Performance Analysis of Dynamic Downlink Cellular Networks

Qiong Liu

THESE DE DOCTORAT DE

L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES RENNES

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : Télécommunications

Par

Qiong LIU

« Performance Analysis of Dynamic Downlink Cellular Networks »

Thèse présentée et soutenue à « Rennes », le « 03/06/2022 »
Unité de recherche : Institut d'electronique et des technologies du numérique (IETR)
Thèse N° : 22ISAR 14 / D22 - 14

Rapporteurs avant soutenance :

Marco DI RENZO Directeur de recherche, CNRS, L2S, CentraleSupelec, Universite Paris-Sacaly
Lina MROUEH Professeure, Institut Supérieur d'Electronique de Paris

Composition du Jury :

Président :	Laurent CLAVIER	Professeur, IMT Nord-Europe
Examinateurs :	Marco DI RENZO	Directeur de recherche, CNRS, L2S, CentraleSupelec
	Lina MROUEH	Professeure, Institut Supérieur d'Electronique de Paris
	Marceau COUPECHOUX	Professeur, Telecom Paris
Co-dir. de thèse :	Philippe Mary	Maitre de conferences, HDR, INSA de Rennes
	Jean-Yves Baudais	Chargé de recherche, HDR, CNRS

Performance analysis of dynamic downlink cellular networks

Qiong Liu

Acknowledgements

First, I am grateful to my supervisor, Dr. Philippe Mary, who provided me with precious opportunities and constant support during my Ph.D. There are many things Philippe has taught me, but nothing was more precious than his enthusiasm to push the boundaries of scientific knowledge. Without his motivation, this study undoubtedly would not have had excellent fruition.

I owe my deepest gratitude to my supervisor, Dr. Jean-Yves Baudais, for sharing his outstanding expertise and providing precious advice in mathematics and telecommunication. He nurtured my passion for pursuing this degree and encouraged me to embark on the path of academics. I especially appreciate his gentlemanliness, which makes people feel like in a spring breeze. Besides, he encouraged me to study French, which opened a new door to understanding the world.

My sincere thanks to Prof. Laurent Clavier for presiding over my Ph.D. defense. All my gratitude to Prof. Marco Di Renzo and Prof. Lina Mroueh for their time reading and commenting on my thesis. I also thank Prof. Marceau Coupechoux be my jury number.

I want to thank my colleague Dr. Mohammadreza Mardani Varmazyar. Thanks to our similar research directions, we discussed many academic details and exchanged many opinions on coding and formulating. Besides, I would also like to thank my colleague, Mrs. Yara Yaacoub (who will be a doctor very soon), an excellent researcher. She kindly helped me with many complex issues that needed to be solved in French, and we had a very enjoyable working period.

Of course, thanks must go to my parents and my little brother. My parents do not speak English and have no idea what telecommunication is, but they tried their best to support and encourage me to do what I wanted. Without their constant support, I would never have had a chance to start and finish a Ph.D.

Finally, I want to thank my Fiance, Shaojie. He changed his job from China to France to support me. He created a comfortable environment allowing me to focus on my work without any worries about housework and finance. He is always the first man to hear my presentations, although he is not a specialist. It would have been impossible for me to go through the last three years without any regrets.

These three and half years are my golden age. Several people have helped and collaborated with me in various ways. It is never until the end that one realizes that science has a broader field to be reclaimed. Here, I respect all those amazing people.

Contents

List of Figures

List of Tables

List of acronyms

ARQ	Automatic repeat request
BPP	Binomial point process
BS	Base station
CCDF	Complementary cumulative distribution function
CDF	Cumulative distribution function
CMDP	Constrained Markov decision process
D2D	Device-to-device
DTMC	Discrete time Markov chain
FIFO	First in first out
Geo	Geometric distribution
GI	General distribution
ISR	Interference-to-signal ratio
i.i.d	Independent and identically distributed
LT	Laplace transform
MAM	Matrix-analytical method
MDP	Markov decision process
MIMO	Multiple input multiple output
MMPP	Markov modulated Poisson process
NOMA	Non-orthogonal multiple access
OFDM	Orthogonal frequency division multiplexing
PDF	Probability density function
PGFL	Probability generating functional
PH	Phase-type distribution
PMF	Probability mass function
PP	Point process
PPP	Poisson point process
QBD	Quasi-birth-and-death
QoS	Quality of service
RAN	Radio access networks
RDP	Relative distance process
RL	Reinforcement learning
r.v.	Random variable
SARSA	Current State, current Reward, next State and next Action

SG	Stochastic geometry
SINR	Signal-to-interference-plus-noise ratio
SIR	Signal-to-interference ratio
SNR	Signal-to-noise ratio
TDMA	Time division multiple access
TD	Temporal difference
UE	User equipment

Mathematical notations and variables

Mathematical notations

$\|x\|$	Spectral norm of variable x
$B \setminus A$	In B but not in A
$B \in \mathbb{R}^d$	B is a subset of $\mathbb{R}^d$
$\|B\|$	Lebesgue measure of the Borel subset B
$\mathbb{C}$	The space of complex numbers
$\mathbb{C}^+$	$\mathbb{C} \setminus \mathbb{R}$
δ_x	Dirac measure
$\Delta(B)$	Intensity measure of a point process in B
e	Exponential function
$\mathbb{E}[x]$	Expectation of r.v.X
$\mathbb{E}_y[x]$	Expectation of X over Y
$\mathbb{E}_y[x\|z]$	Expectation of X over Y and conditioned on z
$f(X)$	Function of x
${}_2F_1(a.b;c;z)$	Gauss hypergeometric function
$\forall$	For all
$G[f]$	Probability generating functional for f
$\mathrm{Im}[z]$	Imaginary part of z
$\mathscr{L}_X(s)$	Laplace transform of variable X
M_k	k-th order moment
$\mathbb{N}$	The space of natural numbers
$\binom{n}{k}$	n choose k
$n!$	Fractional of n
$(\Omega, \mathscr{F}, P)$	Probability space Ω with σ-field $\mathscr{F}$ and measure P
Φ	Poisson point process
$\Phi(B)$	Number of points of Φ in B
$\mathbb{P}^{!x}$	Reduced palm measure of the point process Φ
$\mathbb{P}[A]$	Probability of event A
$\prod_{x \in \Phi} f(x)$	Product of function f evaluated at a point x of the point process Φ
$\mathbb{R}$	The space of real numbers
$\mathbb{R}^d$	Real d-dimensional Euclidean space
$\mathbb{R}^+$	The space of non-negative real numbers

$\sum_{x\in\Phi} f(X)$	Sum of function f evaluated at a point x of Φ
$\text{var}_y[X]$	Variance of X over y
$\text{var}_y[X\|z]$	Variance of X over y and conditioned on z
$\mathbf{X}$	Matrix
$\mathbf{X}^{-1}$	Inverse of matrix X
$\mathbf{X}^{H}$	Hermitian transpose of $\mathbb{X}$
$\mathbf{X}^{T}$	Transpose of $\mathbf{X}$
$\mathbf{X}_{i,j}$	Entry (i,j) of matrix $\mathbf{X}$
$\mathbf{x}$	Vector
$\{\cdot\}$	Elements in a set
$\mathbb{1}_B(x)$	Indicator function of $x \in B$
$\|\cdot\|$	L_2 vector norm

