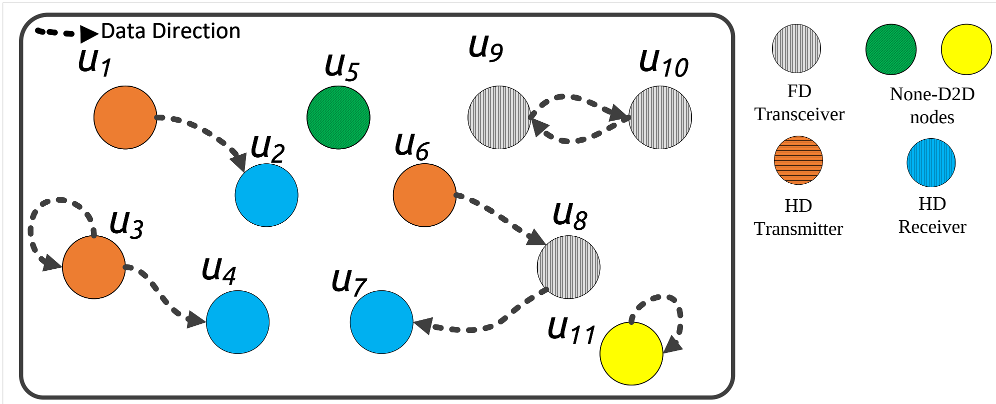

Figure 1.4 FD-Enabled D2D caching network

1.1 Motivations

Technically, the random graph demonstrated in Fig. 1.4 delineates the number of satisfied users that appear randomly in a FD-enabled D2D caching network. More precisely, the caching mechanism dictates the possible transmitting/receiving users which directly impacts the number of users in each mode. For caching to be successful, the user must be able to find the caching content. This depends on several factors such as:

- The caching policy
- The content popularity and
- The type of transmission that is being appeared randomly between devices. This randomness comes from the fact that user devices cache and request different contents.

The caching policy refers to how the content distributed over the caching nodes. For instance, is the distribution of contents in user devices done through a deterministic manner or it is done in a random manner? Also, how the popularity of contents plays a role in content distributions? It is worth to note that the popularity of content refers to the probability of the event of users requesting the same content. Also, it is possible that some users fail to retrieve their desired demand because of the unavailability of the desired content in the nearby users' storage. Therefore, there might be a blocking of the request, depending on how the content caching was distributed. Finally, the type of transmission plays an important role because it induces a graph of directed links between devices and, depending on the direction, directed links may create an extra obstacle for the user to fetch the right content. In terms of modeling, FD transmission represents the best type of transmission possible between two device entities because it creates more content delivery opportunities for all instants of time.

Even though the analysis presented in this thesis would still be valid for classical duplex transmissions, we will keep the FD denomination to follow the trend in the literature that has put in relevance such transmission with caching policies [22].

In what follows, we use the term *node* to denote indistinctly a user or her communicating device. In the existing literature, e.g., see [22–32], the focus has been on the performance analysis of an arbitrary node when it obtains content from a nearby node either by assuming that the requested content is always available nearby or by employing an optimistic caching policy, also referred as the deterministic caching policy [3]. In [22–24], a cache-enabled and cluster-based FD-D2D network is analyzed to investigate network throughput and download time and for the worse case scenario, it is assumed that all users inside the cluster are simultaneously making random requests to retrieve their desired contents. Stochastic Geometry (SG) based analysis is being conducted in [25–28], in which the distribution of transmitters, relays, and receivers are determined with the caching placement probabilities inspired by the *Thinning Theory* [33], and the receiver of interest is associated to the closest transmitter in the vicinity. FD-enabled transmitters in [27, 28] operate as relays to relay contents from the central base station (BS), when the desired content of the receiver of interest is not stored in the nearby transmitter. In the other approach [29–31], it is assumed that the portion of transmitters operate in FD mode, while the rest of users and backhaul wireless access point operate in HD mode. Therefore, the existing work focus on a very specific case out of all possible scenarios that could occur in a cache-enabled D2D networks in which the nodes are empowered with FD capability. Specifically, the utilization of a more realistic caching policy is missing in the context of FD-enabled D2D caching networks.

Differently from the above mentioned literature, from the Quality of Experience (QoE) perspective, we concentrate our analysis on how good is the system from the user point of view. In other words, how can we find the *the user's probability of getting the desired content*. To be able extract that KPI it is necessary to model all possible modes of operation on different caching policies and infer how probable each mode will be in each case. More details about the main contributions are provided next.

1.2 Objectives and Contributions

The main contribution of this thesis is a thorough probabilistic analysis that allows the evaluation of a new KPI that provides the user's quality of experience (QoE) in a cache-enabled FD D2D network. That measure is the probability that a user finds the requested content. This QoE measure can then be converted into another measure that may be useful for network operators: the expected number of satisfied users, that is, the expected number of user that have found their wanted content.

The evaluation of these measures needs the exploration of the probability of all operating modes in the network, which has not been done so far in the literature. More detailed original contributions and outcomes are listed below:

- We study stochastic caching policies for the content placement strategy, according to which every node caches contents randomly from a large set of contents tracked and determined by the central network. Differently from [22–32], in which deterministic caching is being used, the stochastic caching policy causes overlapping between the users' caches, bringing the opportunity to deliver the content of interest to the receiver of interest through different nearby transmitters. We then compare the caching performance of random caching with deterministic caching, in terms of the potential transmitters/receivers and outage users.

- We model and analyze all possible operating modes when nodes have FD capability. Differently from [32], in which operating modes are modeled based on the deterministic caching policy, here we conduct the analysis for all operating modes in a stochastic caching policy and compare with deterministic caching.

- Using the closed-form expressions of the probabilities for all operating modes, we derive the Probability Mass Functions (PMFs) of the number of nodes that randomly operate in different modes. Different from [29–31], in which the density of transmitters, receivers, and relay nodes are determined by the cache placement probabilities thanks to *Thinning Theory*, here we obtain PMFs not only for HD transmitters/receivers and FD transceivers, but also PMFs of the outage users, those that fail to retrieve contents from their nearby nodes. Moreover, differently from [32], in which the PMF is obtained only for the transmitters that actively transmit at any given time by considering deterministic caching policy, here we obtain PMFs for all modes by considering stochastic caching policy and compare the PMFs of both caching policies in terms of system key parameters.

1.3 Thesis Organization

In Chapter 2, state of the art is investigated. System model, caching mechanisms, and operating modes are presented in chapter 3. In Chapter 4, modeling and analysis of the statistics and probabilities for all possible operating modes along with the expected number of satisfied users is presented. Theoretical and simulation results are provided in Chapter 5 and finally Chapter 6 presents the closing remarks and some further insights on the proposed analytical framework. In the rest of thesis, bold capital account for matrices and none-bold capital letters account for sets.

CHAPTER 2 LITERATURE REVIEW

Recent studies in cache-enabled D2D networks took a variety of directions from investigating theoretical bounds and fundamental limits [34] to different caching and delivery strategies [35–54], [22, 23, 25, 32, 55–73]. Fig. 2.1 delineates the overall view of the existing works in cache-enabled D2D networks. In particular, from the caching strategies point of view, recent studies took two major directions, namely:

- Deterministic caching policies [22–32, 41–43, 51–54]
- Stochastic caching policies [44–50]

On the other hand, from the delivery strategies and data transmission point of view, recent studies utilized two major approaches, namely:

- Half-Duplex transmission [55–67]
- Full-Duplex transmission: [22–32, 68–73]

As can be seen from the above categories, there is an overlap between two point of views, namely works categorized with caching strategies and the works categorized with transmission types. More precisely, except the works [22–32], all works in the caching strategies point of view utilized half-duplex transmission types. In the sequel, all these categories along with key differences between them are being explained.

2.1 Caching Strategies

In a nutshell, a caching strategy means the way that user devices store some contents in their local storage and then transmit those contents to nearby users, when needed. Given a variety of applications and scenarios, the caching strategy is the key procedure to secure the availability of contents for the user devices as part of the major goal of the cache-enabled D2D network, which is maximizing the chance of a successful content delivery to users. From the network infrastructure point of view, the more chance to retrieve desired contents through direct D2D links, the more benefit in terms of spectrum reuse, traffic offloading, and congestion control. In what follows, different strategies for caching mechanism are being explored.

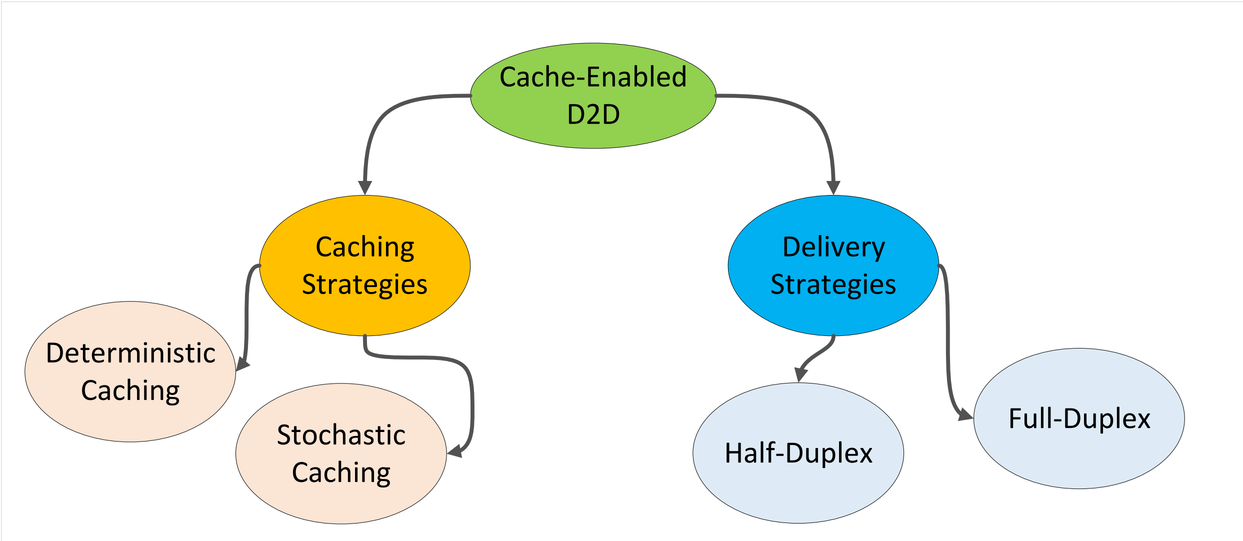

Figure 2.1 Overall view of the existing literature in cache-enabled D2D networks

2.1.1 Deterministic Caching Policies

In deterministic caching policy, the central network determines which contents should be cached in which users devices. Depending on the network scenario, in this policy, the contents are already been tracked and stored by the central network in some libraries and they can be pushed to the local storage of the user devices, when this policy is being established in the network. Technically, there are two major approaches in this policy:

i) **Most Popular Set**: This approach is based on the most popular contents among users which are tracked and stored by the central network in prior times and this procedure typically takes place within the off-peak hours of the network operations [22–32, 42].

ii) **Scenario-Based**: This approach is dictated by some circumstances of the network scenario or application and the cached contents are determined based on these circumstances [41, 43, 51–54].

From the network design point of view, the key difference between the above approaches comes from the caching distribution of the contents, that is, the first case comes with somehow known characteristics or known distributions to store contents, while the latter typically comes with unknown content caching distributions. Here are more details about the above cases:

Most Popular Set

The history of demanded contents by users is being tracked and stored by the central network. In particular, the only feature that determines this set is the popularity of this content also called "cache hit frequency" of the contents while is stored in a library called "universal library" (e.g. see Fig. 2.2). Contents are sorted based on their popularity in a descending way and the most popular content is labeled by the first index in the library. This set can be updated regularly in specific time intervals and, depending on the dynamic characteristics of the target network, this interval can be chosen from somewhere between one hour to days. The advantage of this caching policy is that the central network do not need to make any complex computations to satisfy target network demands and the only logged info is the cache hit frequency of the contents. This approach is being employed for the first time in [3] and many follow-up works (e.g. [22–32, 42]) put the same idea in different performance analysis approaches.

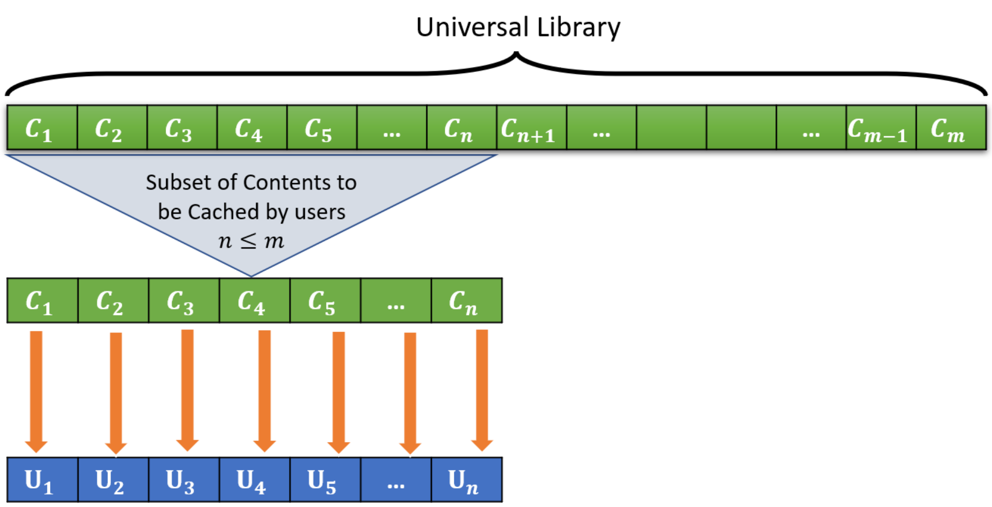

Figure 2.2 Most popular contents cached in subset of users

Scenario-Based:

In contrast to the previous case, this type of determining the cached contents requires computations and feedback from the target network due to specific goal defined by a specific network scenario. For instance, a social-aware D2D caching scenario proposed in [54], which aims to cache contents based on the social relationship between user devices. In this approach, the relationship between user devices is being treated as a graph and pushing contents on user devices is being inspired by this relationship, such that the overall outcome of the network leads to the highest possible cache hit. Another scenario is utilized in [43] in which the users are potentially forming D2D pairs based on the distance threshold and the central network pushes contents in the user devices such that the content is always available in the closest user device. A cluster-centric caching approach is being utilized in [41], in which the contents are always pushed in the users that belong to one cluster.

Further approach is to maximize the energy efficiency of the network and the contents should be determined and pushed to the local storage of the user devices in a way that the overall outcome of the network leads to highest possible energy efficiency. Here, the key performance metric is the energy efficiency and the whole procedure is designed to satisfy this specific goal [52].

Given unknown distributions in scenario-based caching policy and challenges in predicting such distribution, some authors came across the idea of utilizing an approach in which the availability of the content in a user device is modeled by a constant probability value [47–50]. Indeed, constant probability does not represent the unknown distribution, nevertheless, this approach made significant simplifications in analyzing some complex networks.

2.1.2 Stochastic Caching Policy

As can be inferred from the name of this policy, the procedure is stochastic and users participate in content caching procedure independently. Technically, the central network has no influence in the content caching procedure and it is always assumed that users are sufficiently incentivized to participate in such a procedure. One simple assumption is that users cache what they watch, listen, or read, assuming that the size of contents is small or the local storage constraint is not a concern. This strategy has been employed in [3] for the first time and the follow-up works [44–46] utilized this strategy in different scenarios.

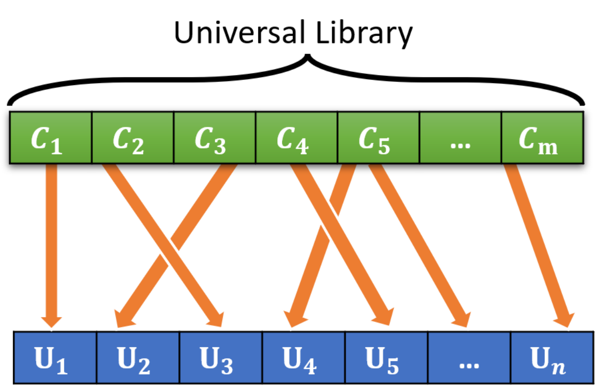

Figure 2.3 Contents are being cached in users randomly

2.2 Transmission Types and Operating Modes

Transmission types refer to the delivery procedure, in which one of the following options can be utilized. i) Half-Duplex and ii) Full-Duplex transmission. In the sequel, the details of these two major transmission types are being explored. It is worth to note that even though the details of the following works is based on their transmission type, nevertheless, the key assumptions on different caching polices with respect to the transmission type is also being investigated.