Variables

α	Path loss exponent
α_t	Learning rate in RL-based algorithm
A	Borel subsets of $\mathbb{R}^d$
$\mathscr{A}$	Action space of RL model
$A_{i,j}$	Coefficients of matrix $\mathbf{A}$
B	Buffer length restriction
$B(t)$	Queue length at the beginning of the time slot t
$b(o,r)$	Ball of radius r centered at the origin o
$\beta_{i,t}$	State indicator of the transmitter located as x
c_d	Volume of the unit ball in a d dimensions
f_m	Transmit success probability in region-m
$\bar{F}(\theta,u)$	Meta distribution of the SINR
η	Discounting factor in reinforcement learning algorithm
γ_t	Dynamic SINR at typical UE at time slot t
$G_n(r)$	Distribution of the distance from origin to the nth nearest node
G_{tx}	Transmit antennas gains
$H_{0,t}$	Fading channel gain of the serving BS at time slot t
$H_{x,t}$	Fading channel gain of the interfering BS x at time slot t
I	Aggregated interference power
I_t	Dynamic aggregated interference power
λ_b	BS density
λ_u	UE density
λ	Lagrange multiplier
M_b	The b-moments
μ_t	Transmit success probability at the typical UE
$p_{i,j}$	Transition probability of a Markov chain

p_t	Dynamic coverage probability
p	Stable coverage probability
p_s	Stable probability
$\tilde{p}_s$	Stable probability for greedy policy
p_l	Lower bound of dynamic coverage probability
p_u	Upper bound of dynamic coverage probability
p_{loss}	Packet loss probability
$p(s', r\|s, a)$	Transition probability to be at next state s' and reward r given any state s and action a.
P_{rx}	Received antenna gains
Φ	Poisson process
Φ_t	Set of BSs that are transmitting in the time slot t
$\tilde{\Phi}$	Limit of the PPP series where the activity of BSs does not evolve with time
$p(s'\|s, a)$	State-transition probabilities
P_{rx}	Received power
P_{tx}	Transmitted power
$\pi(a\|s)$	Probability of the agent choosing action a at the state s
q_t	Probability of an interfering BS is active at time slot t
$q_\pi(s)$	Action-value function starting from s, taking the action a and thereafter following policy π
$q^*_\pi(s)$	Optimal action-value function
$r(s, a)$	Expected value for state-action pairs
(s, a, s')	Expected rewards for state-action-next-state triples
$R(t)$	Immediate reward at time slot t
$\mathscr{R}$	Reward space
$S(t)$	Environment state at time slot t
$\mathscr{S}$	State space
σ^2	Noise power (dB)
t	Time slots t
θ	SINR receiving threshold
$v_\pi(s)$	State-value function starting from s, and thereafter following policy π
$v^*_\pi(s)$	Optimal state-value function
W	Queue delay
$X(t)$	Arrival packet process
$\|x_0\|$	Distance from the typical UE located at origin to the desired BS
$\|x\|$	Distance from the typical UE located at origin to the interfering BS x
ξ	Packet arrival rate
ξ_c	Critical packet arrival rate
ξ_c^l	Lower bound of critical packet arrival rate
ξ_c^u	Upper bound of critical packet arrival rate

Résumé en français

Cette thèse traite de l'étude des performances des réseaux cellulaires dynamiques aléatoires en liaison descendante. La principale question posée dans ce manuscrit est la caractérisation de la région de stabilité d'un réseau aléatoire lorsqu'un modèle de trafic est intégré à la description de la géométrie du réseau. Nous commençons par caractériser la région de stabilité d'un réseau aléatoire, c'est-à-dire l'ensemble des intensités de trafic à partir desquelles les files d'attente des stations de base divergent. À partir de la notion de probabilité de couverture dynamique, nous prenons en compte l'interaction entre les états des files d'attentes dans le réseau à l'aide d'une modélisation par chaîne de Markov discrète des files d'attente, où le taux de service de l'utilisateur typique dépend de la probabilité de couverture dynamique. Les cas des files d'attente à taille finie et infinie sont traités. La région de stabilité indique à partir de quelle intensité de trafic au moins une file d'attente dans le réseau diverge. On souhaite également avoir une description plus fine du phénomène en répondant à la question "quelle est la proportion de files d'attente instables dans le réseau?". Dans ce cas, on a recours à la notion de ϵ-stabilité qui décrit l'ensemble des intensités de trafic pour lesquelles une file d'attente prise au hasard a une probabilité de diverger inférieure à ϵ. Enfin, la caractérisation des régions de stabilité en considérant l'allocation des ressource est très difficile à obtenir, à cause de la dépendance entre la géométrie et la dynamique du réseau et la stratégie d'allocation. Afin de s'affranchir de ce problème, nous proposons d'étudier la région de stabilité à l'aide d'algorithme d'apprentissage par renforcement. La dynamicité du réseau considérée dans cette thèse se prête parfaitement à la description par un processus décisionnel markovien pour lequel des stratégies d'apprentissage par renforcement peuvent être proposées. Nous étudions donc la région de stabilité lorsque la station de base typique peut choisir d'émettre ou de rester silencieuse selon l'état du réseau observé.

Chapitre 1 : introduction

Ce chapitre présente les motivations de la thèse et le contexte des études des réseaux cellulaires dynamiques en liaison descendante. Nous donnons ensuite la structure du manuscrit ainsi que les contribution de la thèse. Le chapitre se termine par la liste des publications relative à ces travaux.

Chapitre 2 : préliminaires mathématiques

Les outils mathématiques utilisés dans cette thèse vont de la géométrie stochastique à l'apprentissage par renforcement en passant par la théorie des files d'attente. Ce chapitre a pour objectif d'introduire les principes et les principaux résultats qui seront utilisés par la suite dans ces trois domaines.

Au niveau de la géométrie stochastique, ce chapitre illustre deux des théorèmes les plus importants de la littérature sur les processus ponctuels de Poisson (PPP) : les théorèmes de Slivnyak et de Campbell. Le théorème de Slivnyak établit qu'une statistique d'un PPP conditionnée sur un point est la même que la statistique du PPP entier. Ce théorème permet de se restreindre au calcul de la statistique du rapport signal à bruit plus interférence à l'origine, et dire que cela est représentatif du reste du réseau. D'autre part, le théorème de Campbell est utilisé pour calculer la valeur moyenne d'une somme de fonctions évaluées à l'emplacement du point de traitement.

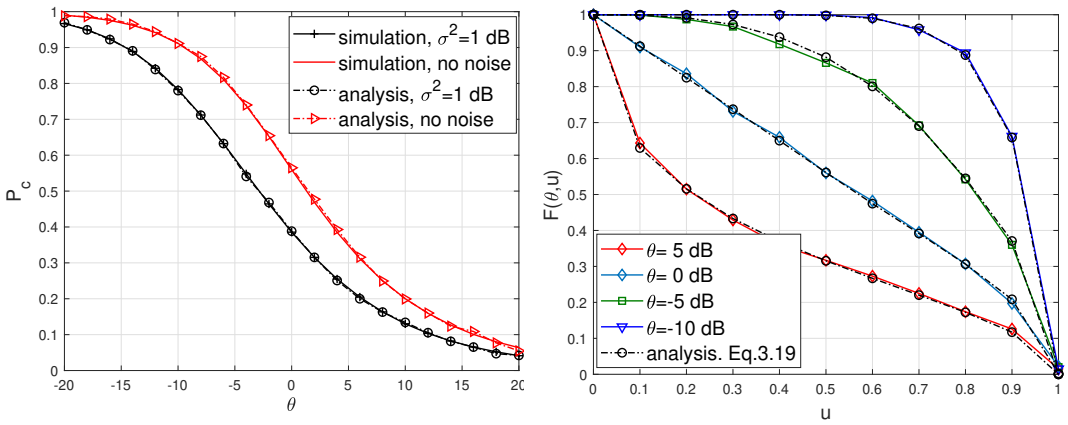
La dynamique des files d'attente est modélisée à l'aide de processus markovien qui sont introduits dans ce chapitre. Les outils d'analyse de la distribution stationnaire d'une chaîne de Markov de type naissance et mort sont introduits. Ils seront utilisés pour calculer les probabilités des états de la file d'attente dans les chapitres 4 et 5.