2.2.1 Half-Duplex

Users can form D2D links subject to availability of the contents in nearby users. In a nutshell, one user can receive its desired demand, when the content of interest is being cached by a nearby user. However, availability of contents depends on the cached contents and correspondingly the caching procedure that is being done in the earlier stage. It is also possible that more than one user demand the content of interest. In that case, one user can target multiple receivers via multicast D2D communications. First, let us delineate what does it mean to have half-duplex operation in a cache-enabled D2D network. Fig. 2.4 delineates two possible options for HD transmission: (a) applies to the case where only one user is demanding the content which is cached in the nearby user, therefore, we have one potential receiver. And option (b) applies to the case where multiple users are demanding the same content, hence, we have a set of potential receivers and one single transmitter can target this group through one multicast transmission. It is worth to recall to that the occurrence of both options depends on two factors: first, what is cached in user devices in prior, and second, what is being demanded by the users. It is obvious that the number of potentially satisfied users in both options is different, correspondingly, the amount of traffic to be offloaded in both cases is different. That is, multicast transmission can offload more data as it targets multiple users with the same demand.

Recent studies [55–67] utilized different analytical tools to model the performance of the cache-enabled network when the communication type is half-duplex. In what follows, details of the modeling along with the target performance metrics for different works are being explored.

Authors in [55] utilized Bipartite Graph Theory, and Stochastic Geometry to model total offloading probability of the cache-enabled D2D system. In particular, the condition to form a D2D connection is satisfied when the distance between the users is less than some predefined threshold. To determine the density of the D2D users, the *Thinning Theory* is

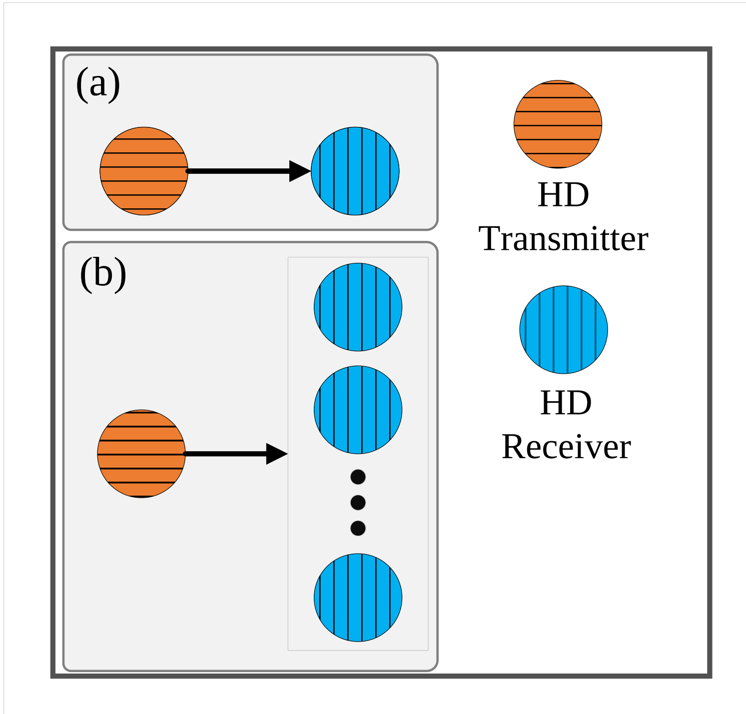

Figure 2.4 Half-Duplex transmission mode: (a): single link transmission and (b): multicast transmission

being considered, in which the density of the users depends on the probability of the content cached in the user of interest. Also, the self-offloading probability is being considered in which the users first check their own cache to see if the desired content can be fetched from their own cache. In some works self-offloading possibility is also called “Self-Request” [3, 22]. The content placement policy is determined by solving an optimization problem that maximizes the total offloading probability, namely, it aims to push contents in user devices that are located close to each other. In this work, such an optimal caching policy leads to highly complex scheduling problem since the system should establish a D2D link that i) is located closer to the receiver and ii) it creates less interference to other receivers as the whole D2D links are sharing the same resource blocks. Due to this complexity, Bipartite Graph Theory is being used to determine the potential D2D pairs and then schedule the links in an iterative way to satisfy the aforementioned conditions (i.e., less interference and shortest distance). The numerical results in terms of the offloading probability and sum rate are being drawn against the case, in which uniform popularity distribution is employed. That is, users follows uniform distribution to store contents rather than the scenario-based caching policy proposed by this work. The idea of exploiting the maximization of offloading probability as the key performance indicator is also investigated in [58] in which the goal is to maximize the user content delivery opportunities through the half-duplex D2D links and the technical aspects of this work is somehow similar to the one as in [55].

D2D communication and BS broadcasting scheme is proposed in [56], in which the goal is to integrate the time-variant characteristic of the popularity of the contents in cache-enabled D2D network. Technically, the final goal is to compare the performance of the system against the case where static caching policy is employed in the network, and that is why the dynamic nature of the network in terms of the mobility and popularity changing is the main concept in this modeling. The central network (here the BS) aims to determine which contents in users' local storage should be updated. Here Self-Request capability is also considered alongside the D2D opportunities. When the desired content is not available via users' own cache, the requests are being delivered to the central BS to see if the desired content can be delivered via D2D communication or via BS broadcasting. Similar to [55], the chance of availability of a content in user devices is modeled by some probabilities and the goal is to determine these probabilities such that the overall number of satisfied users is maximized. The problem also looks into determining the cluster size in terms of the distance that users belong to it, correspondingly, this cluster size can be interpreted as the radio range for the D2D communication since only one single D2D link is allowed per cluster. A similar idea of integrating the cluster size in the problem is also investigated in [3, 57].

Further extension to [56], in which one single D2D link is allowed per cluster, authors in [57] integrate the D2D link data rate requirement in the caching design and the goal is to determine the content placement strategy such that the data rate is being satisfied in D2D links. Since all clusters are sharing the same resource blocks, optimal D2D communication radio range determination is implicitly at the heart of the designing approach, because the interference, accordingly the link capacity, can be managed through controlling the distance from transmitters in other clusters to a typical receiver at one cluster. The authors in [59] modeled a cluster-based D2D content delivery system in which the goal is to maximize the number of D2D links and at the same time, restricting the possible interference imposed from the other clusters. The main difference between this work and other cluster-based modelings as in [56, 56, 57] is that the receiver of interest can potentially receive its desired content from multiple transmitters and there are more than one potential candidate to satisfy the typical user's demand. However, given the constraints in terms of interference threshold, an appropriate is the one that leads to less interference to other D2D links. To obtain the optimal D2D pairs within the network, the problem takes into account the time slotted transmission intervals along with caching size by utilizing Markov chain system.

Another effort was made by [60] to answer the question of how much latency can be reduced by using direct D2D links in a cache-enabled Fog network? In particular, multicast D2D communication is being considered for possible content delivery. The procedure of discovering the potential transmitter to transmit data is being done through broadcast signals made by all users in orthogonal broadcast channels. It is shown that when there is the possibility of achieving high data rate by D2D links, the architecture guarantees the significant reduction in the overall content delivery procedure across the entire network. Hence, interference alignment plays a crucial role in dense D2D network to reduce the amount of the imposed interference on a typical receiver, the authors proposed two different delivery strategies: i) to benefit from the assistance of the fronhaul transmissions to the edge nodes in a optimally designed time interval, and ii) to establish the D2D links for the remaining packet deliveries.

An approach of D2D pair selection is being employed in [62] by considering probabilistic content placement. Similar to [55], the availability of contents in a potential D2D transmitters is modeled by some probability parameters. These probabilities are further being used to make a *thinning* from the dense D2D network. The goal is to find an optimal cache placement strategy that ensures the maximum achievable rate in the system. Distance threshold and closest neighbor are the key factors to determine which contents should be cached in which users. And similar to previously mentioned works, none-D2D self-request is also considered.

Another graph-based approach is proposed in [66] to maximize the number of satisfied users when there is more than one candidate to satisfy a demand made by a typical user. The concept in this work is similar to the one proposed in [55] yet with the difference that the final outcome is focusing on obtaining a graph representation that leads to maximized number of requesting users rather than choosing the transmitters (by pushing contents on them) based on some predefined system thresholds as being considered in [55].

Further multi-source distance-based D2D content delivery system is being proposed by [67] in which the problem of interest is being explored from the energy cost perspective. In this approach, fraction of contents are being cached by user devices rather than caching the whole contents. Therefore, a desired content by a typical user can be satisfied by transmitting different fractions stored locally at different potential transmitters. When multiple transmitters target a single typical receiver, the fractions of the contents are bounded together in the receiver to make a whole content available to the typical user. This approach is also called many-to-one in similar works.

Further multicast graph-based D2D content delivery is proposed in [61], in which the goal is to minimize the potential number of cached contents by user devices while satisfying the desired demand. Since the caching phase by the users itself imposes load on the cellular network, this work targets a minimum amount of caching to reduce the caching load while satisfying the user demands via multicast transmissions. To model the potential D2D connections, a random graph is built and is being updated in different time stamps due to the dynamic behavior of the involving users such as mobility pattern.

Similar to [61], different multicast scenarios are being proposed in [63–65] by considering the overall system delay as the key performance indicator [63,64] and multicast coverage rate [64] to establish D2D multicast transmissions in a close proximity.

2.2.2 Full-Duplex

Full-Duplex communication means transmission and reception at the same time. Let us first illustrate the capability of FD communication in cache-enabled D2D network. If one user demands a content which is cached in a nearby user and at the same time, some other user(s) demand the content which is being cached in that user, both content delivery satisfactions can be done through full-duplex communication. Fig. 2.5 illustrates this capability for three different options.

- Two users might demand the contents that are being cached in each other, so they can exchange data simultaneously, e.g., see (a) in Fig. 2.5.
- One user might demand a content which is being cached in a nearby user and at the same time, one user (e.g. see (b) in Fig. 2.5) or multiple users (e.g. see (c) in Fig. 2.5) might demand the content which is being cached by that user. In the latter case, namely (c), FD user can target other receivers through multicast transmission.

It is noted that in this demonstration, the intermediate FD transceiver is not operating as a relay, namely, no data is being relayed between the users. However, there are some works [69–73] that considered FD capability as a relaying operation in cache-enabled D2D network and will be discussed in the sequel.

This type of communication for the cache-enabled D2D networks is being proposed by [22] and follow-up works utilized this capability in different scenarios [23–32, 68–73]. From the FD transmission perspective, existing works utilized FD capability either in none-relaying [23–26, 29–32, 68] or relaying [27, 28, 69–73] scenarios. In the sequel, the key aspects of these works are being explored.

None-Relaying Utilization

As discussed earlier, Fig. 2.5 (c) refers to the case in which the data is not being relayed over the FD node. The goal is to bring the benefit of the FD capability in the cache-enabled

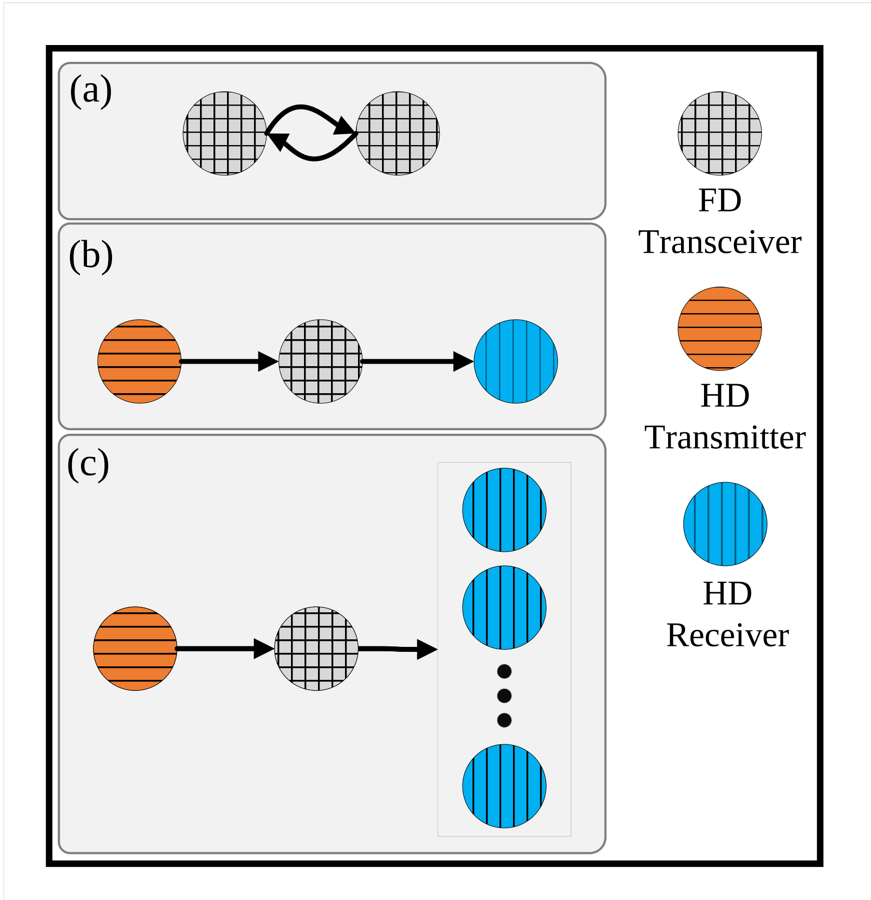

Figure 2.5 Full-Duplex Transmission Mode: (a): Bidirectional FD operation, (b): Half-Duplex operation, and (c): Three Node FD operation

network and these advantages are being explored from different perspectives. Authors in [22, 23] utilized FD capability in a cluster-based D2D caching scenario in which multiple links are allowed to be established within a cluster in contrast with the single link per cluster scenario as in [3]. It is shown that the FD capability in cache-enabled D2D network reduces the overall download time and increases the system sum throughput. Further investigation is conducted in [25] in which the significance of FD capability is being investigated in an ultra dense D2D network. It is shown that given a careful self-interference cancellation, FD-enabled caching network outperforms the conventional HD network in terms of the capacity yet with a slightly reduction in success probability due to extra interference imposed by more transmitters compared to its HD counterpart. Self-Request is also taken into account in these works. Further, the authors in [32] proposed an analytical framework for the success probability analysis of the FD-enabled caching D2D network in which the amount of interference is dictated by the number of transmitters in network.

Similar *Thinning* approach as in [55, 62] is being utilized in [26, 68] in which the density of full-duplex D2D pairs is being obtained based on some predefined distance threshold. It is assumed that the content of interest for a typical user can be fetched from its closed neighbor subject to meet some predefined distance threshold. Another effort made by [24] to investigate the probabilities of FD and HD transmission operations in a cluster-based caching D2D network. In particular, the goal is to obtain the coverage probability of the typical user and cluster spectral efficiency.

Inspired from [25, 32], further studies made by [26, 29–31] to analyze the impact of FD capability in cache-enabled D2D network. In particular, the authors utilized stochastic geometry and proposed a framework that takes into account the self-request and FD-powered opportunity to obtain the closed form expressions for content offloading probability and overall energy consumption. Numerical results reveal that first, FD capability achieves the largest content caching probability in comparison with the scheme that users mostly rely on cellular links to retrieve data, which means the largest. Second, FD-enabled network outperforms other schemes in terms of successful offloading probability apart from the scenario with large user density and heavy load. Third, the schemes with FD are always preferable to the schemes without FD. The authors in [29–31] aim to target a common goal, which is analyzing delivery time when nodes are empowered with FD capability. To achieve the goal, an optimal power control based on the linear precoding design is proposed to minimize the system delivery time. It is shown that given careful self-interference cancellation scheme, the FD-enabled network outperforms its HD counterpart

From the caching mechanism perspective, deterministic caching policy is being utilized in [22–25, 29–32, 68].

Relaying Utilization

In contrast to none-relaying utilization of the FD capability in cache-enabled D2D network, in which no data is being relayed over FD nodes, another research direction as in [27, 28, 69–73] considered relaying approach to improve the performance of the network. The idea of utilizing FD radios in a relaying operation can help improve the overall system throughput by delivering contents to the users that are either out of coverage from the central BS or the signal strength is weak, therefore the intermediate nodes can relay their desired content from the main transmitter. Authors in [27, 28] assumed that the content distribution within an area consisting of users follows Poisson process, therefore the density of relaying nodes are extracted using *Thinning Theory*. In this approach, the system considers the relaying mechanism as the resort when there is no nearby user to satisfy the request made by another user. Given these assumptions, the number of transmitters and receivers follows the same densities as it is been extracted using *Thinning Theory*.

An optimal target selection scenario is proposed in [69] to minimize the content delivery time. In this work, it is assumed that the number of transmitters and receivers are fixed and the caching policy does not have any impact of the number of transmitters and receivers. Relaying nodes are being selected based on the optimal distance from the main transmitter dictated by an optimization problem, in which the goal is to choose appropriate relays that leads to minimum possible content delivery time.

When there is no chance to deliver the desired content through a single relaying node, multi-hop D2D relaying scheme can be used to minimize the outage probability. This approach is being investigated in [70] to analyze throughput-outage tradeoff of the network when multiple hops are needed to deliver the desired content to the target user. Stochastic geometry is used to model the location of user devices while the content placement strategy is designed based on the popularity of the contents and availability of the intermediate nodes involved in the multi-hop transmission opportunities. The numerical results are compared with the single-hop case as in [72] and it is shown that the proposed multi-hop approach achieves better performance, providing a careful content placement strategy that maximizes the availability of the contents across the entire network. Another multi-hop approach is being investigated in [73] in which the goal is to optimize spectrum sharing and power allocation for D2D transmitters. In this work, relays are being selected by applying the interference-aware multi-hop path selection schemes to maximize the spectrum reuse and minimize the power consumption. It is assumed that the number of transmitters and receivers are fixed and the caching placement does not dictate the density of transmitters or receivers.