Enfin, l'apprentissage par renforcement est introduit comme une solution à un problème de décision séquentiel, dès lors que l'interaction de l'agent avec son environnement est modélisée par un processus décisionnel markovien [3]. L'intérêt de cette approche est qu'elle permet de trouver une politique de transmission optimale, dans un environnement incertain sans modèle physique explicite de la communication pour effectuer l'allocation des ressources, mais seulement par le biais d'essais et d'erreurs de la part de l'agent. À la fin de ce chapitre, nous donnons les principes fondamentaux de l'apprentissage par renforcement avec l'équation de Bellman, ainsi que deux algorithmes classiques d'implémentation : Q-learning et SARSA. Ces algorithmes seront ensuite appliqués pour étudier des stratégies de transmission associées à un problème d'optimisation.

Chapitre 3 : état de l'art

L'utilisation de la géométrie stochastique pour l'analyse des performances des réseaux aléatoires a fait son apparition il y a deux décennies avec l'émergence des réseaux ad-hoc mobiles. À partir de 2011, ce domaine a reçu un intérêt croissant très important de la part de la communauté scientifique, grâce à des résultats analytiques relativement simples, issus de l'étude des réseaux de points de Poisson. Depuis, les contributions théoriques et applicatives (celles qui considèrent des modèles de réseaux de plus en plus complexes) n'ont cessé de se développer. La plupart des résultats de la littérature supposent que les emplacements des stations de base sont les points d'un PPP homogène et fonctionnant à pleine charge. Cette hypothèse de modélisation est faite en raison de la simplicité mathématique de la formulation et cela permet d'obtenir une borne pessimiste de la probabilité de couverture dans les réseaux cellulaires.

Dans ce chapitre, on introduit d'abord les résultats sur la probabilité de couverture [4], et la notion de méta-distribution dans un réseau cellulaire [5]. La méta-distribution est la



(a) Probabilité de couverture

(b) Méta-distribution

FIGURE 1 – Performances en cas de pleine charge.

probabilité qu'un noeud choisi aléatoirement dans le réseau soit couvert, ou de manière équivalente, étant donné une réalisation du PPP, la proportion de noeuds couverts dans le réseau. La figure 1a représente la probabilité de couverture d'un réseau à pleine charge avec et sans bruit, resimulée à partir de [4]. La figure 1b est la méta-distribution du SINR dans un PPP, simulée à partir de [5]. Cependant, et comme nous l'avons mentionné, le cas de pleine charge peut être justifié pour les macro-cellules aux heures de pointe mais s'avère inexacte sur un système réel soumis à des variations temporelles du trafic.

Dans la deuxième partie de ce chapitre, nous présentons le modèle spatio-temporel, ainsi que les avantages et les faiblesses de la littérature traitant de ce modèle. Les travaux existants sur les modèles dynamiques spatio-temporels peuvent être divisés en deux catégories. La première catégorie est celle des modèles avec mobilité, cf. [6, 7, 8]. La deuxième catégorie est celle des réseaux de Poisson statiques, où l'emplacement des émetteurs et des récepteurs est fixe pendant toute une époque temporelle, cf. [9, 10, 11]. Dans ce travail, nous avons étudié les réseaux de Poisson statiques.

Pour simplifier l'analyse, certains travaux [11, 10] supposent que l'évolution des files d'attente est indépendante et identiquement distribuée entre toutes les stations de base. Cependant, dans les systèmes pratiques, les états des files d'attente sont corrélés temporellement et spatialement entre les émetteurs. La nouveauté de notre travail réside dans la prise en compte de la corrélation entre les interférences créées par toutes les stations de bases et l'état des files d'attente au niveau des transmetteurs au cours du temps. Les travaux en [12, 13] se sont intéressés à l'interaction entre la dynamicité des files d'attente et la topologie du réseau. Cependant, les analyses de performance sont axées sur la probabilité de couverture. Les questions de stabilité du système ne sont pas abordées, ce qui est l'objet de notre travail. Il manque encore une description analytique simple pour étudier les performances de réseaux dynamiques à grande échelle en tenant compte des corrélations spatio-temporelles. La figure 2 illustre une réalisation d'un processus ponctuel de Poisson homogène, où chaque émetteur

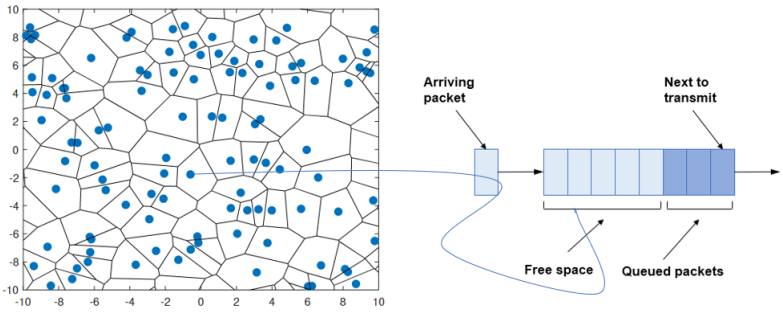


FIGURE 2 – Réalisation d'un processus de Poisson homogène avec évolution de la file d'attente dans chaque émetteur.

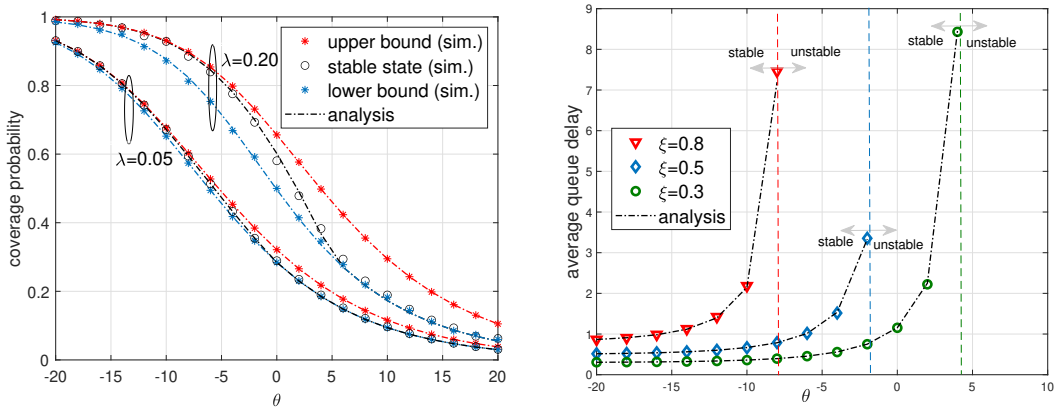
est équipé d'un buffer pour stocker les paquets.