A joint cache-enabled Unmanned Aerial Vehicles (UAVs) and D2D communication is proposed in [71], in which the FD capability is applied to UAVs. Specifically, cache-enabled UAVs and D2D users jointly provide content transfer. Technically, if the caching of UAVs or user devices have the requested content locally, a local D2D communication link or a UAV downlink can be established. Otherwise, UAVs first fetch the desired contents from the BS and then deliver them to the users. Here, the distribution of the number of relays, transmitters, and receivers are not being dictated by the caching policy, rather, the transmitters/relays are being selected based on the chance of delivering the content to the target user.

It is also worth mentioning that apart from the above works that utilized FD capability in cache-enabled D2D system, there are some works that utilized None Orthogonal Multiple Access (NOMA) [74, 75] with the same goal of maximizing the cache-hit probability. While the FD and NOMA are different from the design perspective yet from the content delivery perspective, they are doing the same job to transmit and receive at the same time.

2.3 Objective of the Thesis in Comparison with the Existing Works

To outline the shortcomings in the existing literature and clarify the objective of the thesis, a small network example consisting of 7 number nodes is considered.

A given caching policy, let us assume that all those users in this small network make requests at random(the details about the caching and request procedures will be discussed later in the next chapter). There are many possible outcomes in this small network (example) as illustrated in Fig. 2.6. In this demonstration, direction of the arrows indicates the data direction. The resulting outcomes is a hypothetical random graph and the connections between the nodes in this graph is dictated by the following factors:

- Distribution of the contents on user devices
- Random requests of the users
- Number of users

As can be observed from this graph, the number of transmitters, and, correspondingly, the number of satisfied users changes because of the random nature of the above factors. Having knowledge on the possible outcomes can help get insights on the system behavior in terms of the expected number of the satisfied users. As discussed earlier, in the existing works, FD capability is being utilized in different scenarios and the advantages of the FD-enabled D2D caching network is being investigated from different perspectives. However, none of the existing works quantified these possible outcomes and it is not clear how precisely the aforementioned factors change the number of transmitters and receivers. Also, as can be observed from this graph, there are other possibilities rather than transmission and reception, such as cache-hit outage and self-request. Even though the self-request case is widely considered in the existing works, it is not clear how the density of these users changes due to aforementioned factors.

In fact, we can see that nodes in all outcomes randomly choose one of the modes as delineated in Fig. 2.7. These modes will be discussed in the next chapter, however, just from the direction of the arrows perspective, we can imagine the possible outcomes for a typical node.

Now, we came across to the following questions that forms the basis of the objectives in this thesis:

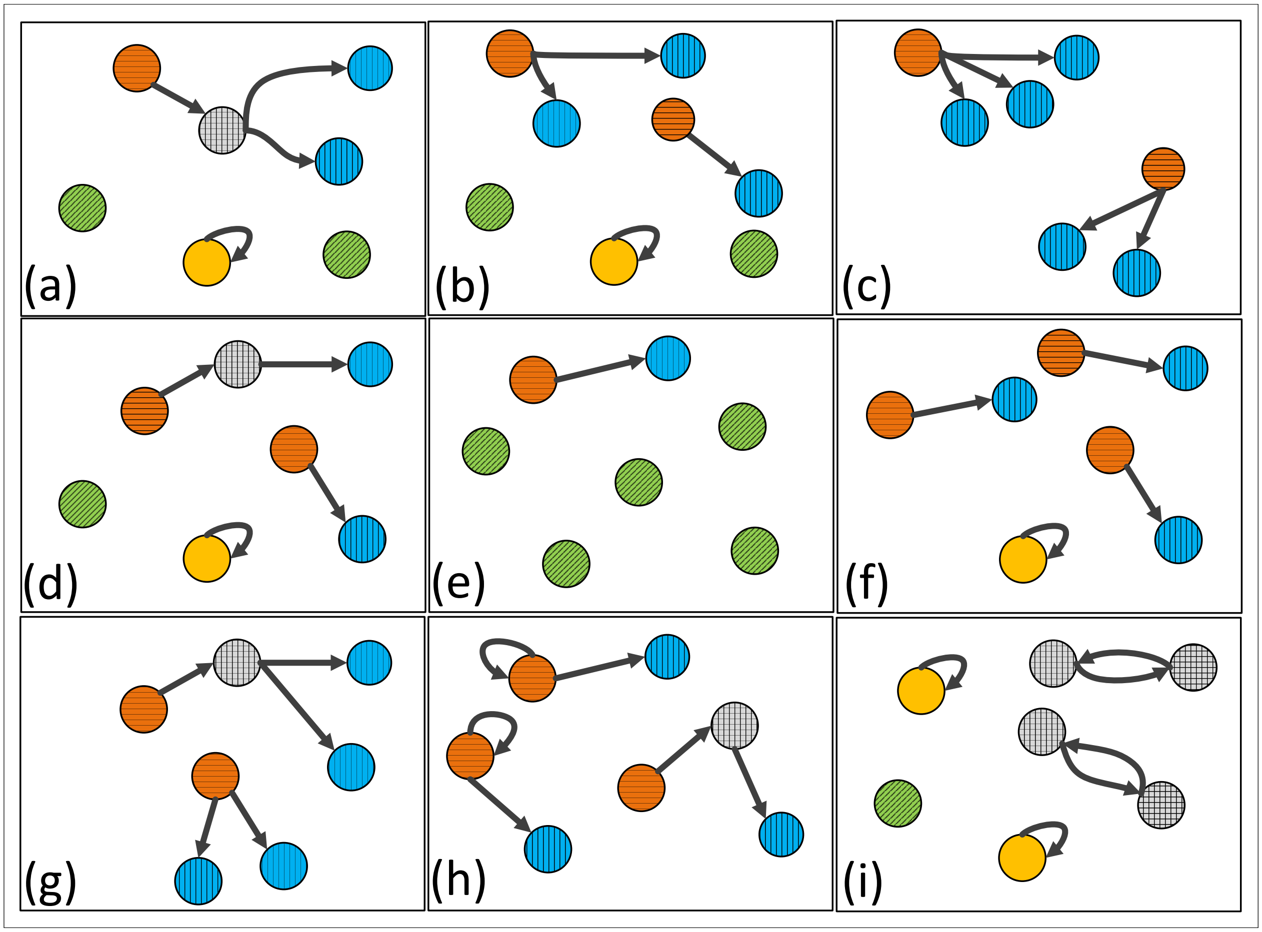

Figure 2.6 Some possible outcomes out of many[1]

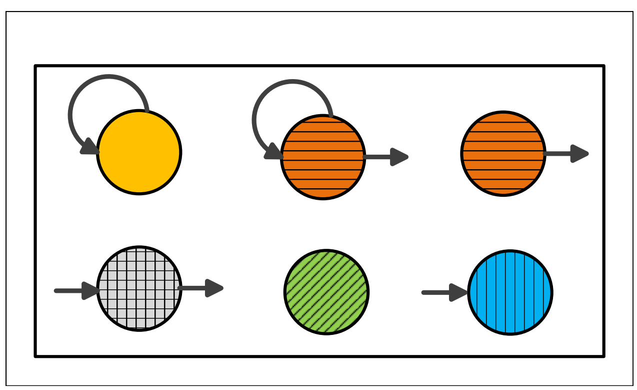

Figure 2.7 All possible outcome from the observations for a small network consisting 7 nodes

[1]There is no existing closed-form expression to obtain the number of possible outcomes for a given small network. However, from the observations, one can conclude that this number is extraordinarily large.

- Considering a typical node within the network, what is the probability of different outcomes for this typical node?
- How the caching strategy impacts the shape of the random graph and which factors have the most dominant impact on all possible outcomes?
- How the number of nodes within the network plays the role and how this role is connected to the caching policy?
- What is the expected number of satisfied users given the number of users, caching strategy and relevant distributions involved in the caching strategy?
- What is the expected number of transmitters in the network?
- How to quantify the role of FD capability in increasing the content delivery opportunities?

Given the above questions, in summary, Table 2.1 outlines the details on the comparison between the thesis objective and existing works. In this thesis, both deterministic and stochastic caching policies are being utilized, and all their key features are being studied through the analytical approach and the impact of all involving parameters are being studied in the numerical results. In contrast to the existing works in which the expected number of satisfied users, expected number of transmitters and receivers, are not being studied, in this thesis, all these factors are being taken into account in the analysis approach and quantified through the simulations.

Table 2.1 Objective of the thesis in comparison with the existing FD-D2D caching works

Ref.	**Self Request**	**Cache Outage**	**# of TXs**	**Sat. UEs**	**Caching Policy**	**Inv. of Caching Par.**
[32]	✓	✓	✓	×	Det.	×
[22]	✓	×	×	×	Det.	×
[23]	✓	×	×	×	Det.	×
[69]	✓	×	×	×	Det.	×
[70]	×	×	×	×	Det.	×
[71]	×	×	×	×	Det.	×
[72]	✓	×	×	×	Det.	×
[73]	×	×	×	×	Det.	×
[24]	×	×	✓	×	Det.	×
[25]	✓	✓	×	×	Det.	×
[26]	✓	×	✓	×	Det.	×
[27]	×	×	×	×	Det.	×
[28]	×	×	×	×	Det.	×
[29]	×	×	×	×	Det.	×
[30]	×	×	×	×	Det.	×
[31]	×	×	×	×	Det.	×
[68]	×	×	×	×	Det.	×
Thesis	✓	✓	✓	✓	Det. & Sto.	✓

CHAPTER 3 SYSTEM MODEL

3.1 System Overall Structure

The system is a D2D network of N fixed nodes overlaying a cellular network. All nodes are assumed to have FD capability. No constraints are imposed on the cache size of devices so that every device has enough storage to cache a number of contents.

In the network, there is a central entity that has full knowledge of the cached content and their random requests. That entity is called the central network. Details on how this system functions are given below, but first let us provide a summary of the notation used throughout the thesis that is also necessary to grasp the system explanations in Table 3.1.

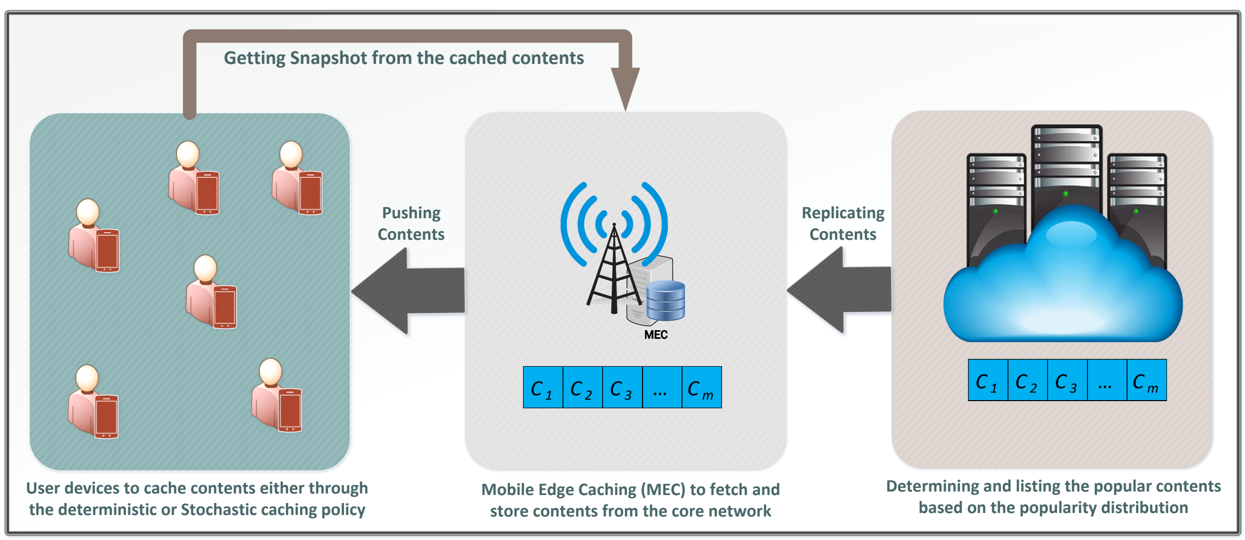

Figure 3.1 Demonstration of the centrally-controlled D2D caching network

3.2 Caching Model

It is assumed that the central network already tracks and stores a set of popular contents, as shown in Fig. 3.1. Two different caching policies called deterministic and stochastic are being considered and are both controlled through the central network, which is represented on this figure by a central BS. Note, however, that while the central BS plays the major role in determining the deterministic caching policy, it does not make decisions regarding the stochastic policy. However, it has knowledge on the cached contents and the users' random

Table 3.1 Summary of notation being used throughout the thesis

Notation	**Description**
m	Total number of contents
N	Number of users
L	Set of all contents
U	Set of users
L_N	Set of contents cached in N users in deterministic caching
c_ℓ	The ℓ-th content in the library
ρ_κ	Popularity of the κ-th content in deterministic caching
μ_ν	Probability of caching content c_ν in user u_ν
γ_r	Skew exponent in deterministic caching
γ_c	Skew exponent in stochastic caching
$\mathbf{A}$	Total permutations
$\mathbf{M}_{i\times j}(\mathbb{N})$	Matrix of size $i \times j$ in the field domain of $\mathbb{N}$
δ	Caching policy indicator; $\delta \in \{\mathrm{o}, \mathrm{s}\}$
Δ	User operating mode indicator
$\mathcal{P}_\Delta^\delta$	Probability of operating in mode Δ with a caching policy δ
$\mathcal{P}_{\mathrm{hit}}^\delta$	Hitting probability in a caching policy δ
$\mathbf{\Omega}^\nu$	a Block matrix associate to an arbitrary user u_ν
ω_{ij}^ν	Elements of a block matrix $\mathbf{\Omega}^\nu$
$a_{j\nu}$	Rows of indices of a block matrix $\mathbf{\Omega}^\nu$
$a_{j\nu}$	Columns of indices of a block matrix $\mathbf{\Omega}^\nu$
$r_i(.)$	Operator of taking i^{th} row of a matrix
$\mathcal{W}_i$	Event of missing some contents in the permutations
$\mathcal{X}_\nu$	Event of caching content u_ν in user u_ν
$\mathcal{Y}_\nu$	Event of caching the rest of contents in the rest of the users except u_ν
$\mathcal{Z}_\nu$	Event of requesting content c_ν cached in user u_ν
$\phi_{ij\nu}$	Event of occurrence of all events $\mathcal{W}_i$, $\mathcal{X}_\nu$, $\mathcal{Y}_\nu$, and $\mathcal{Z}_\nu$

requests. The assumption of centrally controlled caching policies is widely considered in the existing literature [2–4, 8, 10, 24, 26, 32, 76–80]. Let us denote the library of popular contents of size m by $\mathrm{L} = \{c_\ell \mid \ell \in \mathbb{N}, 1 \leq \ell \leq m\}$. Each content, i.e., c_ℓ, has an associated popularity score, which is characterized by the user requests during the period that the central network tracked those requests. This library is stored in terms of the popularity score descending order, i.e., content c_{m-1} is less popular than that of c_m. Similar to [22–32], it is assumed that all contents in the library are tracked and stored beforehand, and that the popularity distribution for all contents remains constant during the delivery procedures. Each user has a unique identity defined as u_κ, which is a member of set $\mathrm{U} = \{u_\kappa \,|\, \kappa \in \mathbb{N}, 1 \leq \kappa \leq N\}$. Now, to determine which contents are cached in each user, two different caching policies are being utilized as discussed in the sequel.

3.2.1 Deterministic Caching

In this policy, the central BS is responsible for pushing contents in user devices in a deterministic manner such that each user caches one content from the library. This assumption is widely considered in the existing literature [3,4,32]. According to this policy, each content is associated with a single user, which means that there is no overlap between cached contents in user devices. While each user has the capability of storing multiple contents, for the sake of simplicity, single content caching is assumed per user. Under these assumptions, user u_κ is assumed to cache content c_κ, where c_κ is different across users. For instance, users u_1, u_2, and u_3 store contents c_1, c_2, and c_3, respectively. It is clear that there is only one choice to push a subset of contents $\mathrm{L}_N = \{c_\ell \,|\, \ell \in \mathbb{N}, 1 \leq \ell \leq N\}$, where $\mathrm{L}_N \subseteq \mathrm{L}_m$ and $N \leq m$, in a set of users U consisting of N users. It is noted that the elements of sets L_N and U are mutually ordered. Without loss of generality, it is assumed that there is no preference in users identification labeling, since the users are sorted randomly within the set U; this assumption is justified in the next chapter. Zipf distribution is being utilized, which is a special case of the Riemann Zeta function because it is widely used in the existing literature [3,32,79]. According to the Zipf distribution, the popularity of content c_κ is equivalent to the probability of requesting content c_κ and this probability is denoted by ρ_κ, which is defined as follows [3,32,79]:

$$\rho_\kappa = \kappa^{-\gamma_r} \left(\sum_{\eta=1}^{m} \eta^{-\gamma_r} \right)^{-1}, \tag{3.1}$$

where κ is the content index and the parameter γ_r is the skew exponent and characterizes the popularity distribution by controlling the popularity of the contents for a given library size m. It is clear that $\left(\sum_{\kappa=1}^{N} \rho_\kappa = 1\right) \iff (N = m)$; otherwise, $\sum_{k=1}^{N} \rho_\kappa < 1$ for $N < m$.

3.2.2 Stochastic Caching

In this policy, users are caching content in a stochastic manner based on the given distribution (i.e., Zipf distribution) [81]. Even though the central BS does not push specific content into user devices and the users are caching content in a stochastic manner, we still assume that the central BS has a snapshot of the cached content in the user devices. This means that, at the end, the BS is aware of the cached contents through the feedback exchanged between the

BS and the users. The same library L and user set U is considered in this policy, as described in 3.2.1. Since each user caches content randomly from the library, there is the possibility of having overlap between users' caches. User u_ν caches content c_ν, $\nu \in \mathrm{L}$, and it is chosen at random based on a caching distribution, which will be discussed in the sequel. Choosing a set of random contents from the corresponding set of libraries produces permutations with repetition. There are m^N choices of subsets to assign N contents, with $N \leq m$ to N users randomly. Now, let us denote $\mathbf{A} \in \mathbf{M}_{m^N \times N}(\mathbb{N})$, $1 \leq a_{ij} \leq m$, as the permutations matrix with repetition. Every row of this matrix is a vector denoted by $r_i(\mathbf{A})$ with entries given by the i^{th} row of $\mathbf{A}$ and is called a permutation. For example, if the size of the library is 3, i.e., $\mathrm{L} = \{c_1, c_2, c_3\}$, and assuming that there are only two users, $\mathrm{U} = \{u_1, u_2\}$, all possible $m^N = 3^2 = 9$ permutations to store two contents (one per each user) in two users are as follows:
$r_1(\mathbf{A}) = [c_1, c_2]$, $r_2(\mathbf{A}) = [c_2, c_1]$, $r_3(\mathbf{A}) = [c_1, c_1]$, $r_4(\mathbf{A}) = [c_2, c_2]$, $r_5(\mathbf{A}) = [c_1, c_3]$, $r_6(\mathbf{A}) = [c_3, c_1]$, $r_7(\mathbf{A}) = [c_3, c_3]$, $r_8(\mathbf{A}) = [c_2, c_3]$, $r_9(\mathbf{A}) = [c_3, c_2]$. The probability of caching content c_ν in user u_ν is similar to the popularity distribution as in Eq. (3.1) with a different skew exponent γ_c. This probability is denoted as μ_ν, which is given by $\mu_\nu = \nu^{-\gamma_c} \left(\sum_{\eta=1}^{m} \eta^{-\gamma_c} \right)^{-1}$. It should also be noted that the whole analysis is flexible and can be used with any other caching distribution.

3.3 Modeling User Operating Modes

It is assumed that a user randomly requests content from the library according to the popularity distribution given by (3.1). A pair of users can potentially initiate a D2D connection if one of them finds its desired content in the other user. Based on the information of the cached contents and users' requests, there are different operating modes for an arbitrary user, as shown in Fig. 3.2. Definitions of the operating modes are given below. In what follows, $\delta \in \{\mathrm{d}, \mathrm{s}\}$ denotes the caching policy where "d" and "s" stand for the deterministic and stochastic, respectively.

- **Self-request (SR)**: in this mode, an arbitrary user can find its desired content in its own cache. Let $\mathcal{P}_{\mathrm{SR}}^{\delta}$ be the probability of this mode.
- **Self-request and HD transmission (SR-HDTX)**: according to this mode, an arbitrary user can find its desired content in its own cache and can concurrently serve for other users' demand. $\mathcal{P}_{\mathrm{SR-HDTX}}^{\delta}$ denotes the probability of this mode.
- **Full-Duplex Transceiver (FDTR)**: In this mode, an arbitrary user can find its

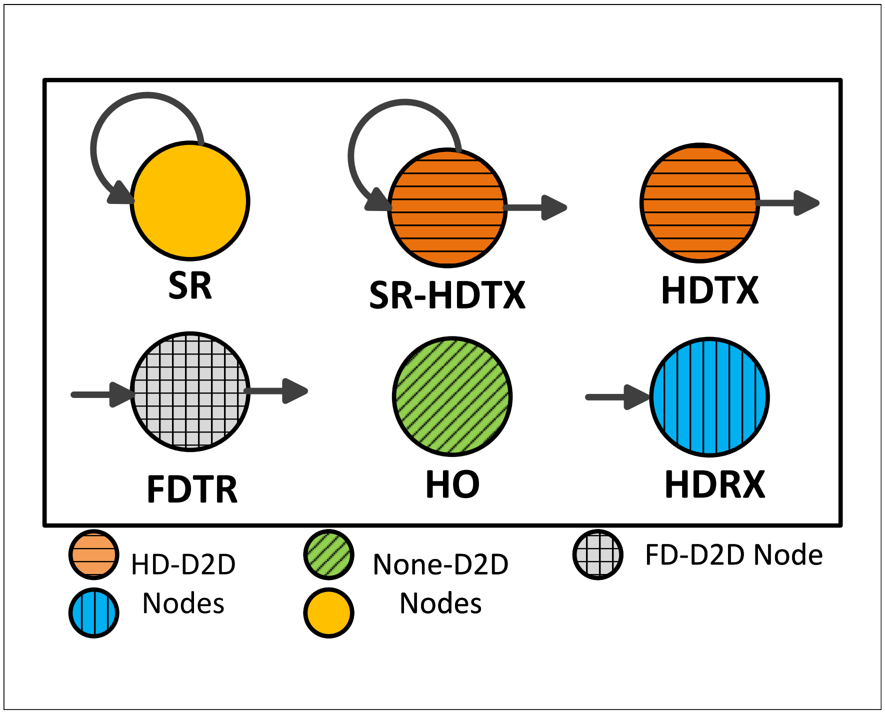

Figure 3.2 All possible operating modes for an arbitrary node

desired content in its vicinity through a D2D link and can concurrently serve other users' demand. Let us denote $\mathcal{P}^{\delta}_{\text{FDTR}}$ as the probability of this mode.

- **Half-Duplex Transmitter (HDTX)**: According to this mode, an arbitrary user cannot find its desired content either in its vicinity or in its own cache; however, it can serve other users' demand. Let us denote $\mathcal{P}^{\delta}_{\text{HDTX}}$ as the probability of this mode.

- **Half-Duplex Receiver (HDRX)**: An arbitrary user can receive its desired content via a D2D link, and there are no user(s) that demand(s) the content that is cached in this user. Let us denote $\mathcal{P}^{\delta}_{\text{HDRX}}$ as the probability of this mode.

- **Hitting Outage (HO)**: This mode takes place when an arbitrary user cannot find its desired content in its vicinity or its own cache, and there is(are) no user(s) that demand(s) its cached content. Let us denote $\mathcal{P}^{\delta}_{\text{HO}}$ as the probability of this mode.

In the next chapter, the analysis of each mode is conducted with respect to different caching policies.

CHAPTER 4 ANALYZING NODE OPERATING PROBABILITIES

In this chapter, the goal is to derive closed-form expressions for all operating modes by considering both deterministic and stochastic caching policies.

4.1 Deterministic Caching

Given an arbitrary node and denoting $\mathcal{P}_{\Delta}^{\mathrm{d}}$ as the probability of operating mode $\Delta \in \{\mathrm{SR}, \mathrm{SR-HDTX}, \mathrm{FDTR}, \mathrm{HDTX}, \mathrm{HDRX}, \mathrm{HO}\}$ when considering the deterministic caching policy, we have the theorem as follows:

Theorem 1. *In a cache-enabled FD-D2D network endowed by a deterministic caching policy, the probabilities of all possible operating modes for an arbitrary user are as follows:*

$$\mathcal{P}_{\mathrm{SR}}^{\mathrm{d}} = \frac{1}{N} \sum_{\kappa=1}^{N} \rho_\kappa \left(1 - \rho_\kappa\right)^{N-1}, \tag{4.1}$$

$$\mathcal{P}_{\mathrm{SR-HDTX}}^{\mathrm{d}} = \frac{1}{N} \sum_{\kappa=1}^{N} \rho_\kappa \left(1 - \left(1 - \rho_\kappa\right)^{N-1}\right), \tag{4.2}$$

$$\mathcal{P}_{\mathrm{FDTR}}^{\mathrm{d}} = \frac{1}{N} \sum_{\kappa=1}^{N} \left(\mathcal{P}_{\mathrm{hit}}^{\mathrm{d}} - \rho_\kappa\right) \left(1 - \left(1 - \rho_\kappa\right)^{N-1}\right), \tag{4.3}$$

$$\mathcal{P}_{\mathrm{HDRX}}^{\mathrm{d}} = \frac{1}{N} \sum_{\kappa=1}^{N} \left(\mathcal{P}_{\mathrm{hit}}^{\mathrm{d}} - \rho_\kappa\right) \left(1 - \rho_\kappa\right)^{N-1}, \tag{4.4}$$

$$\mathcal{P}_{\mathrm{HDTX}}^{\mathrm{d}} = \frac{1}{N} \sum_{\kappa=1}^{N} \left(1 - \mathcal{P}_{\mathrm{hit}}^{\mathrm{d}}\right) \left(1 - \left(1 - \rho_\kappa\right)^{N-1}\right), \tag{4.5}$$

$$\mathcal{P}_{\mathrm{HO}}^{\mathrm{d}} = \frac{1}{N} \sum_{\kappa=1}^{N} \left(1 - \mathcal{P}_{\mathrm{hit}}^{\mathrm{d}}\right) \left(1 - \rho_\kappa\right)^{N-1}, \tag{4.6}$$

where ρ_κ is given in Eq. (3.1) and $\mathcal{P}_{\mathrm{hit}}^{\mathrm{d}} = \sum_{\kappa=1}^{N} \rho_\kappa$, which is the hitting probability or the probability at which a given random request from an arbitrary user hits the given subset of the probability space defined in subsection 3.2.1.

Proof. From Fig. 3.2, it can be inferred that the probability of occurrence of each mode at an arbitrary user u_κ depends on two different events: i) The request of the user u_κ (we denote this event by $\mathcal{V}$ and the probability of this event by $\mathcal{P}_{\Delta}^{v,\kappa}$) and ii) The requests from other users for the content cached in user u_κ. Let us denote this event by $\mathcal{W}$ and the probability of this event by $\mathcal{P}_{\Delta}^{w,\kappa}$. The joint probability of both events, i.e., $\mathbb{P}(\mathcal{V}, \mathcal{W})$ gives the probability of the operating mode Δ for a specific node u_κ. Since the requests of all users are independent from each other, hence, $\mathbb{P}(\mathcal{V},\mathcal{W}) = \mathbb{P}(\mathcal{V})\mathbb{P}(\mathcal{W}) = \mathcal{P}_{\Delta}^{v,\kappa}\mathcal{P}_{\Delta}^{w,\kappa}$. Now, by using the *law of total probability* and given the deterministic caching policy, the probability of operating mode Δ denoted by $\mathcal{P}_{\Delta}^{\mathrm{d}}$ for an arbitrary node can be defined as follows:

$$\mathcal{P}_{\Delta}^{\mathrm{d}} = \sum_{\kappa=1}^{N} \mathcal{P}_{\Delta}^{v,\kappa}\mathcal{P}_{\Delta}^{w,\kappa}\mathcal{P}_{u_\kappa}, \tag{4.7}$$

where $\mathcal{P}_{u_\kappa}$ is the probability of choosing an arbitrary user among N users uniformly at random. The proof is provided for the HDTX mode; however, the approach remains the same for the other modes. Now, let us define two binary random variables $\mathcal{I}_\kappa$ and $\mathcal{J}_{\tau,\kappa}$ for u_κ as follows:

$$\mathcal{I}_\kappa = \begin{cases} 0 & ; u_\kappa \text{ cannot find its desired content} \\ 1 & ; u_\kappa \text{ can find its desired content,} \end{cases} \tag{4.8}$$

$$\mathcal{J}_{\tau,\kappa} = \begin{cases} 0 & ; u_\tau \text{ does not demand for content } c_\kappa \\ 1 & ; u_\tau \text{ demands for content } c_\kappa. \end{cases} \tag{4.9}$$

The probability $\Pr(\mathcal{I}_\kappa = 0)$ is equivalent to the situation where u_κ demands content that is not cached by other nodes in its vicinity, i.e., $\Pr(\mathcal{I}_\kappa = 0) = 1 - \mathcal{P}_{\mathrm{hit}}^{\mathrm{d}}$, which corresponds to the parameter $\mathcal{P}_{\mathrm{HDTX}}^{v,\kappa}$, i.e.,

$$\mathcal{P}_{\mathrm{HDTX}}^{v,\kappa} = \Pr(\mathcal{I}_\kappa = 0), \tag{4.10}$$

and the probability $\Pr(\mathcal{J}_{\tau,\kappa} = 0)$ is equivalent to

$$\Pr(\mathcal{J}_{\tau,\kappa} = 0) = 1 - \rho_\kappa. \tag{4.11}$$

Now, the parameter $\mathcal{P}_{\mathrm{HDTX}}^{w,\kappa}$ is equivalent to the probability that there is at least one node that demands content c_κ, hence

$$\begin{aligned}\mathcal{P}_{\mathrm{HDTX}}^{w,\kappa} &= 1 - \Pr\left(\bigcup_{\tau=1,\tau\neq\kappa}^{N} \mathcal{J}_{\tau,\kappa} = 0\right) \\ &\overset{(a)}{=} 1 - \prod_{\tau=1,\tau\neq\kappa}^{N} \Pr(\mathcal{J}_{\tau,\kappa} = 0) \\ &\overset{(b)}{=} 1 - (1-\rho_\kappa)^{N-1},\end{aligned} \tag{4.12}$$

where (a) follows the fact that the requests of all users are independent from each other and (b) follows directly using Eq. (4.11). It is clear that choosing an arbitrary node out of N

nodes is equal to $\frac{1}{N}$, which is valid for all modes; hence, $\mathcal{P}_{u_\kappa}$ can be defined as follows:

$$\mathcal{P}_{u_\kappa} = \frac{1}{N}. \tag{4.13}$$

For any arbitrary node, the probability for all modes can be defined by taking expectation over all possible values for κ. Now, by substituting the Eqs. (4.10), (4.12), and (4.13) in Eq. (4.7), the final expression can be obtained as in Eq. (4.3) as follows:

$$\mathcal{P}^{\mathrm{d}}_{\mathrm{HDTX}} = \frac{1}{N}\sum_{\kappa=1}^{N}\left(1-\mathcal{P}^{\mathrm{d}}_{\mathrm{hit}}\right)\left(1-\left(1-\rho_\kappa\right)^{N-1}\right). \tag{4.14}$$

□

4.2 Stochastic Caching

Given an arbitrary node and denoting $\mathcal{P}^{\mathrm{s}}_{\Delta}$ as the probability of operating mode Δ when considering the stochastic caching policy, we have the theorem as follows:

Theorem 2. *In a cache-enabled FD-D2D network endowed with a stochastic caching policy, the probabilities of all possible operating modes for an arbitrary user are as follows:*

$$\mathcal{P}^{\mathrm{s}}_{\mathrm{SR}} = \sum_{x\in L^1}\rho_x\left(1-\rho_x\right)^{N-1}\mu_x, \tag{4.15}$$

$$\mathcal{P}^{\mathrm{s}}_{\mathrm{SR-HDTX}} = \sum_{x\in L^1}\rho_x\left(1-\left(1-\rho_x\right)^{N-1}\right)\mu_x, \tag{4.16}$$

$$\mathcal{P}^{\mathrm{s}}_{\mathrm{FDTR}} = \mathbf{R}^{\mathrm{T}}\mathbf{C}_\alpha\left(\mathbf{e}-\mathbf{H}\right)\mathbf{C}_\beta, \tag{4.17}$$

$$\mathcal{P}^{\mathrm{s}}_{\mathrm{HDRX}} = \left(\mathbf{e}-\mathbf{R}\right)^{\mathrm{T}}\mathbf{C}_\alpha\left(\mathbf{e}-\mathbf{H}\right)\mathbf{C}_\beta, \tag{4.18}$$

$$\mathcal{P}^{\mathrm{s}}_{\mathrm{HDTX}} = \mathbf{R}^{\mathrm{T}}\mathbf{C}_\alpha\mathbf{H}\mathbf{C}_\beta, \tag{4.19}$$

$$\mathcal{P}^{\mathrm{s}}_{\mathrm{HO}} = \left(\mathbf{e}-\mathbf{R}\right)^{\mathrm{T}}\mathbf{C}_\alpha\mathbf{H}\mathbf{C}_\beta, \tag{4.20}$$