Nous terminons ce chapitre en revenant sur les applications de l'apprentissage par renforcement dans le domaine des communications numériques. Ces dernières années, l'apprentissage par renforcement a été appliqué à des systèmes point à point afin de minimiser l'énergie consommée sous contrainte de délai, par exemple [14, 15], ou à des réseaux cellulaires réguliers pour gérer la question de la planification des connexions des utilisateurs, e.g. [16, 17]. Au cours des deux dernières années, un petit nombre de travaux ont combiné des outils de géométrie stochastique et d'apprentissage par renforcement pour étudier les problèmes d'allocation des ressources pour les réseaux aléatoires, par exemple [18, 19], ou pour les problèmes d'association d'utilisateurs, par exemple [20]. Cependant, aucun travail jusqu'à présent n'a encore combiné le RA avec des réseaux dynamiques aléatoires pour évaluer la performance moyenne, ce qui est la principale direction que nous présentons dans le chapitre 6. Cependant, aucun travail jusqu'à présent n'a encore combiné l'apprentissage par renforcement avec la géométrie stochastique dans un réseau dynamique pour l'évaluation de la performance moyenne de la RA, ce qui est la principale direction que nous présentons dans le chapitre 6.A la fin de ce chapitre, nous listons les difficultés possibles pour exploiter ensemble la SG et la RL pour étudier les réseaux cellulaires dynamiques.

Chapitre 4 : analyse de couverture des réseaux cellulaires dynamiques en liaison descendante

La probabilité de couverture dynamique et la région de stabilité des réseaux cellulaires dynamiques sont traitées dans ce chapitre en considérant des files d'attente de taille infinie et finie. L'évolution des files d'attente est modélisé par un processus de Markov à temps discret pour capturer l'interaction entre la probabilité de couverture et l'évolution de l'état des files d'attente. En régime stable, c'est-à-dire lorsque les files d'attente ne divergent pas, la probabilité de couverture suit une équation du point fixe. Nous dérivons également le délai d'attente moyen ainsi que la probabilité de perte de paquets lorsque la taille des files d'attente est finie.

La figure 3 illustre la probabilité de couverture et l'attente moyenne lorsque la taille des



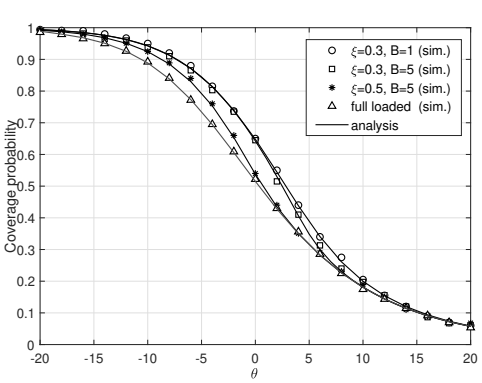
(a) Probabilité de couverture avec buffer infini.

(b) Délai moyen de la file d'attente.

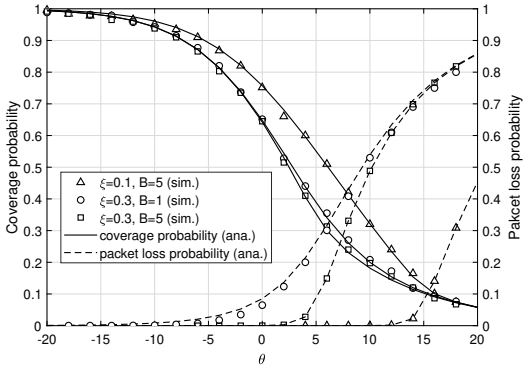
FIGURE 3 – Performance du cas de buffer infini.

files d'attente est infinie. La figure 3a donne la probabilité de couverture, en régime stable, en fonction du seuil de couverture, pour deux densités de réseau différentes, i.e. $\lambda = 0.05$ et $\lambda = 0.2$, ainsi que les bornes supérieure et inférieure de la probabilité de couverture. Dans cette figure, on compare également les résultats analytiques avec ceux de la simulation où l'on peut remarquer une parfaite adéquation, validant les modèles théoriques proposés. On peut observer que la probabilité de couverture se rapproche de la limite supérieure lorsque le seuil θ est faible, alors qu'elle se rapproche de la limite inférieure lorsque θ est élevé. Cela vient du fait que la probabilité de succès de transmission est inversement proportionnelle à la valeur de θ. De plus, la région entre les limites supérieure et inférieure diminue lorsque la densité λ des stations de bases diminue. En effet, le niveau d'interférence d'un utilisateur type diminue lorsque la densité des stations de base diminue, de sorte que la limite supérieure est proche de la limite inférieure. La figure 3b trace le délai moyen dans la file d'attente en fonction du seuil de couverture pour différents taux d'arrivée de paquets. Comme le montre la figure, le délai moyen croît exponentiellement avec le seuil de couverture, et tend vers l'infinie lorsque le seuil approche une valeur critique délimitant la région de stabilité du réseau.

La figure 4 illustre les performances lorsque la taille de la file d'attente est bornée. L'impact des différents taux d'arrivée de paquets et de la taille de la file d'attente sur la probabilité de couverture, ainsi que sur la probabilité de perte de paquets, est illustré. On constate que la taille de la file d'attente a peu d'influence sur la probabilité de couverture : en effet, si le réseau est stable le nombre de paquets stockés dans la file d'attente reste faible. D'autre part, lorsque le seuil de couverture augmente et pour un certain taux d'arrivée des paquets, les nouveaux paquets entrant sont écartés à cause de la saturation de la file d'attente. Enfin, les simulations de Monte-Carlo et l'analyse théorique montrent un très bon accord.



(a) Probabilité de couverture.



(b) Probabilités de couverture et de perte de paquets.

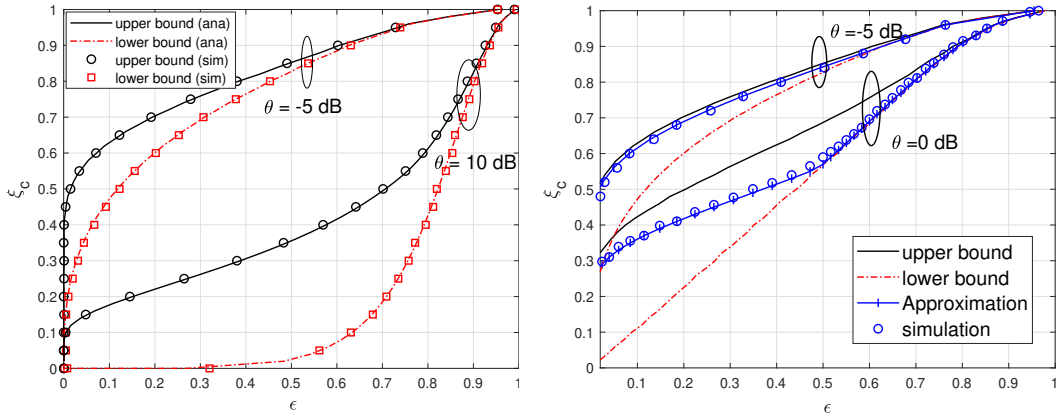
FIGURE 4 – Probabilités de couverture et de perte de paquets avec une file d'attente finie.

Chapitre 5 : Analyse de la région ϵ-stable

Dans ce chapitre, on considère un réseau dynamique où chaque station de base à une file d'attente infinie. Nous nous intéressons à la notion de région ϵ-stable, qui est définie comme l'ensemble des taux d'arrivées tels que la proportion de files d'attente instables est inférieure à ϵ.