Proof. The proof is conducted for $\mathcal{P}^{s}_{\mathrm{HDTX}}$, but the methodology for the rest of the modes in Theorem 2 is similar. Let us first consider an example where there are three users (i.e., $\mathrm{U} = \{u_1, u_2, u_3\}$) and a library of contents with size $m = 4$ (i.e., $\mathrm{L} = \{c_1, c_2, c_3, c_4\}$). Let us build a table of outcomes along with the associated permutations matrix $\mathbf{A}$ and column vectors $r_i(\mathbf{A})$, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ as in Table 4.1. All those matrices will be explained in the sequel. As discussed in section 3.2.2, there are m^N possible outcomes for all users that cache contents at random, and these permutations for this specific example are indicated in the columns of $\mathbf{A}$ (i.e., columns c_1, c_2, and c_3 are associated with three users u_1, u_2, and

In Theorem 2,

$$\mathbf{H}=\begin{bmatrix}\Pr\left(\mathcal{W}_1=0\mid \mathrm{r}_1(\boldsymbol{\Omega}^{\mathbf{1}})\right) & \Pr\left(\mathcal{W}_2=0\mid \mathrm{r}_2(\boldsymbol{\Omega}^{\mathbf{1}})\right) & \dots & \Pr\left(\mathcal{W}_{\mathrm{m}^{\mathrm{N}-1}}=0\mid \mathrm{r}_{\mathrm{m}^{\mathrm{N}-1}}(\boldsymbol{\Omega}^{\mathbf{1}})\right)\\ \vdots & \vdots & \vdots & \vdots \\ \Pr\left(\mathcal{W}_1=0\mid \mathrm{r}_1(\boldsymbol{\Omega}^{\mathbf{m}})\right) & \Pr\left(\mathcal{W}_2=0\mid \mathrm{r}_2(\boldsymbol{\Omega}^{\mathbf{m}})\right) & \dots & \Pr\left(\mathcal{W}_{\mathrm{m}^{\mathrm{N}-1}}=0\mid \mathrm{r}_{\mathrm{m}^{\mathrm{N}-1}}(\boldsymbol{\Omega}^{\mathbf{m}})\right)\end{bmatrix}_{m\times m^{N-1}},$$

$$\mathbf{e}=\begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix}_{m^N\times 1},\quad \mathbf{C}_\alpha=\begin{bmatrix}\mu_1 & 0 & \dots & 0\\ 0 & \mu_2 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \mu_m\end{bmatrix}_{m\times m},\quad \mathbf{C}_\beta=\begin{bmatrix}\Pr\left(\mathcal{Y}_1\mid \mathrm{r}_1(\boldsymbol{\Omega}^{\mathbf{1}})\right)\\ \vdots \\ \Pr\left(\mathcal{Y}_{\mathrm{m}^{\mathrm{N}-1}}\mid \mathrm{r}_{\mathrm{m}^{\mathrm{N}-1}}(\boldsymbol{\Omega}^{\mathbf{1}})\right)\end{bmatrix}_{m^{N-1}\times 1},$$

$$\mathbf{R}=\begin{bmatrix}\Pr\left(\mathcal{Z}_1\mid \mathrm{r}_1(\boldsymbol{\Omega}^{\mathbf{1}})\right)\\ \vdots \\ \Pr\left(\mathcal{Z}_{\mathrm{m}}\mid \mathrm{r}_1(\boldsymbol{\Omega}^{\mathbf{m}})\right)\end{bmatrix}_{m\times 1},$$

$\Pr\left(\mathcal{W}_{\mathrm{i}}=0\mid \mathrm{r}_{\mathrm{i}}(\mathbf{A})\right)=1-\sum_{\substack{\{\theta_{\mathrm{j}}\}\in \mathrm{r}_{\mathrm{i}}(\mathbf{A})\\ \theta_{\mathrm{j}}\neq\theta_{\mathrm{k}}}}\rho_{\theta_{\mathrm{j}}}$, $\Pr(\mathcal{Y}_\nu\mid \mathrm{r}_{\mathrm{j}}(\boldsymbol{\Omega}^\nu))=\prod_{\{\theta_{\mathrm{i}}\mid\theta_{\mathrm{i}}\neq\nu\}\in \mathrm{r}_{\mathrm{j}}(\boldsymbol{\Omega}^\nu)}\mu_{\theta_{\mathrm{i}}}$, $\Pr(\mathcal{Z}_\nu\mid \mathrm{r}_{\mathrm{j}}(\boldsymbol{\Omega}^\nu))=1-(1-\rho_\nu)^{\mathrm{N}-1}$, $\boldsymbol{\Omega}^\nu=\mathbf{A}(a_{j\nu};b_k)$, $a_{j\nu}=\{(i+\mathbb{1}_\nu(\nu-1))m^{N-1}\mid i\in\{1,2,\dots,m\}\}$, $b_k=\{1,2,\dots,N\}$, and $\mathbb{1}_\nu=\begin{cases}0 & ;\nu=1\\ 1 & ;o.w.\end{cases}$

---

u_3, respectively). The number of rows in this table is $m^N=4^3=64$, which are associated with the vector $r_i(\mathbf{A})$, $i\in\{1,2,\dots,m^N\}$. Now, let us pick up an arbitrary node among $N=3$ nodes, namely, one of users u_1, u_2, and u_3. According to explanations in subsection 3.2.2, since the cached contents and the random requests are independent and identically distributed (i.i.d.) among all users, there is no preference in choosing an arbitrary node. Thus, without loss of generality, let us pick up user u_1 with cache content c_1. The rest of the column vectors, namely, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, are formed based on the values given by the permutations matrix $\mathbf{A}$, and they will be explained in the sequel. Table 4.1 is divided into four equally sized blocks. Each block covers $m^{N-1}=4^2=16$ rows. First, let us aim to formulate matrix blocking for any size of $\mathbf{A}$. Denote $\boldsymbol{\Omega}^\nu$, $\nu\in\{1,2,\dots,m\}$, as the *block matrix*, where $\boldsymbol{\Omega}^\nu\in\mathbf{M}_{m^{N-1}\times N}(\mathbb{N})$, $1\leq\omega_{ij}^\nu\leq m$, $N\leq m$. This block matrix can be obtained as follows:

$$\boldsymbol{\Omega}^\nu=\mathbf{A}(a_{j\nu};b_k),\tag{4.21}$$

where ν is the content index as described in subsection 3.2.2, $b_k=\{1,2,\dots,N\}$ is the column index, and $a_{j\nu}$ accounts for the row index and can be formulated as follows:

$$\begin{aligned}a_{j\nu}=&\{1+\mathbb{1}_\nu(\nu-1)m^{N-1},2+\mathbb{1}_\nu(\nu-1)m^{N-1},\\ &\dots,(1+\mathbb{1}_\nu(\nu-1))m^{N-1}\}\\ =&\{(i+\mathbb{1}_\nu(\nu-1))m^{N-1}\mid i\in\{1,2,\dots,m\}\},\end{aligned}\tag{4.22}$$

Table 4.1 Table of Caching Permutations for N=3 and m=4

$r_i(\mathbf{A})$	c_1	c_2	c_3	**W**	**X**	**Y**	**Z**
	A						
1	1	1	1	$\rho_2+\rho_3+\rho_4$	μ_1	$\mu_1\mu_1$	$1-(1-\rho_1)^2$
2	1	1	2	$\rho_3+\rho_4$	μ_1	$\mu_1\mu_2$	$1-(1-\rho_1)^2$
3	1	1	3	$\rho_2+\rho_4$	μ_1	$\mu_1\mu_3$	$1-(1-\rho_1)^2$
4	1	1	4	$\rho_2+\rho_3$	μ_1	$\mu_1\mu_4$	$1-(1-\rho_1)^2$
$\vdots$	1	$\vdots$	$\vdots$	$\vdots$	μ_1	$\vdots$	$\vdots$
16	1	4	4	$\rho_2+\rho_3$	μ_1	$\mu_4\mu_4$	$1-(1-\rho_1)^2$
17	2	1	1	$\rho_3+\rho_4$	μ_2	$\mu_1\mu_1$	$1-(1-\rho_2)^2$
18	2	1	2	$\rho_3+\rho_4$	μ_2	$\mu_1\mu_2$	$1-(1-\rho_2)^2$
19	2	1	3	ρ_4	μ_2	$\mu_1\mu_3$	$1-(1-\rho_2)^2$
20	2	1	4	ρ_3	μ_2	$\mu_1\mu_4$	$1-(1-\rho_2)^2$
$\vdots$	2	$\vdots$	$\vdots$	$\vdots$	μ_2	$\vdots$	$\vdots$
$\vdots$	2	2	2	$\rho_1+\rho_3+\rho_4$	μ_2	$\mu_2\mu_2$	$1-(1-\rho_2)^2$
$\vdots$	2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
32	2	4	4	$\rho_1+\rho_3$	μ_2	$\mu_4\mu_4$	$1-(1-\rho_2)^2$
33	3	1	1	$\rho_2+\rho_4$	μ_3	$\mu_1\mu_1$	$1-(1-\rho_3)^2$
34	3	1	2	ρ_4	μ_3	$\mu_1\mu_2$	$1-(1-\rho_3)^2$
35	3	1	3	$\rho_2+\rho_4$	μ_3	$\mu_1\mu_3$	$1-(1-\rho_3)^2$
36	3	1	4	ρ_2	μ_3	$\mu_1\mu_4$	$1-(1-\rho_3)^2$
$\vdots$	3	$\vdots$	$\vdots$	$\vdots$	μ_3	$\vdots$	$\vdots$
$\vdots$	3	3	3	$\rho_1+\rho_2+\rho_4$	μ_3	$\mu_3\mu_3$	$1-(1-\rho_3)^2$
$\vdots$	3	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
48	3	4	4	$\rho_1+\rho_2$	μ_3	$\mu_4\mu_4$	$1-(1-\rho_3)^2$
49	4	1	1	$\rho_2+\rho_3$	μ_4	$\mu_1\mu_1$	$1-(1-\rho_4)^2$
50	4	1	2	ρ_3	μ_4	$\mu_1\mu_2$	$1-(1-\rho_4)^2$
51	4	1	3	ρ_2	μ_4	$\mu_1\mu_3$	$1-(1-\rho_4)^2$
52	4	1	4	$\rho_2+\rho_3$	μ_4	$\mu_1\mu_4$	$1-(1-\rho_4)^2$
$\vdots$	4	$\vdots$	$\vdots$	$\vdots$	μ_4	$\vdots$	$\vdots$
64	4	4	4	$\rho_1+\rho_2+\rho_3$	μ_4	$\mu_4\mu_4$	$1-(1-\rho_4)^2$

where $\mathbb{1}_\nu$ is an indicator function and defined as follows:

$$\mathbb{1}_\nu = \begin{cases} 0 & ;\nu = 1 \\ 1 & ;o.w. \end{cases} \tag{4.23}$$

Now, let us explain the vector columns in Table 4.1.

W: This column vector indicates the possible hitting probability where user u_1 cannot find their desired content in their vicinity. The reason for this indication comes from the definition of mode HDTX, as discussed earlier in subsection 3.3. According to the definition of this mode and given any row of permutations $r_i(\mathbf{A})$, user u_1 should request any content from library L except the contents that appear in the i^{th} permutation row, e.g., by considering the first permutation $r_1(\mathbf{A})$, [1 1 1]. User u_1 should request any content except c_1, which means that it should request one of the contents c_2, c_3, and c_4. The probability of this event is the summation of all probabilities associated with those contents missing in this permutation, namely, $\rho_2 + \rho_3 + \rho_4$, which is the first element of column vector $\mathbf{W}$, i.e., $r_1(\mathbf{W})$. Now, let us define a binary random variable $\mathcal{W}_i$ as the event when missing some content indices in the permutation $r_i(\mathbf{A})$, which is given as follows:

$$\mathcal{W}_i = \begin{cases} 0 & ;\ \text{there are missing indices in } r_i(\mathbf{A}) \\ 1 & ;\ \text{all indices are found in } r_i(\mathbf{A}). \end{cases} \tag{4.24}$$

Now, the conditional probabilities of events $\mathcal{W}_i = 0$ and $\mathcal{W}_i = 1$ are, respectively defined as follows:

$$\begin{aligned} \Pr\left(\mathcal{W}_{\mathrm{i}} = 0 \mid \mathrm{r_i}(\mathbf{A})\right) =& \sum_{\nu=1}^{m} \rho_\nu - \sum_{\substack{\{\theta_j\} \in r_i(\mathbf{A}) \\ \theta_j \neq \theta_k}} \rho_{\theta_j} \\ =& 1 - \sum_{\substack{\{\theta_j\} \in r_i(\mathbf{A}) \\ \theta_j \neq \theta_k}} \rho_{\theta_j} \end{aligned} \tag{4.25}$$

$$\Pr\left(\mathcal{W}_{\mathrm{i}} = 1 \mid \mathrm{r_i}(\mathbf{A})\right) = 1 - \Pr\left(\mathcal{W}_{\mathrm{i}} = 0 \mid \mathrm{r_i}(\mathbf{A})\right). \tag{4.26}$$

One can apply expressions in Eqs. (4.24), (4.25), and (4.26) for the permutations of block matrix $\mathbf{\Omega}^\nu$, which are used later when we aim to build matrix expressions of the above equations.

X: Given user u_ν, this column indicates the probability of the event that this user cached the first content of the permutation $r_j(\mathbf{\Omega}^\nu)$, i.e., the probability of caching content associated with the first column of Table 4.1. We denote this event by $\mathcal{X}_\nu$ and the conditional probability of this event by $\Pr(\mathcal{X}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu))$, which is as follows:

$$\Pr(\mathcal{X}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu)) = \mu_\nu, \tag{4.27}$$

Y: Excluding user u_ν from the permutation $r_j(\mathbf{\Omega}^\nu)$, the event of caching the rest of the contents associated with the rest of the columns of the vector $r_j(\mathbf{\Omega}^\nu)$ is denoted by $\mathcal{Y}_\nu$ and the conditional probability of this event is denoted by $\Pr(\mathcal{Y}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu))$ and is given as follows:

$$\Pr(\mathcal{Y}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu)) = \prod_{\{\theta_\mathrm{i} \mid \theta_\mathrm{i} \neq \nu\} \in \mathrm{r_j}(\mathbf{\Omega}^\nu)} \mu_{\theta_\mathrm{i}}. \tag{4.28}$$

Z: This vector column indicates the probability of the event that there is no user in the vicinity of user u_ν that demands content c_ν. We denote this event by $\mathcal{Z}_\nu$ and the conditional probability of this event by $\Pr(\mathcal{Z}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu))$. The methodology to obtain this probability is similar to the strategy that is being used to obtain Eq. (4.12) in the proof of the Theorem 1. Hence, we can write the equation as follows:

$$\Pr(\mathcal{Z}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu)) = 1 - (1 - \rho_\nu)^{\mathrm{N}-1}. \tag{4.29}$$

All events $\mathcal{W}_i$, $\mathcal{X}_\nu$, $\mathcal{Y}_\nu$, and $\mathcal{Z}_\nu$ are independent from each other. We denote $\phi_{ij\nu}$ as the event of occurrence of all events jointly, and the conditional probability of this event is $\Pr(\phi_{\mathrm{ij}\nu} \mid \mathrm{r_i}(\mathbf{A}), \mathrm{r_j}(\mathbf{\Omega}^\nu))$, which can be calculated as follows:

$$\begin{aligned}
\Pr(\phi_{ij\nu} \mid r_i(\mathbf{A}), r_j(\mathbf{\Omega}^\nu)) &= \Pr(\mathcal{W}_\mathrm{i}, \mathcal{X}_\nu, \mathcal{Y}_\nu, \mathcal{Z}_\nu \mid \mathrm{r_i}(\mathbf{A}), \mathrm{r_j}(\mathbf{\Omega}^\nu)) \\
&\overset{(a)}{=} \Pr(\mathcal{W}_\mathrm{i} = 0 \mid \mathrm{r_i}(\mathbf{A})) \Pr(\mathcal{X}_\nu, \mathcal{Y}_\nu, \mathcal{Z}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu)) \\
&\overset{(a)}{=} \Pr(\mathcal{W}_\mathrm{i} = 0 \mid \mathrm{r_i}(\mathbf{A})) \Pr(\mathcal{X}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu)) \Pr(\mathcal{Y}_\nu \mid \mathrm{r_j}(\mathbf{\Omega}^\nu)) \\
&\quad \times Pr(\mathcal{Z}_\nu \mid r_j(\mathbf{\Omega}^\nu)) \\
&\overset{(b)}{=} \mu_\nu \left(1 - \sum_{\substack{\{\theta_j\} \in r_i(\mathbf{A}) \\ \theta_j \neq \theta_k}} \rho_{\theta_j} \right) \times \prod_{\{\theta_j \mid \theta_j \neq \nu\} \in r_j(\mathbf{\Omega}^\nu)} \mu_{\theta_j} \\
&\quad \times \Big(1 - (1 - \rho_\nu)^{N-1} \Big),
\end{aligned} \tag{4.30}$$

where (a) follows the independence of all aforementioned events and (b) follows a substitution of Eqs. (4.25), (4.27), (4.28), and (4.29) in the follow-up equation of (a). In our example, by substituting values $i = 1$, $j = 1$, and $\nu = 1$, the following expression is obtained, which is associated with the event of permutation $r_1(\mathbf{A}) = [1\ 1\ 1]$.

$$\Pr(\phi_{111} \mid \mathcal{W}_1, \mathcal{X}_1, \mathcal{Y}_1, \mathcal{Z}_1) = \mu_1^3 \Big(1 - \rho_1\Big) \Big(1 - (1 - \rho_1)^2\Big), \tag{4.31}$$

where $\Big(1 - \rho_1\Big) = \rho_2 + \rho_3 + \rho_4$ as explained in Eq. (4.25). The expression in Eq. (4.31) gives the conditional probability of interest for the specific event ϕ_{111} among all possible permutations. To obtain the final expression for the probability of interest $\mathcal{P}^s_{\mathrm{HDTX}}$ for the

HDTX mode, we need to take expectation over all possible outcomes of event $\phi_{ij\nu}$, as follows:

$$\mathcal{P}^{s}_{\text{HDTX}} = \sum_{i=1}^{m^N}\sum_{j=1}^{m}\sum_{\nu=1}^{N} \Pr\left(\phi_{\text{ij}\nu} \mid \mathcal{W}_{\text{i}}, \mathcal{X}_{\nu}, \mathcal{Y}_{\nu}, \mathcal{Z}_{\nu}\right). \tag{4.32}$$

For our example, the final expression for the HDTX operating mode is

$$\begin{aligned}\mathcal{P}^{s}_{\text{HDTX}} =& \mu_1 p_1[\mu_1^2(\rho_2+\rho_3+\rho_4)+\cdots+\mu_4^2(\rho_2+\rho_3)]\\ &+\mu_2 p_2[\mu_1^2(\rho_3+\rho_4)+\cdots+\mu_4^2(\rho_1+\rho_3)]\\ &+\mu_3 p_3[\mu_1^2(\rho_2+\rho_4)+\cdots+\mu_4^2(\rho_1+\rho_2)]\\ &+\mu_4 p_4[\mu_1^2(\rho_2+\rho_3)+\cdots+\mu_4^2(\rho_1+\rho_2+\rho_3)].\end{aligned} \tag{4.33}$$

The above expression can be easily seen and written in terms of the matrix demonstration as in Theorem 2.

□

From the above results, we can obtain the probability that an arbitrary node operates in the transmitting mode and receiving mode, which are given by the following Corollary. These probabilities are used in the next sections.

Corollary 1. *While the detailed analysis of each operating mode provides insight into the caching performance with respect to all possibilities, it is also interesting to quantify the caching performance by considering more general operating modes, such as transmitter and receiver and FD and HD modes. The new metrics of interest, namely, the probability of operating in a transmitter mode denoted by $\mathcal{P}^{\delta}_{\text{TX}}$, a receiver mode denoted by $\mathcal{P}^{\delta}_{\text{RX}}$, an FD mode denoted by $\mathcal{P}^{\delta}_{\text{FD}}$, and an HD mode denoted by $\mathcal{P}^{\delta}_{\text{HD}}$ are defined as follows:*

$$\mathcal{P}^{\delta}_{\text{HD}} = \mathcal{P}^{\delta}_{\text{SR}-\text{HDTX}} + \mathcal{P}^{\delta}_{\text{HDRX}} + \mathcal{P}^{\delta}_{\text{HDTX}}, \tag{4.34}$$

$$\mathcal{P}^{\delta}_{\text{FD}} = \mathcal{P}^{\delta}_{\text{FDTR}}, \tag{4.35}$$

$$\mathcal{P}^{\delta}_{\text{TX}} = \mathcal{P}^{\delta}_{\text{SR}-\text{HDTX}} + \mathcal{P}^{\delta}_{\text{HDTX}} + \mathcal{P}^{\delta}_{\text{FDTR}}, \tag{4.36}$$

$$\mathcal{P}^{\delta}_{\text{RX}} = \mathcal{P}^{\delta}_{\text{FDTR}} + \mathcal{P}^{\delta}_{\text{HDRX}}. \tag{4.37}$$

Proof. A transmitting node should operate either in SR-HDTX, HDTX, or FDTR mode, which means $\mathcal{P}^{\delta}_{\text{TX}} = \mathcal{P}^{\delta}_{\text{SR-HDTX}} + \mathcal{P}^{\delta}_{\text{HDTX}} + \mathcal{P}^{\delta}_{\text{FDTR}}$. Similar logic applies for the rest. □

Additionally, one can say:
$\mathcal{P}^{\delta}_{\text{SR}} + \mathcal{P}^{\delta}_{\text{SR}-\text{HDTX}} + \mathcal{P}^{\delta}_{\text{FDTR}} + \mathcal{P}^{\delta}_{\text{HDTX}} + \mathcal{P}^{\delta}_{\text{HDRX}} + \mathcal{P}^{\delta}_{\text{HO}} = 1.$

Remark 1. *In the stochastic caching policy, it is assumed that users follow the Zipf distribution for both the caching and the requesting, but with different skew exponents. The reasons to use the Zipf distribution can be summarized as follows: 1) The real-world analysis of the popularity of contents in the world's largest and well-known media sharing networks, such as YouTube, proves that the popularity of contents is very similar to the power-law Reiman Zita Function with an exponential cutoff, alternatively known as the Zipf distribution [81]. It is proven that the popularity of the content is determined by the frequency of the requests of the same content. Alternatively, the number of hits to the same content represents the popularity of that content despite the time characteristics (timing in caching and requesting) or characteristics of the users in terms of age, sex, etc. 2) Zipf distribution has extensively been used in the literature since 2012. The Zipf distribution not only has probabilistic features in nature but also allows us to distinguish between caching and requesting by considering different skew exponents and the number of contents used in the caching policies.*

4.3 Probability Mass Functions (PMFs) and Satisfied UEs

Technically, having knowledge on the number of transmitters and receivers in a wireless D2D network is useful to determine i) how much physical resources are needed to establish the D2D links, ii) how much traffic can be offloaded when the D2D links are initiated, and iii) how much throughput can be achieved on the network. Hence, in the sequel, we aim to obtain the probability mass function (PMF) of all operating modes along with the PMFs of transmitters and receivers.

4.3.1 PMFs

Here, we are interested in obtaining the PMFs of each operating mode. The number of nodes operating in each mode denoted by X_Δ is a binomial random variable, i.e., $X_\Delta \sim B(N, \mathcal{P}_\Delta^\delta)$, and its PMF is as follows:

$$\begin{aligned} f(N, n_\Delta, \mathcal{P}_\Delta^\delta) &= \Pr(X_\Delta = n_\Delta) \\ &= \binom{N}{n_\Delta} (\mathcal{P}_\Delta^\delta)^{n_\Delta} \left(1 - \mathcal{P}_\Delta^\delta\right)^{N-n_\Delta}, \end{aligned} \tag{4.38}$$

where $\binom{N}{n_\Delta} = \frac{N!}{(N-n_\Delta)!n_\Delta!}$.

Proof. The probability of obtaining exactly n_Δ (interpreted as the number of nodes operating in mode Δ) succeeds in N independent Bernoulli trials, leading to the binomial distribution. In fact, n_Δ successes occur with probability $\mathcal{P}_\Delta^\delta$ and $N - n_\Delta$ failures occur with probability $(1 - \mathcal{P}_\Delta^\delta)^{N-n_\Delta}$. However, it is possible to have n_Δ successes out of N trials, where there are $\begin{pmatrix} N \\ n_\Delta \end{pmatrix}$ different ways of distributing n_Δ in a sequence of N trials. □

4.3.2 Expected Number of Satisfied Users

Using the functions obtained in the previous subsections, the closed-form expressions of the expected number of satisfied users in the FD-enabled network can be derived. To obtain that measure, we need to seek the number of nodes that appear in the specific modes that allow the fetching of the requested content. From the observation of Fig. 3.2, we realize that modes SR, SR-HDTX, FDTR, and HDRX are the ones where a user is able to fetch its desired content. It is obvious that multiplying the total number of nodes N by the probability of each node gives us the number of nodes in that mode. Hence, denoting $\mathcal{N}_{\text{sat}}^\delta$ as the expected number of satisfied users with a caching policy δ, we have the equation as follows:

$$\mathcal{N}_{\text{sat}}^\delta = N \left(\mathcal{P}_{\text{SR}} + \mathcal{P}_{\text{SR-HDTX}} + \mathcal{P}_{\text{FDTR}} + \mathcal{P}_{\text{HDRX}} \right). \tag{4.39}$$

CHAPTER 5 NUMERICAL RESULTS

We concentrate our analysis on three types of results. First, we analyze the probabilities of encountering a particular mode. Second, we analyze the behavior of the PMFs with respect to the different system parameters and, finally, the expected number of satisfied users in the system.

5.1 Analysis of mode probability

5.1.1 By type of mode

This subsection contains the results obtained to find the probability behavior of each type of mode when the number of users increases, for the two caching policies.

Fig. 5.1 illustrates the probability of each operating mode versus the number of user devices N when the *deterministic caching* policy is employed. These results are delineated for a typical value of the skew exponent $\gamma_r = 0.8$, while a broad range of values are being utilized in future results, which we will discuss in the sequel. As can be observed in Fig. 5.1, there is a clear distinction in the probability behavior between the group of modes SR, SR-HDTX, HO and HDTX and the group FDTR and HDRX. When increasing the number of user devices, the probabilities of the former group tend to zero. In the first group, the highest chance occurs when N is very low for two main reasons: i) the lower number of users leads to less diversity in available cached contents; correspondingly, the highest rank of the contents is of more interest to a very low number of users; ii) the very low number of users eliminates the chance to form other operating modes since there are a few users in the vicinity. In contrast, the chance of forming an operating mode HDRX is increasing since there is more diversity in the cached contents, while the chance of FDTR and HDTX decay after reaching the maximum peak for the values of N less than 50. The HDRX mode is the dominating operating mode among all modes when N increases, and more users obtain the chance to receive their desired contents. Fig. 5.2 and Fig. 5.3 demonstrate the results when the *stochastic caching* policy is being employed. The difference between these two figures is the number of users N. Due to high complexity order of the equations as follows: (4.15), (4.16) (4.17), (4.19), (4.18), and (4.20), Fig. 5.2 is delineated with limited values of N and M, while Fig. 5.3 demonstrates the simulation results with higher values of N and M. The overall behavior of the *stochastic caching* policy is similar to *deterministic caching*, there are key differences between them as follows. First, the HO mode is more accented in *stochastic caching*, which means that in this realistic caching policy, the outage probability is higher than that of the case when the *deterministic caching* policy is being employed. Another key difference is the major difference between both caching policies with respect to the FDTR mode. As can be observed, in stochastic caching, the probability of the FDTR mode always increases with N. The main reason comes from the high chance of overlapping cached contents when the *stochastic caching* policy is being employed. In other words, because of the overlap, there is a high chance of finding users' desired content in their vicinity, and because we consider FD transmission, many potential delivery opportunities are created.

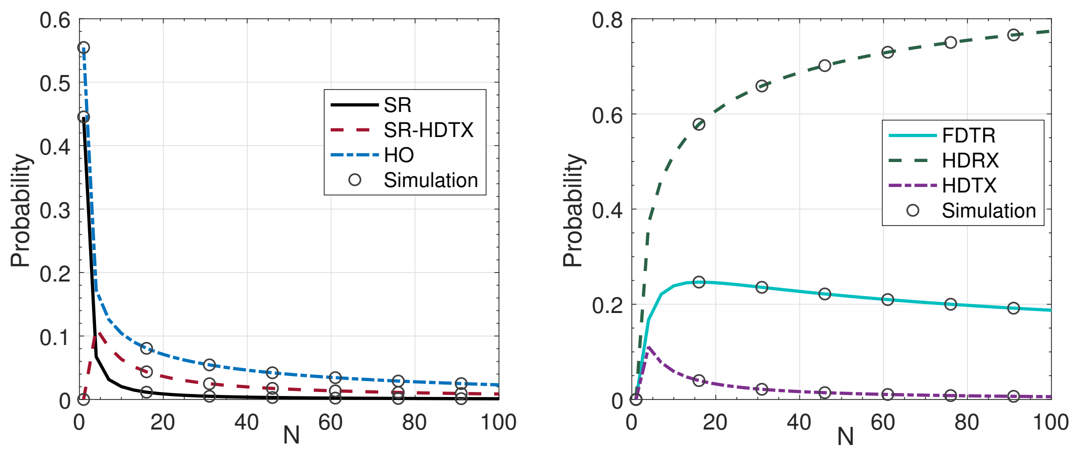

Figure 5.1 Deterministic caching with $M = 500$, $\gamma_r = 0.8$

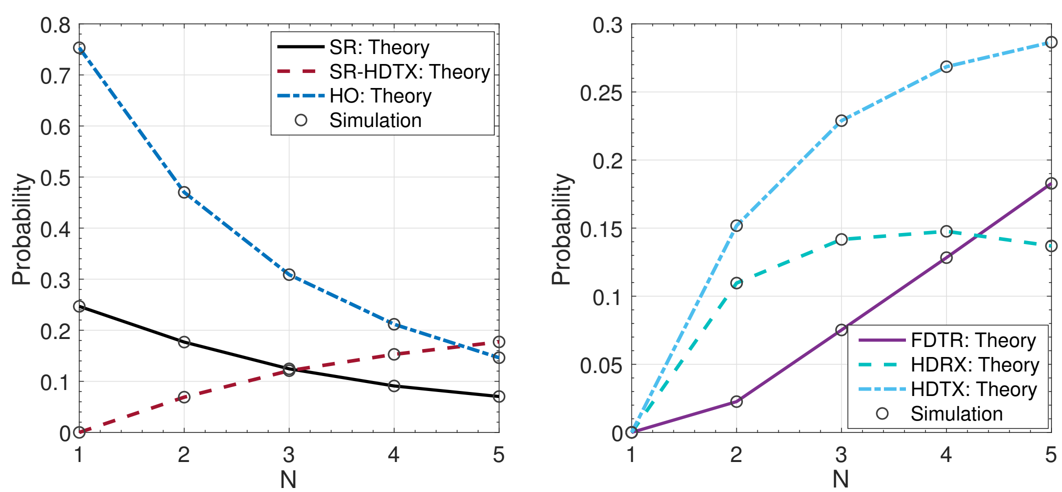

Figure 5.2 Limited results of the stochastic caching with $M = 7$, $\gamma_r = 0.8$, $\gamma_c = 1.6$

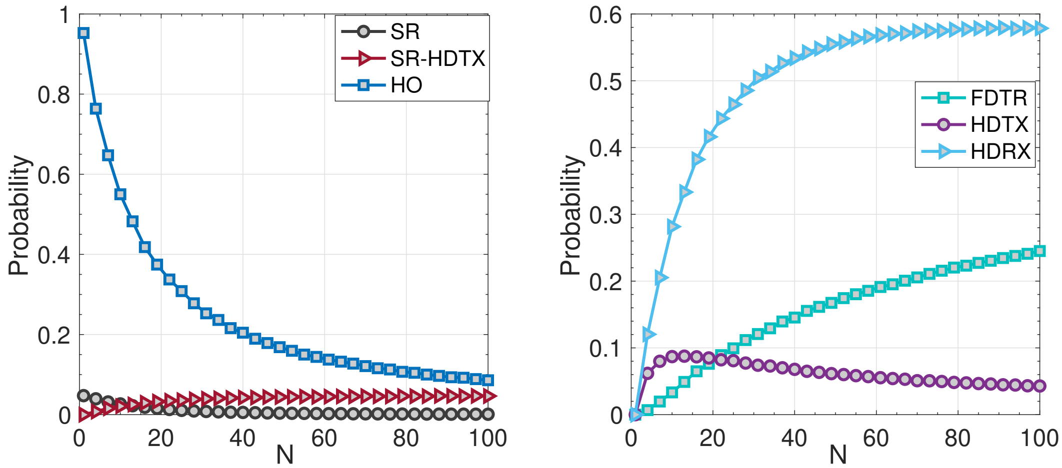

Figure 5.3 Stochastic caching with $M = 500$, $\gamma_r = 0.8$, $\gamma_c = 1.6$

5.1.2 By type of caching features

In this subsection, we want to provide insights into how the mode probabilities change with respect to the type of caching, but also with respect to the popularity index and library size. To perform the analysis, we regroup the modes in a group called HD and FD or TX and RX, as defined in Corollary 1. Fig. 5.4 portrays the different curves obtained using Eqs.(4.34), (4.35), (4.36), (4.37). This figure includes i) a comparison of FD and HD modes in both caching policies labeled by HD-δ and FD-δ (figures in the first column); ii) a comparison of operating in transmitter and receiver modes labeled by TX-δ and RX-δ, respectively (figures in the second column); and iii) a comparison of SR and HO modes in both caching policies labeled by SR-δ and HO-δ (figures in the third column). The simulation parameters used for this figure are summarized in Table 5.1. The values of the simulation parameters are being utilized in the existing literature, hence, the simulations in this thesis is conducted based on well-known range of parameters. According to Fig. 5.4, the probability of the HD mode is

Table 5.1 Simulation Parameters in Fig. 5.4

$\mathcal{P}_{\Delta}^{\delta}$ vs.	Parameters		
	N	M	γ_c
γ_r	100	10^4	1.6
	N	M	γ_r
γ_c	100	10^4	2.5
	N	γ_r	γ_c
M	50	0.8	1.6

higher than that of the FD mode with respect to all key parameters γ_r, γ_c, and M. The higher the values of γ_r and γ_c are, the higher the redundancy in requests and cached contents. In other words, the higher the γ_r is, the higher the chance of requesting highly ranked contents, and similarly, the higher the γ_c is, the higher the chance of caching high ranked contents. For the size of the library M, the higher M is, the higher the diversity in the popularity of the contents. Increasing M has a minor impact on the probabilities of the operating modes in comparison with the impact of parameters γ_r and γ_c. Similar explanations are valid for the TX and RX modes.