Tout d'abord, nous dérivons des expressions analytiques les bornes supérieure et inférieure. Ensuite, nous dérivons une approximation précise du taux d'arrivée critique, i.e. le taux d'arrivée à partir duquel la probabilité de divergence d'une file d'attente dépasse ϵ, en négligeant le régime transitoire dans le calcul de la probabilité de coupure. En particulier, le modèle de chaîne de Markov discrète du chapitre 4 est utilisé pour gérer l'interaction entre la probabilité de succès de la transmission et l'évolution de la file d'attente, afin d'obtenir une approximation précise du taux d'arrivée critique.

La figure 5 représente les bornes et l'approximation de la région ϵ-stable. Les résultats numériques révèlent que l'approximation proposée est plus informative que les bornes, qui sont relativement lâches. En outre, l'écart entre les bornes supérieure et inférieure est faible lorsque le seuil de couverture est petit. Les résultats révèlent également qu'un léger changement dans le taux d'arrivée peut grandement affecter la fraction de files d'attente instables dans le réseau. Les résultats montrent également que, dans certaines configurations de réseau, de petites modifications du taux d'arrivée peuvent affecter considérablement le pourcentage de files d'attente instables dans le réseau. Par exemple, lorsque le seuil de SIR $\theta = -5$dB, le taux maximal d'arrivée des paquets autorisé est de 0.7 si 20 % des files d'attente sont autorisées à être instables ; lorsque toutes les files d'attente doivent être stables, le taux maximal d'arrivée des paquets autorisé est de 0.



(a) Bornes. (b) Approximation.

FIGURE 5 – Limites et approximation de la région ϵ-stable.

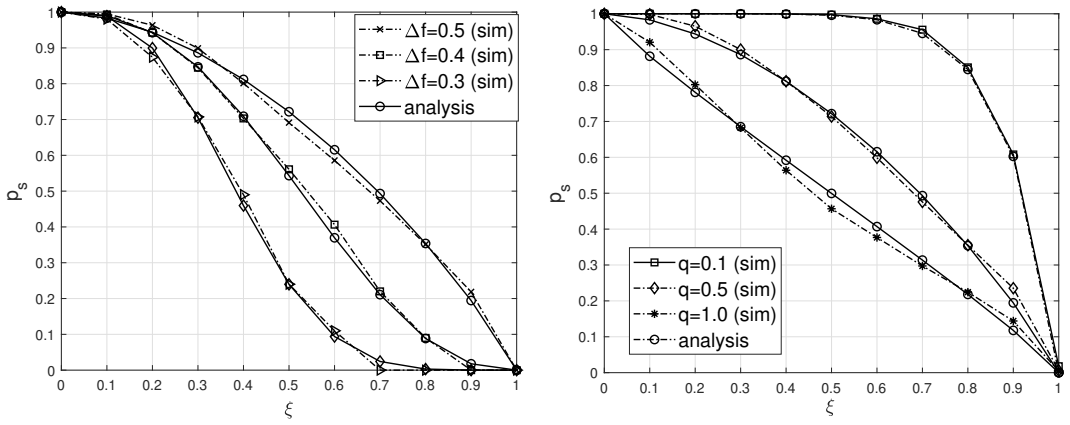
Chapitre 6 : Politiques basées sur l'apprentissage par renforcement dans les réseaux cellulaires dynamiques

Dans ce chapitre, nous proposons des politiques de transmission tenant compte de l'information sur l'état du canal, de l'état des files d'attente et de l'interférence globale dans les réseaux cellulaires dynamiques en liaison descendante. L'objectif est de trouver la politique de transmission afin minimiser les coûts de transmission tout en limitant le coût d'attente dans le buffer. Le problème est formulé à l'aide d'un processus de décision de Markov à horizon infini et est résolu en ligne par apprentissage par renforcement. Nous analysons d'abord la probabilité de stabilité de la politique gloutonne, i.e. le transmetteur émet dès qu'un paquet est présent dans la file d'attente, qui sert de politique de référence pour étudier celles obtenues par apprentissage. La politique gloutonne fournit une limite supérieure de la probabilité de stabilité par rapport à la politique basée sur l'apprentissage par renforcement. Nous montrons qu'il existe un compromis entre la probabilité de stabilité et les coûts de transmission qui dépend de l'intensité du trafic. Les résultats numériques révèlent que les politiques basées sur l'apprentissage maintiennent la même région de stabilité par rapport à l'algorithme glouton, mais avec un coût de transmission inférieur.

La figure 6 montre la probabilité de stabilité $\tilde{p}_s$ en fonction de ξ sur Δf et q, où Δf est la probabilité de succès de transmission avec un SIR croissant dans différents intervalles et q est la probabilité d'activité d'une station de base interférente choisie au hasard. On constate que la probabilité de stabilité $\tilde{p}_s$ diminue lorsque ξ augmente. De même, étant donné ξ dans la figure 6a, on peut observer que $\tilde{p}_s$ est décalé vers une valeur plus élevée lorsque $\Delta f = 0.3$ par rapport à $\Delta f = 0.5$. Cela signifie qu'une plus grande probabilité de succès de transmission assure une meilleure probabilité de stabilité. En outre, comme dans la figure 6b, plus l'activité de la station de base interférente est faible, plus la probabilité de stabilité est élevée. Cela vient du fait qu'un q plus grand entraîne un niveau de brouillage plus élevé, ce qui diminue le SIR

au niveau de l'utilisateur type et réduit encore la probabilité de stabilité.

La figure 7 compare le coût total et la probabilité de stabilité des politiques en fonction de l'intensité du trafic. Pour la même configuration de réseau, la politique basée sur l'apprentissage est capable de maintenir la même région de stabilité à un coût de transmission inférieure à la politique gloutonne. Il n'y a pas de différence significative de performance entre les algorithmes Q-learning et SARSA qui convergent tous les deux vers la politique optimale. Nous observons qu'il existe un compromis entre la probabilité de stabilité et le coût total des politiques basées sur l'apprentissage. En effet, à mesure que l'intensité du trafic augmente, la probabilité de stabilité diminue et l'agent a tendance à être plus actif dans l'envoi de paquets, ce qui augmente d'autant le coût de transmission. Les politiques basées sur l'apprentissage par renforcement permettent à l'agent d'ajuster de manière flexible la politique de transmission en fonction de l'intensité du trafic et de la configuration du réseau, tandis que la politique gourmande n'est sensible qu'aux états de la mémoire tampon.



(a) Effet de Δf.

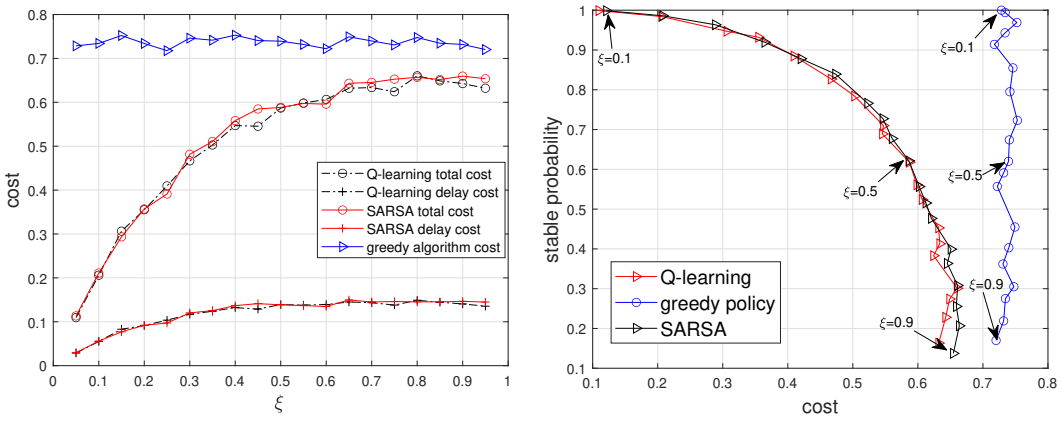
(b) Effet de q.