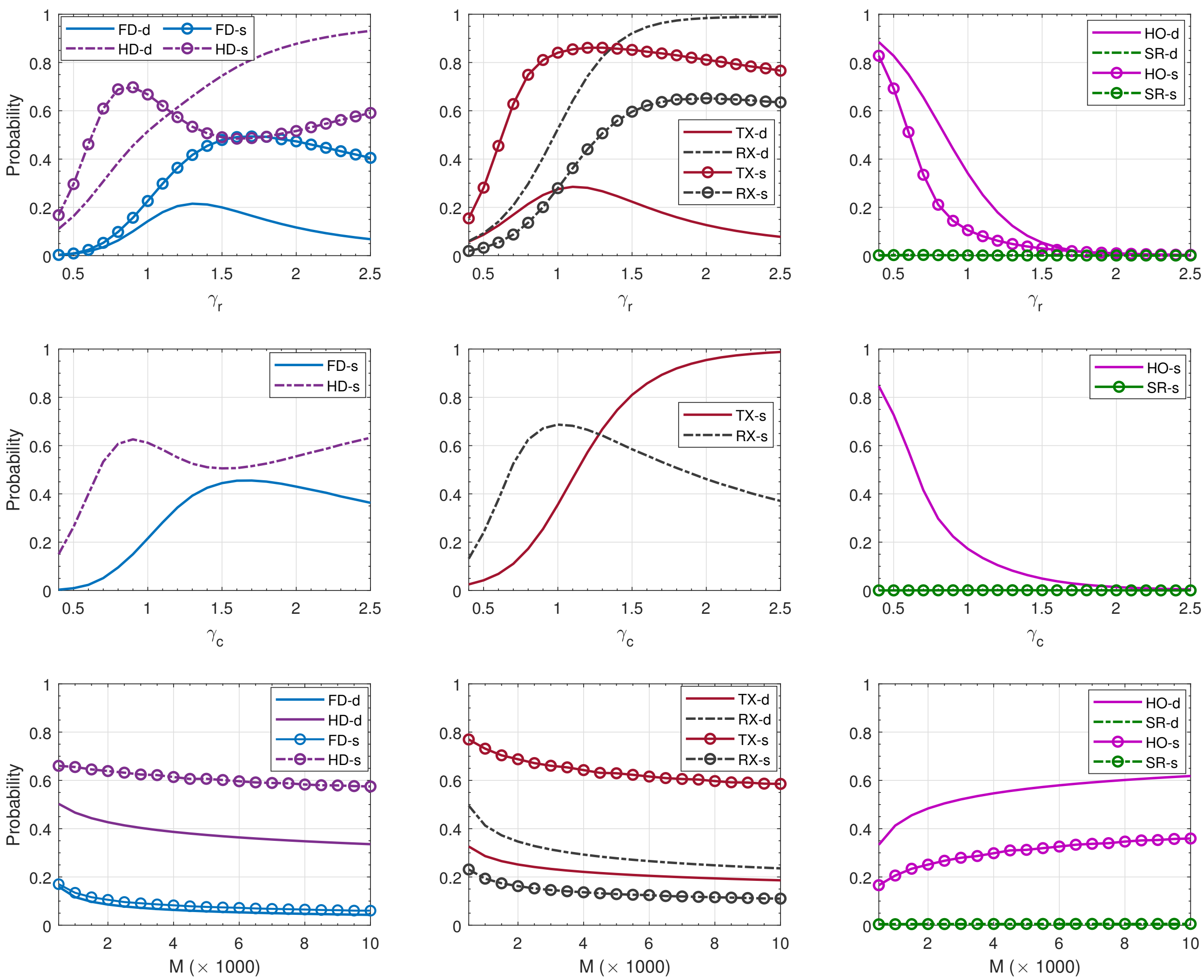

Figure 5.4 Comparison of probabilities versus γ_r , γ_c, and M

5.1.3 PMF analysis

Fig. 5.5 illustrates the PMF of the operating modes. For the sake of visibility and presentation, all results are shown as continuous curves, while the original values are discrete. In this figure, the set of subfigures in the first row demonstrates the comparison between *deterministic caching* and *stochastic caching* policies, while the second and third rows illustrate the PMFs in *stochastic caching* with respect to skew exponent γ_c and library size M, respectively. As seen from this figure, the expected number of HD users is more than that of FD users for all cases. Additionally, the expected number of users operating in a *deterministic caching* policy is higher than that of FD users. The reason for this can be interpreted from the results in Fig. 5.1 and Fig. 5.3. Since the probability of HD mode is higher than that of the FD modes, the same explanations of those figures are valid here. Furthermore, higher redundancy in the cached contents increases the number of collaborating HD and FD users, i.e., caching performance with $\gamma_c = 1.6$ provides more nodes in content exchange in comparison with the case when $\gamma_c = 0.8$. For the TX and RX modes, the role of the skew exponent γ_c is the same as the role in the HD and FD modes. Furthermore, increasing the library size M yields a reduction in the number of nodes operating in the HD and FD and TX and RX modes, conversely increasing the number of HO and SR modes.

5.1.4 Expected number of satisfied users

As previously mentioned, a measure of QoE for the user of the system is the probability of the user finding their content. Now, from the operator standpoint, the metric should be the expected number of satisfied users. In Fig. 5.6, we illustrate how this metric changes with the number of users, the popularity of the cached and requested content, and the size of the content library when a deterministic or a stochastic caching policy is used. The number of satisfied users is clearly dictated by the content redundancy exponents (γ_r and γ_c). For instance, given a fixed and lower value of the request skew exponent γ_r (lower redundancy in requests), lower values of the caching skew exponent γ_c (lower redundancy in caching) provide less satisfaction in the FD-enabled network in comparison with the case where the higher redundancy is being employed by the caching skew exponent. The reason for this phenomenon is that more users are requesting highly ranked contents when γ_c is high, which increases the chance of both FD and multicast transmission opportunities occurring. A similar observation can be made from the figure where the caching exponent γ_c varies with the two different numbers of users and a fixed request exponent γ_r. In this figure, when γ_c increases, users are caching more of highly ranked content, while they are requesting lower ranked content. However, for the values between $\gamma_c = 0.8$ and $\gamma_c = 1$, we observe the

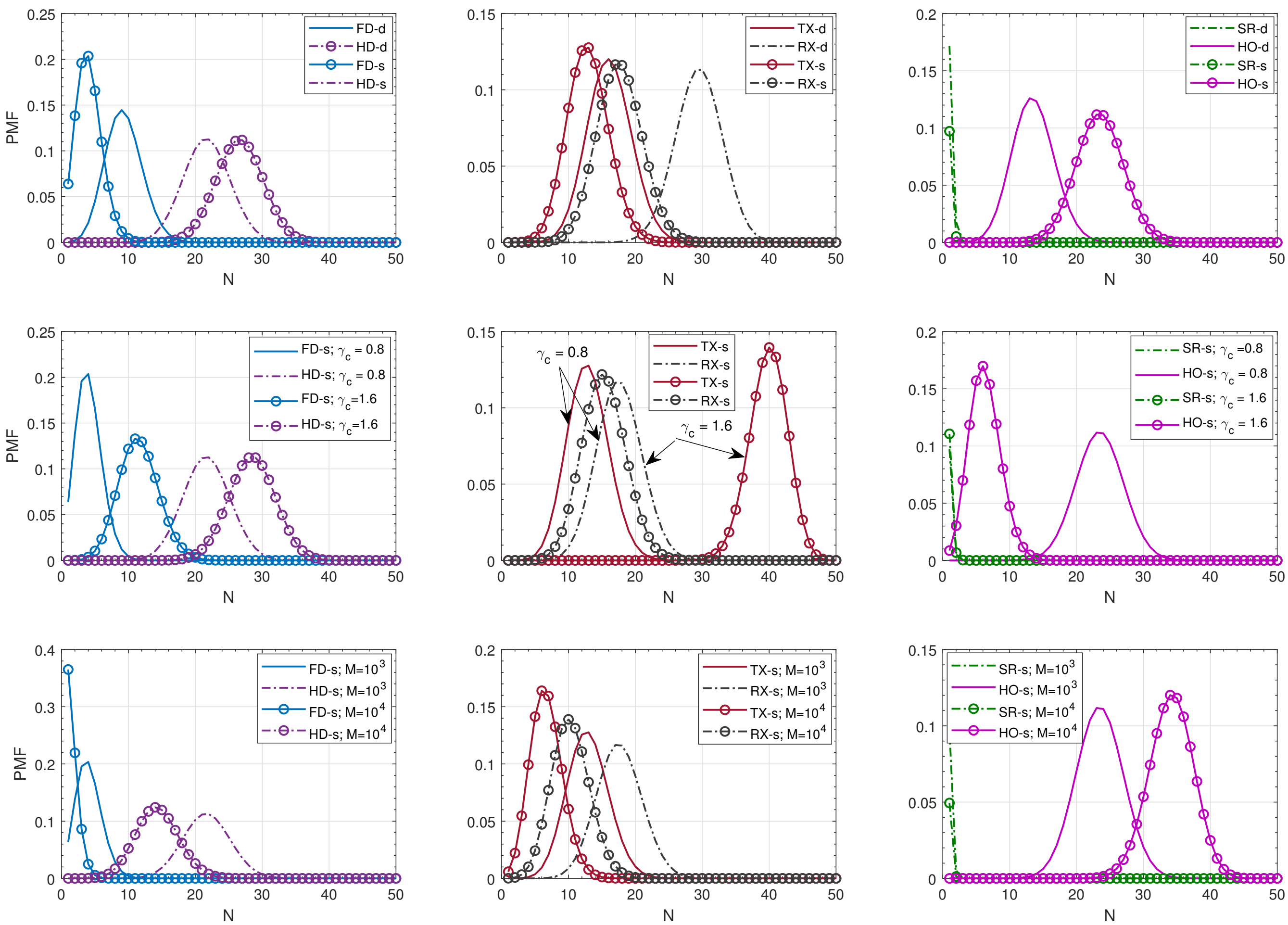

Figure 5.5 Comparison of PMFs with respect to γ_c and M

highest number of satisfaction, which is because of the lower difference between the caching exponent and the request exponent, meaning that the cached and requested content has the same ranking. Similar explanations of the impact of diversity in library (M) in deterministic caching are valid for the stochastic caching policy.

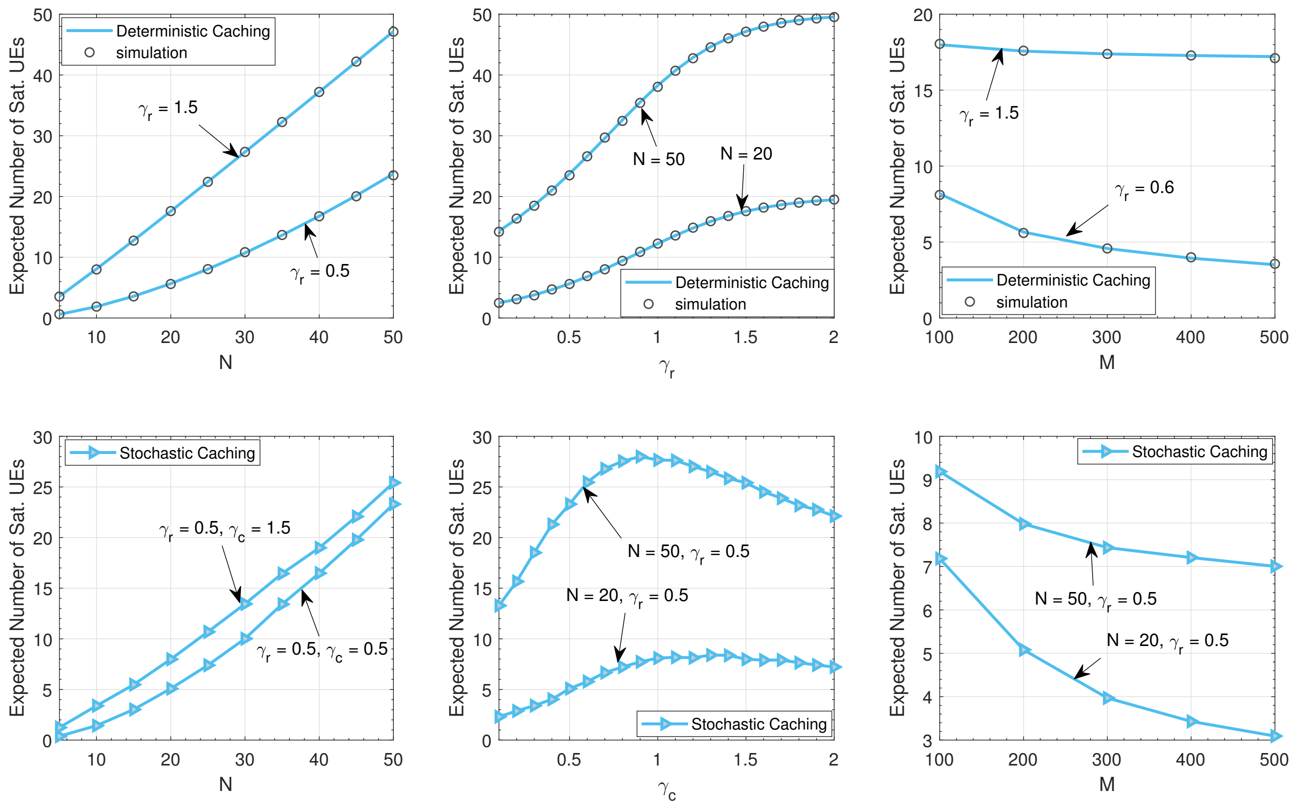

Figure 5.6 Expected number of satisfied users for both deterministic and stochastic caching policies

CHAPTER 6 CONCLUSION AND RECOMMENDATIONS

In this thesis, a thorough study of the probability of occurrence of different modes in a cache-enabled FD-D2D network is being proposed. The proposed analysis takes into account a hypothetical random graph which includes all possible modes. Two different caching policies are being exploited in the framework, which are deterministic and stochastic caching policies. The probability of all operating modes are being analyzed with respect to both caching policies and numerical results drawn with respect to all involving system parameters such as number of users in the network and popularity distribution parameters. General probability rules are being employed to derived the closed-form probability expressions. Then, the obtained probabilities are being utilized to obtain the probability mass functions (PMFs) for all operating modes. Numerical results for PMFs are drawn with respect to all involving parameters by considering both caching policies. Further, the obtained probabilities are being used to obtain the expected number of satisfied users in network. The satisfied users are obtained based on the modes that there is at least one incoming link to the typical user, namely, when the typical user operates in Self-Request, HDRX, and FDTR modes. Therefore, the expected number of satisfied users changes with these probabilities.

6.1 Some Insights on the Proposed Analysis

It has been observed that the probability of occurrence of different modes strongly tied to caching characteristics such as popularity of contents, redundancy in requests, and number of users. In particular:

- The probability of the event that a user could fetch its desired content via its own cache (i.e., Self-Request mode) is very low unless the number of users is very low. This mode is interesting since it delivers the content of interest immediately without the need for any transmission, however, numerical results revealed that with deterministic caching policy, and only when the number of users is low and the redundancy in the requests is high, the probability of this mode has the highest possible probability.
- Cache hit outage is being considered as a separate mode. It is shown that when there are few users in the network, this mode has its highest possible probability, namely, it is more likely to fail to retrieve desired content since there is a few cached contents in the vicinity. The last resort for the users that appear in this mode is to retrieve their desired content through traditional cellular downlink. Similar to self-request mode, the

PMF of this mode is dictated by the caching policy characteristics. In summary, to have a less outage in system, alongside having more users in the vicinity, high redundancy in caching and requests is needed and this is valid for both caching policies.