FIGURE 6 – Probabilité de stabilité avec la politique gloutonne.

Chapitre 7 : conclusions et travaux futurs

Ce chapitre conclut le travail et présente des extensions potentielles des travaux réalises dans cette thèse. Par exemple :

- Le chapitre 6 se focalise sur la présence d'un seul agent dans le réseau, tandis que les autres fonctionnent toujours avec une politique gloutonne, ce qui ne rend pas le problème symétrique. L'extension à l'apprentissage par renforcement multi-agent est une étape nécessaire pour analyser les performances du réseau avec cette stratégie. Cependant, l'extension est non-triviale car des problèmes d'instabilité peuvent apparaître lorsqu'on envisage un apprentissage distribué décentralisé. La formulation théorique d'un apprentissage multi-agents et les conditions de convergence restent à explorer.



(a) Les coûts des différentes politiques.

(b) Compromis probabilité de stabilité-coût.

FIGURE 7 – Performance politique basée sur l'AR.

- Les travaux actuels portent sur les réseaux statiques, ce qui signifie que les émetteurs et les récepteurs ne changent pas de position pendant la transmission. Dans une prochaine étape, la mobilité pourrait être introduite ce qui posent la question de la modélisation des flux de charge à travers les cellules.

Chapter 1

Introduction

The massive connectivity of mobile devices, the demand for high data rates, and spectrum limitations have imposed complicated interference situations on modern wireless systems. In order to specify the quality of service of devices in large-scale wireless networks, coverage probability is undoubtedly an important metric. To provide a unified mathematical framework to characterize the coverage probability, stochastic geometry has been widely used in various wireless systems, e.g., the device-to-device networks [21], the broadcast networks [22], and the heterogeneous networks [23, 24], etc.

On the other hand, with the development of the multi-media and the evolution of mobile applications, a large amount of data from different sources are transmitted simultaneously in the same communication system. In these systems, the impact of the data traffic on the network services is no longer negligible. The conventional approaches of network performance analysis [4, 25, 26] heavily rely on the full buffer assumption, i.e., each base station has a backlog of packets and is always transmitting. The performance of dynamic cellular networks, while taking into account that interactive queues to be characterized, is the focus of this thesis.

This thesis is focused on characterizing the coverage probability and the ϵ-stable region for downlink cellular networks with traffic-aware. However, we also revisit the performance analysis of a fully-loaded downlink cellular network as part of the literature review. The initial work starts with constructing tractable mathematical models to describe the coverage probability, queue delay, and packet loss probability, considering different application scenarios with infinite and finite buffers. Subsequently, the ϵ-stable region is studied, which is the set of arrival rates such that the proportion of unstable queues is not more significant than ϵ. Finally, transmission strategies are studied by introducing the Markov decision process and solving it online using reinforcement learning.

1.1 Background and motivation

Stochastic geometry successfully provided a unified mathematical framework to model different types of large-scale wireless networks by characterizing the statistics of the signal to interference plus noise ratio (SINR) of a randomly chosen user [27]. Stochastic geometry captures the spatial randomness of the wireless systems and can take into account fading,

shadowing, and power control [27, 4, 5]. In the last decade, stochastic geometry has been combined with more complex network models taking into account frequency reuse, multiple antennas, multiple-tiers, or load-aware protocols, to cite a few.[28, 29, 12].

On the other hand, real systems are subjected to temporal traffic variations and sources generate packets according to some stochastic processes [28]. The full-load hypothesis, which assumes that each cell uses the same frequency band and is always transmitting, is not enough to address the performance of practical systems because a transmitter in a cell is active only if there are packets in its queue. Therefore, load-awareness is essential for practical performance assessment. However, the interaction between queues, i.e., the state of one queue depends on the state of the others, makes the problem mathematically rather involved [30].

A first attempt combining stochastic geometry and queuing theory has been granted in [7] by considering a double-stochastic network, i.e., the users appear randomly in space and time when they have a packet to transmit. However, the interaction between the queues at different BSs are ignored. The works in [12, 13, 31, 32] pushed further the analysis of the interaction between the queue dynamics and the topology of the network. A traffic-aware spatio-temporal model for IoT devices supported by cellular uplink connectivity has been developed in [12]. The scalability and stability of the network, i.e., its ability to support a large number of devices while the queues are not diverging, have been studied. Similarly, [33] assumed that the traffic is generated at random spatial regions, rather than modeling the flow at each independent user. In [13], a novel spatio-temporal mathematical framework is provided to analyze the preamble transmission success probability of a cellular IoT network, where the number of accumulated packets in the queues is approximated by a Poisson distribution. However, the theoretical findings are not validated by simulations.

We develop a comprehensive approach to handle the interaction between the coverage probability and the queueing state evolution using discrete time Markov chain (DTMC). A simple model is considered, but contrarily to the state of the art [29, 8, 34], closed-form expressions are given that make the bridge between the coverage probability and the fraction of active base stations (BS) under conditional stable state. We also characterize the explicit upper and lower bounds on the dynamic coverage probability. Besides, to the best of our knowledge, all the works mentioned before studied the coverage probability with infinite queue lengths. However, the packet loss probability is also an important performance measure needed for the design of telecommunication networks. The quantity of interest is the probability of a new packet is dropped when the buffer has a finite size [35, 36]. Subsequently, we proposed a tractable mathematical model to analyze the coverage probability and packet loss probability considering the buffer restriction. Particularly, we derive the closed-form expression of the coverage probability that depends on the activity probability of a randomly chosen BS which is related to the buffer length. We also characterize the packet loss probability of a randomly chosen BS when the network convergence, i.e., DTMC works at stationary regime.

On the other hand, the primal consideration in queueing systems is about stability. For a point-to-point system with random arrival and departure processes, the stable region requires that the service rate be larger than the arrival rate. However, traffic conditions are more complicated in a large-scale network with multiple queues since the service rate depends on the state of all transmitters in the network. Then, sufficient and necessary conditions for

system stability have been introduced and studied in [29], and meta-stability in [8], where the network appears stable for possibly a long time and then suddenly exhibits instability. Remarkably, the stochastic geometry and queuing theory have been merged to give sufficient and necessary conditions for the stability of interacting queues [29]. However, this work has considered a peer-to-peer network with constant link distances. On the other hand, the stability and meta-stability of uplink random access network considering the data traffic have been studied in [8]. The analysis in this work was based on a double-stochastic networks, i.e., space-time Poisson call arrivals. However, a single cell network is considered in this work, thereby ignoring the interaction between the queues at different BSs.

We characterize the ϵ-stable region in a large scale dynamic downlink cellular network, with multi-cells and random link distances, contrarily to [29]. Moreover, the ϵ-stable region provides fine-grained information on the network stability and answers question such that "what is the set of arrival rates such that the proportion of unstable queues in the network is below ϵ, if the required signal-to-interference ratio (SIR) is θ?". The characterization of the ϵ-stable region relies on the use of meta-distribution [5]. We provide closed-form expressions for upper and lower bounds considering modified systems and Markov inequalities. Besides, we propose an alternative definition of the ϵ-stable region and derive a tight approximation of the critical arrival rate accordingly.

Our original works mentioned previous are based on simple transmission schemes, thus the key performance metrics, such as coverage, delay, and ϵ-stable region can be characterized as exact and tractable expressions. Decoupling the SIR analysis and the statuses of the queues at all BSs is difficult when we consider adaptive transmission policies. The difficulty of the performance analysis lies in the fact that the SIR relies on the statuses of the queues at all BSs, and on the other hand the SIR at the users also affects the statuses of the queues.

Reinforcement learning is one of the most important research directions of machine learning, which has significantly impacted the development of artificial intelligence over the last 20 years. RL is a learning process in which an agent can periodically make decisions, observe the results, and then automatically adjust its strategy to achieve the optimal policy [37, 38]. However, since the nature of the problems studied with SG or RL is so fundamentally different, it is rare to find common ground where the strength of these tools can be jointly leveraged. In the end of this manuscript, we investigate the transmission policies considering the channel state information and queue states and interference in dynamic downlink cellular networks. The problem is formulated with a countable state, infinite horizon, discounted cost Markov decision process (MDP) with infinite buffer assumption, and solving it online using RL. The goal is to minimize the transmission costs possible while limiting the delay cost in the buffer of typical BS. In the end, we compare the performance with greedy algorithm.

1.2 Structure of the thesis and contributions

The manuscript is organized as follows:

- Chapter 2. We review the important definitions, properties, lemmas, theorems, etc. from stochastic geometry, queuing theory as well as reinforcement learning that are used

later in subsequent chapters.

- Chapter 3. We first review the coverage probability and meta-distribution of downlink cellular networks with base stations operating at full capacity. We then present the spatio-temporal model with traffic awareness and the advantages and disadvantages of the literature dealing with this problem. We conclude the chapter by returning to the application of reinforcement learning in digital communications and listing the difficulties that can be encountered when combining SG and RL to study dynamic cellular networks.

- Chapter 4. We develop a comprehensive approach to deal with the interaction between coverage probability and traffic evolution using discrete-time Markov chains. Contrary to the state of the art, we give closed expressions for the stable coverage probability and build bridges to the probability of being active at an arbitrary interfering base station. In addition, we describe upper and lower bounds on the dynamic coverage probability. When the base stations are equipped with infinite buffer, we calculate the average queue delay. The numerical results show that when the network is stable, the average delay stabilizes at a small value, and when the network is unstable, the average delay plummets to infinity. In the end, when base stations are equipped with finite buffer, we characterize the effect of buffer length on the coverage and packet loss probability.

- Chapter 5. We study the stability region of packet arrival rates in dynamic downlink cellular networks, where each base station has an infinite buffer. We introduce the ϵ-stable region concept and derive the corresponding upper and lower bounds. Furthermore, we propose an alternative definition of the ϵ stability and derive the approximation of the critical arrival rate accordingly. In particular, the DTMC model in the chapter 4 is used to deal with the interaction between transmission success probability and queue evolution to obtain an approximation of the critical arrival rate.

- Chapter 6. We consider the problem of reinforcement learning-based transmission policies considering channel state information, queueing state, and dynamically aggregated interference in large-scale networks with multiple cells and random link distances. The goal is to minimize the transmission cost while limiting the waiting cost in the buffer. First, we give closed-form expressions for the stable probability based on different policies, and the results show that the greedy policy provides an upper bound on the stable probability than the RL-based transmission policy. Second, there is a tradeoff between the stable probability and the transmission cost. However, numerical results show that the RL-based strategy can achieve the same stable probability but lower transmission cost than the greedy algorithm.

- Chapter 7. This chapter provides the conclusions and future perspective of our works.

1.3 Publications

[C1] **Q. Liu**, J. -Y. Baudais and P. Mary, "A Tractable Coverage Analysis in Dynamic Downlink Cellular Networks," 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).

[C2] **Q. Liu**, J. -Y. Baudais and P. Mary, "Queue Analysis with Finite Buffer by Stochastic Geometry in Downlink Cellular Networks," 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring).

[C3] **Q. Liu**, J. -Y. Baudais and P. Mary, "Analysis of the ϵ-stable region in dynamic downlink cellular networks," 2022 IEEE 95rd Vehicular Technology Conference (VTC2022-Spring).

[C4] **Q. Liu**, P. Mary, J. -Y. Baudais, "Politiques de transmission basées sur l'apprentissage par renforcement dans les réseaux cellulaires dynamiques et aléatoires," submitted to GRETSI 2022.

[C5] **Q. Liu**, P. Mary, J. -Y. Baudais, "Stability analysis based on reinforcement learning and stochastic geometry for dynamic downlink cellular networks," a journal on the preparing.

Chapter 2

Mathematical Background

2.1 Stochastic geometry

Cellular networks are based on the concept of replacing a single cell with high power BS by several cells with low power BSs for higher system capacity [39]. Performance of cellular networks depends largely on the location of BSs and users. Stochastic geometry (SG) is a field of applied probability that aims at providing tractable mathematical models and appropriate statistical methods to study and analyze random phenomena on the plane $\mathbb{R}^2$ or in larger dimension [40]. Besides, SG enables to study the behaviours of wireless networks averaged over random spatial realization. Our work includes the modelling the network nodes by Poisson point process (PPP) leveraging the tools from stochastic geometry, particularly the point process theory. We review important definitions, properties and theorems related to Poisson point process from [25, 41], which we will used in the subsequent chapters.

2.1.1 Poisson point process

A Poisson point process is a random collection of points and plays a fundamental role in the description of the network geometry. A Poisson point process can be defined as follows.

Definition 2.1 ([41])**.** *A point process with intensity measure Λ is a Poisson point process (PPP) if for every compact $B \in \mathbb{R}^d$, $\Phi(B)$ has a Poisson distribution with mean $\Lambda(B)$, that is*

$$\mathbb{P}\left(\Phi(B) = k\right) = e^{-\Lambda(B)} \frac{\Lambda(B)^k}{k!} \tag{2.1}$$

If Λ admits a density λ, the Poisson distribution can be expressed as

$$\mathbb{P}(\Phi(B) = k) = \exp\left(-\int_B \lambda(x)dx\right) \cdot \frac{(\int_B \lambda(x)dx)^k}{k!} \tag{2.2}$$

A consequence of this definition is the independence property: If $B_1, B_2, \cdots, B_n$ are disjoint compact sets, then $\Phi(B_1), \Phi(B_2), \cdots, \Phi(B_n)$ are independent.

Due to its analytical tractability and analytical flexibility, PPP has been the "model of choice" in many researches in the last decades [27]. In the following, we will discuss key proprieties underlying such tractability.

Property 2.1 (Superposition). *The superposition of two independent PPP Φ_1 with density λ_1 and Φ_2 with density λ_2, is another PPP with density $\lambda_1+\lambda_2$.*

Property 2.2 (Independent thinning). *Independent selection of points in a PPP with probability p results in another PPP with density $p\lambda$. This is known as independent thinning.*

Property 2.3 (Stationary PPP). *A PPP Φ is stationary if its distribution is translation invariant, i.e., if $\Lambda(B)=\lambda|B|$ for all B. A stationary PPP can also be called uniform or homogeneous.*

When PPP is stationary, the statistical characteristics seen from a homogeneous PPP are independent of the observation location. In other words, the interference characterized on an arbitrary test point is equivalent to the interference characterized at any other location in $\mathbb{R}^2$ including the points in Φ.