- It is shown that the FD capability brings more content delivery opportunities to the D2D caching network. This capability increases the expected number of satisfied users when nodes appear to operate in reception and transmission at the same time. The results revealed that even when the number of FD nodes is low, due to multicast transmission in the transmission side, still FD nodes provide more opportunities to the potential receivers. Even though the FD capability potentially doubles the number of expected users, providing that both users find their desired content in each other, however, due to multicast opportunities in the transmission side, the benefit of the FD capability can be magnified since multiple users can receive their desired content by a single transmission from the FD node. It is shown that having more users in the network leads to more content delivery opportunities empowered by FD nodes, namely, the number of FD nodes increases with the number of users.

- The numerical results for the expected number of satisfied users revealed that it seems there is an optimal value for the caching exponent when stochastic caching policy is employed. Namely, it is observed that for the values around $\gamma_c \approx [0.8 \quad 1]$, the number of satisfied users has its highest possible value. Indeed the peak point might change by changing the number of users of skew exponent in the requests, however, in the simulation results, typical values of request skew exponent are being utilized as observed by many studies in the past. Therefore, it can be concluded that from the design perspective to employ stochastic caching policy, these values are good candidates to maximize the expected number of satisfied users.

- The investigation on the impact of diversity in the contents, namely the library size, also revealed interesting results. It is shown that the higher diversity increases the cache hit outage but has almost no impact on self-request probability. Also, higher diversity slightly decreases both FD and HD operation probabilities. However, when diversity comes with high redundancy in requests and cached contents (higher skew exponents), considerable change has been observed in the expected number of satisfied users. In summary, to reverse the negative impact of diversity in the expected number of satisfied users, higher skew exponents are the key factors to tune.

There are some interesting directions that can be further taken to utilize or customize the proposed analysis. In the sequel, some of the potential directions is discussed.

6.2 Potential Research Directions

6.2.1 Physical Layer Investigations

In this thesis, the analysis and simulations are conducted based on key assumptions that do not take the physical layer aspects into account. By integrating the physical layer aspects of the problem in the proposed framework, there are some open problems that need to be investigated as an extension to the research proposed in this thesis. Activating a set of transmitting nodes will create mutual interference which can make affect the links capacity, amount of offloaded traffic, and overall throughput. Here are some potential extensions:

- Channel Modeling: In this thesis, it is assumed that the D2D links are always available to establish the content deliveries between users. However, in the real-world scenarios, channel conditions between transmitter and receiver plays a crucial role in D2D links. Therefore, channel characteristics of the D2D links need to be considered to investigate the quality of D2D links.

- Full-Duplex Impairments: The main impairment for FD radios is experiencing self-interference (SI) at the transceiver. For the FD-D2D nodes, SI need to be considered to investigate more realistic link capacity since the SI can degrades the link quality by increasing the amount of interference imposed on the transceiver.

6.2.2 Scheduling Problem

Due to overlap in stochastic caching, it is possible that one user can retrieve its desired content through multiple candidates in the vicinity. In that case, it is interesting to investigate the problem of choosing the appropriate link, which leads to some system target KPIs. If a typical user can receive its desired content from two nodes such that one of them appear in HD transmission mode and the other one appear in the FD mode, choosing the first option can eliminate SI impact on the FD counterpart in the latter case. Therefore, an appropriate scheduling strategy need to be designed to tackle this problem. The scheduling problem is also important from the physical layer perspective. Technically, the amount of resources to schedule the D2D links can be affected by the number of transmitters and receivers, hence, it is important to have an efficient scheduling algorithm for the D2D links to meet certain constraints in the network.

6.2.3 Satisfying Cache Hit Outage Mode

As discussed in this thesis, it is possible that users cannot find their desired content within the network and at the same time, there is no potential user that demands the content cached by these users. In this mode, these users are completely out of D2D collaborations and their demand are not being satisfied. These users need to obtain their desired content through the traditional downlink. Therefore, it is interesting to consider a cellular Heterogeneous Network (HetNet) in which the outage users can obtain their desired content either through small base stations (SBS) or macro base stations (MBS). Hence, given the probability of the typical user that appear in the outage mode, it is worth to study a traditional mobile network in which the users are extracted based on the obtained probabilities. The possible research approach can be integrated into a more comprehensive user satisfaction metric, in which the probability of satisfying a typical user is combination of FD-D2D opportunities and cellular downlink.

6.2.4 Security Aspects

Content deliveries between D2D links can suffer from potential security threads. One of the well-known threats in such networks are Denial of Service (DoS) attacks that can block the links and avoid efficient content deliveries in the network. Hence, it is important to integrate security mechanisms in the D2D links to secure content deliveries among users.

REFERENCES

[1] "Cisco Annual Internet Report (2018-2023) White Paper," https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, accessed: May 3, 2022.

[2] J. Yao, T. Han, and N. Ansari, "On mobile edge caching," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2525–2553, 2019.

[3] N. Golrezaei *et al.*, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, 2014.

[4] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 176–189, 2016.

[5] Y. Adachi *et al.*, "Cloud-assisted dynamic content sharing among vehicles," in *2016 IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, 2016, pp. 516–523.

[6] V. S. Varanasi and S. Chilukuri, "Adaptive differentiated edge caching with machine learning for v2x communication," in *International Conference on Communication Systems & Networks (COMSNETS)*, 2019, pp. 481–484.

[7] R. Lu *et al.*, "UAV-assisted data dissemination with proactive caching and file sharing in v2x networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[8] K. Liu *et al.*, "Network-coding-assisted data dissemination via cooperative vehicle-to-vehicle/-infrastructure communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1509–1520, 2015.

[9] D. Grewe *et al.*, "Adept: Adaptive distributed content prefetching for information-centric connected vehicles," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–5.

[10] X. Han *et al.*, "Incentive mechanism with the caching strategy for content sharing in vehicular networks," in *2019 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2019, pp. 1–6.

[11] K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 3025–3041, 2019.

[12] D. Korpi *et al.*, "Full-duplex mobile device: Pushing the limits," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 80–87, 2016.

[13] S. K. Sharma *et al.*, "Dynamic spectrum sharing in 5G wireless networks with full-duplex technology: Recent advances and research challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 674–707, 2018.

[14] K. S. Ali, H. ElSawy, and M. Alouini, "Modeling cellular networks with full-duplex d2d communication: A stochastic geometry approach," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4409–4424, 2016.

[15] L. Marandi *et al.*, "Delay analysis in full-duplex heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 9713–9721, 2019.

[16] G. Liu *et al.*, "Cooperative noma broadcasting/multicasting for low-latency and high-reliability 5G cellular v2x communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7828–7838, 2019.

[17] B. Wang *et al.*, "Joint precoding and user scheduling for full-duplex cooperative mimo-noma v2x networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–6.

[18] M. Yang, S. Jeon, and D. K. Kim, "Interference management for in-band full-duplex vehicular access networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1820–1824, 2018.

[19] A. Bazzi *et al.*, "Enhancing cooperative driving in ieee 802.11 vehicular networks through full-duplex radios," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2402–2416, 2018.

[20] J. Zang, V. Towhidlou, and M. Shikh-Bahaei, "Collision avoidance in v2x communication networks," in *2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*, 2019, pp. 1–6.

[21] D. Zhang *et al.*, "Performance analysis of fd-noma-based decentralized v2x systems," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5024–5036, 2019.

[22] M. Naslcheraghi, S. A. Ghorashi, and M. Shikh-Bahaei, “FD device-to-device communication for wireless video distribution,” *IET Communications*, vol. 11, no. 7, pp. 1074 – 1081, January 2017.

[23] M. Naslcheraghi, S. A. Ghorashi, and M. Shikh-Bahaei, “Full-duplex device-to-device collaboration for low-latency wireless video distribution,” in *2017 24th International Conference on Telecommunications (ICT)*, 2017, pp. 1–5.

[24] K. T. Hemachandra, O. Ochia, and A. O. Fapojuwo, “Performance study on cache enabled full-duplex device-to-device networks,” in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.

[25] M. Naslcheraghi, S. A. Ghorashi, and M. Shikh-Bahaei, “Performance analysis of inband fd-d2d communications with imperfect si cancellation for wireless video distribution,” in *2017 8th International Conference on the Network of the Future (NOF)*, 2017, pp. 176–181.

[26] H. Qu *et al.*, “Performance analysis of the content dissemination mechanism with proactive content fetching and full-duplex d2d communication: an evolutionary perspective,” *Mobile Networks and Applications*, vol. 24, no. 2, pp. 532–555, 2019.

[27] R. Q. Hu, Y. Qian *et al.*, “D2d communications in heterogeneous networks with full-duplex relays and edge caching,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4557–4567, 2018.

[28] R. Q. Hu, L. Hanzo *et al.*, “Heterogeneous networks relying on full-duplex relays and mobility-aware probabilistic caching,” *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5037–5052, 2019.

[29] T. X. Vu *et al.*, “Linear precoding design for cache-aided full-duplex networks,” in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.

[30] T. X. Vu *et al.*, “Full-duplex enabled mobile edge caching: From distributed to cooperative caching,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 1141–1153, 2019.

[31] T. X. Vu *et al.*, “Cache-aided full duplex: Delivery time analysis and optimization,” *IEEE Wireless Communication Letters*, 2018.

[32] M. Naslcheraghi, M. Afshang, and H. S. Dhillon, "Modeling and performance analysis of full-duplex communications in cache-enabled d2d networks," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[33] M. Haenggi, *Stochastic geometry for wireless networks.* Cambridge University Press, 2012.

[34] 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, 2015.

[35] Z. Chen and M. Kountouris, "D2d caching vs. small cell caching: Where to cache content in a wireless network?" in *2016 IEEE 17th international workshop on signal processing advances in wireless communications (SPAWC).* IEEE, 2016, pp. 1–6.

[36] M. Rim and C. G. Kang, "Cache partitioning and caching strategies for device-to-device caching systems," *IEEE Access*, vol. 9, pp. 8192–8211, 2021.

[37] A. Ibrahim, A. Zewail, and A. Yener, "Device-to-device coded caching with distinct cache sizes, submitted to ieee trans. commun," *arXiv preprint arXiv:1903.08142*, 2019.

[38] R. Sun *et al.*, "Cost-oriented mobility-aware caching strategies in d2d networks with delay constraint," *IEEE Access*, vol. 7, pp. 177 023–177 034, 2019.

[39] N. Deng and M. Haenggi, "The benefits of hybrid caching in gauss–poisson d2d networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1217–1230, 2018.

[40] M. Rim and C. G. Kang, "Content prefetching of mobile caching devices in cooperative d2d communication systems," *IEEE Access*, vol. 8, pp. 141 331–141 341, 2020.

[41] Z. Ma *et al.*, "Deployment model and performance analysis of clustered d2d caching networks under cluster-centric caching strategy," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4933–4945, 2020.

[42] M.-C. Lee, H. Feng, and A. F. Molisch, "Dynamic caching content replacement in base station assisted wireless d2d caching networks," *IEEE Access*, vol. 8, pp. 33 909–33 925, 2020.

[43] G. Kollias and A. Antonopoulos, "Joint consideration of content popularity and size in device-to-device caching scenarios," in *ICC 2020-2020 IEEE International Conference on Communications (ICC).* IEEE, 2020, pp. 1–6.

[44] Q. Li *et al.*, "D2d-assisted caching on truncated zipf distribution," *IEEE Access*, vol. 7, pp. 13 411–13 421, 2019.

[45] H. Nikbakht *et al.*, "Stochastic d2d caching with energy harvesting nodes," in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2020, pp. 1–8.

[46] X. Zhang *et al.*, "Analysis and optimization of service delay for multiquality videos in multitier heterogeneous network with random caching," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2451–2462, 2021.

[47] 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, 2016.

[48] J. Ji *et al.*, "Probabilistic cache placement in uav-assisted networks with d2d connections: performance analysis and trajectory optimization," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6331–6345, 2020.

[49] R. M. Amer *et al.*, "Performance analysis and optimization of cache-assisted comp for clustered d2d networks," *IEEE Transactions on Mobile Computing*, 2020.

[50] J. Xu *et al.*, "Probabilistic model based caching strategy for device-to-device communications," in *2021 IEEE/CIC International Conference on Communications in China (ICCC)*, 2021, pp. 1001–1005.

[51] E. Baccour *et al.*, "Ce-d2d: Dual framework chunks caching and offloading in collaborative edge networks with d2d communication," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2019, pp. 1550–1556.

[52] Y. Fu *et al.*, "Caching efficiency maximization for device-to-device communication networks: A recommend to cache approach," *IEEE Transactions on Wireless Communications*, 2021.

[53] X. Huang, G. Zhao, and Z. Chen, "Segment-based random caching in device-to-device (d2d) caching networks," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2015, pp. 731–735.

[54] K. Zhu *et al.*, "Social-aware incentivized caching for d2d communications," *IEEE Access*, vol. 4, pp. 7585–7593, 2016.

[55] N. Zhao *et al.*, "Caching d2d connections in small-cell networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12 326–12 338, 2018.

[56] M.-C. Lee, H. Feng, and A. F. Molisch, "Dynamic caching content replacement in base station assisted wireless d2d caching networks," *IEEE Access*, vol. 8, pp. 33 909–33 925, 2020.

[57] M.-C. Lee and A. F. Molisch, "Individual preference aware caching policy design in wireless d2d networks," *IEEE transactions on wireless communications*, vol. 19, no. 8, pp. 5589–5604, 2020.

[58] J. Rao *et al.*, "Optimal caching placement for d2d assisted wireless caching networks," in *2016 IEEE international conference on communications (ICC)*. IEEE, 2016, pp. 1–6.

[59] A. Ahmed, H. Shan, and A. Huang, "Modeling the delivery of coded packets in d2d mobile caching networks," *IEEE Access*, vol. 7, pp. 20 091–20 105, 2019.

[60] R. Karasik, O. Simeone, and S. S. Shitz, "How much can d2d communication reduce content delivery latency in fog networks with edge caching?" *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2308–2323, 2019.

[61] M. Peer, V. A. Bohara, and A. Srivastava, "Cache selection in dynamic d2d multicast networks using inhomogeneous markov model," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3235–3245, 2020.

[62] J. Ma *et al.*, "Scalable video transmission in cache-aided device-to-device networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4247–4261, 2020.

[63] L. Zhong *et al.*, "Adaptive delay optimization of multicast-enabled coded caching in device-to-device networks," *China Communications*, vol. 17, no. 1, pp. 42–48, 2020.

[64] Q. Fu *et al.*, "Extensive cooperative content caching and delivery scheme based on multicast for d2d-enabled hetnets," *IEEE Access*, vol. 9, pp. 40 884–40 902, 2021.

[65] Q. Li *et al.*, "Analysis of a cooperative caching-multicast strategy in d2d-aided networks," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2614–2618, 2021.

[66] T. Fang *et al.*, "Content delivery in edge caching networks: A hypergraph game-theoretic approach," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 12 248–12 252, 2021.

[67] S. Kumar and S. Misra, "Enabling multi-source device-to-device content delivery in cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10 853–10 863, 2021.

[68] S. A. Taheri and M. Rasti, "Caching placement and offloading in d2d-assisted wireless networks with in-band full duplex," in *2020 28th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2020, pp. 1–5.

[69] S. Pang *et al.*, "Optimal target user selection policy for d2d wireless caching networks," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2665–2678, 2021.

[70] M.-C. Lee, M. Ji, and A. F. Molisch, "Optimal throughput-outage analysis of cache-aided wireless multi-hop d2d networks," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2489–2504, 2020.

[71] J. Ji *et al.*, "Joint trajectory design and resource allocation for secure transmission in cache-enabled uav-relaying networks with d2d communications," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1557–1571, 2020.

[72] M.-C. Lee, A. F. Molisch, and M. Ji, "Throughput-outage scaling laws for wireless single-hop d2d caching networks with physical models," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[73] Q.-N. Tran *et al.*, "Spectrum sharing and power allocation optimised multi-hop multi-path D2D video delivery in beyond 5G networks," *IEEE Transactions on Cognitive Communications and Networking*, 2021.

[74] K. Z. Shen and D. K. So, "Power allocation for d2d noma in cache-aided networks," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–6.

[75] K. Z. Shen, T. E. A. Alharbi, and D. K. C. So, "Cache-aided device-to-device non-orthogonal multiple access," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–6.

[76] Y. Hui *et al.*, "Content in motion: An edge computing based relay scheme for content dissemination in urban vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3115–3128, 2018.

[77] E.-s. Kim, S.-y. Park, and J.-i. Jung, "Performance comparison of the multi-hop broadcast schemes for safety message in v2v communication," in *2009 IEEE International*

Conference on Network Infrastructure and Digital Content. IEEE, 2009, pp. 1023–1028.

[78] W. Drira and F. Filali, "Ndn-q: An ndn query mechanism for efficient v2x data collection," in *2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking Workshops (SECON Workshops)*. IEEE, 2014, pp. 13–18.

[79] E. Bastug, 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, 2014.

[80] Y. Chen *et al.*, "Cooperative full duplex content sensing and delivery improves the offloading probability of d2d caching," *IEEE Access*, vol. 7, pp. 29 076–29 084, 2019.

[81] M. Cha *et al.*, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, pp. 1–14.