In Poisson networks, the probability densities of the distances from a point to its nth nearest neighbor are given in a simple form [42]. We focus on a ball $b(o,r)$ of radius r centered at the origin o.

Property 2.4 (Distances). *The distribution of the distance from origin to the nth nearest node is*

$$\begin{aligned} G_n(r) &= 1-\mathbb{P}(\Phi(b(o,r))<n) \\ &= 1-\exp(\lambda c_d r^d)\sum_{k=0}^{n-1}\frac{(\lambda c_d r^d)^k}{k!} \end{aligned} \tag{2.3}$$

where $c_d=|b(o,r)|=\frac{\pi^{d/2}}{\Gamma(d/2+1)}$ is the volume of the unit ball in d dimensions, and $\Phi(b(o,r))$ is the number of points within the ball $b(o,r)$. When taking the derivation, the probability density is the generalized gamma distribution

$$g_n(r)=\exp(-\lambda c_d r^d)\frac{d(\lambda c_d r^d)^n}{r\Gamma(n)} \tag{2.4}$$

when $n=1,d=2$, (2.4) reduces to Rayleigh distribution with mean $\frac{1}{2\sqrt{\lambda}}$.

The main observation to obtain (2.3) is that the nth-nearest node is at at distance larger than r if there are at most $n-1$ nodes in $b(o,r)$. The distribution of the interpoint distances is important in the performance evaluation of wireless networks. For example, when user terminals and BSs follow two independent PPP, and each user connect to its nearest BS, the probability density function (PDF) of the distance from user to its serving BS can by easily obtained by Property 2.4 when $n=1$.

2.1.2 Slivnyak-Mecke theorem

Slivnyak's theorem states that for a Poisson point process Φ, since all points are independent of each other, conditioning on a point at x does not change the distribution of the rest of the process.

Theorem 1 (Palm distribution)**.** *The reduced Palm distribution equals the distribution of the PPP itself and can be written as*

$$\mathbb{P}^{!x} \equiv \mathbb{P} \tag{2.5}$$

where $\mathbb{P}$ *is the distribution of the PPP, and* $\mathbb{P}^{!x}$ *is the reduced Palm probability, defined as the distribution of the PPP assuming a point at* $x \in \mathbb{R}^d$*.*

It is often desirable to make statistical statements about a randomly selected (or "typical") point in a point process. However, we cannot simply pick a point uniformly from an infinite number of points, and if we define a rule on how to pick such a point, we introduce biasing because the rule will have to depend on the point's surroundings. To solve this issue, we usually define the point process to have a point at a specific location, i.e., the Palm distribution [41].

2.1.3 Functions of point process

The mean of sum and product of functions evaluated at points of a point process have a wide range of applications in wireless communications, and they are described in the following.

Theorem 2 (Campbell's theorem [25])**.** *For a PPP* Φ *with intensity measure* Λ *and a measurable function* $f : \mathbb{R}^d \to \mathbb{R}^+$*, the expectation of the sum of* f *is*

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x,\Phi)\right] = \int_{\mathbb{R}^d} \mathbb{E}\left[f(x,\Phi\cup x)\right]\Lambda(dx) \tag{2.6}$$

and

$$\mathbb{E}\left[\sum_{x\in\Phi} f(x,\Phi\backslash x)\right] = \int_{\mathbb{R}^d} \mathbb{E}\left[f(x,\Phi)\right]\Lambda(dx) \tag{2.7}$$

The first form is obtained because conditioning on a point at x is the same as adding that point, and the second (reduced) form follows from (2.5) because $\mathbb{P}^{!x} \equiv \mathbb{P}$. Note that the expectation on the left side of the equation cannot be placed inside the summation, since the summation is random and is a function of Φ.

The expected product over a point process is called *Probability generating functional* (PGFL), named equivalently the Laplace functional, defined as follows.

Definition 2.2 (PGFL)**.** *For a measurable function* $f(x) : \mathbb{R}^d \to [0,1]$ *such that* $1-f$ *has bounded support, the PGFL* $G[f]$ *of the point process* Φ *is defined as*

$$G[f] \triangleq \mathbb{E}\left[\prod_{x\in\Phi} f(x)\right] \tag{2.8}$$

Theorem 3 (PGFL of PPP [25])**.** *For a PPP* Φ *with intensity measure* Λ *and a measurable function* $f : \mathbb{R}^d \to [0,1]$*, the PGFL of the PPP is*

$$G[f] = \mathbb{E}\left[\exp\left(-\int_{\mathbb{R}^d}(1-f(x)\Lambda(dx))\right)\right] \tag{2.9}$$

PGFL is often used in the evaluation of Laplace transform of $\sum_{x\in\Phi} f(x)$, which can be expressed as

$$\mathbb{E}\left[\exp(-s\sum_{x\in\Phi} f(x))\right] = \mathbb{E}\left[\prod_{x\in\Phi}\exp(-sf(x))\right] = \exp\left(-\int_{\mathbb{R}^d}\left(1-e^{-sf(x)}\Lambda(dx)\right)\right) \tag{2.10}$$

Since the signal-to-interference ratios are determined by relative distances, it is sometimes convenient to work directly with the point process of relative distances, introduced in [43]. For the stationary PPP, the distance relative to the nearest point and the PGFL function are tractable also and given next.

Theorem 4 (PGFL of relative distance process (RDP) [43])**.** *For a stationary point process Φ, let $x_0 = \arg\min\{x\in\Phi : \|x\|\}$ be the point closest to the origin. The relative distance process (RDP) is defined as*

$$\mathscr{R} \triangleq \left\{x\in\Phi\backslash\{x_0\} : \frac{\|x_0\|}{\|x\|}\in(0,1]\right\}. \tag{2.11}$$

When Φ is a PPP, the PGFL of the RDP is

$$G_{\mathscr{R}}[f] = \frac{1}{1+2\int_0^\infty (1-f(x))x^{-3}dx}, \tag{2.12}$$

for functions $f:[0,1]\to[0,1]$ such that the integral in the denominator of (2.12) is finite.

In the next chapters, Campbell's theorem and PGFL theorem are widely applied to calculate the expectation of interference functional with the form $\mathbb{E}[e^{-sI}]$, where I is the interference power.

2.1.4 Moment measures

Definition 2.3 (Moment measures)**.** *The nth moment measure of a point process Φ is defined as the expected product of the number of points falling in regions $B_1, B_2, \cdots, B_n \in \mathscr{B}$, where $\mathscr{B}$ is the Borel σ-algebra.*

$$\mu^{(n)}(B_1\times B_2\times\cdots\times B_n) \triangleq \mathbb{E}[\Phi(B_1)\Phi(B_2)\cdots\Phi(B_n)] \tag{2.13}$$

The nth moment measure can also be viewed as the intensity measure of the product point process $\Phi^{(n)} = \underbrace{\Phi\times\cdots\Phi}_{n \text{ times}}$. The element of $\Phi^{(n)}$ are the n-tuples $(x_1, x_2, \cdots, x_n)\in\mathbb{R}^{nd}$, where $x_k\in\Phi$ is a point of a point process.

For $n=1$, the moments measure reduce to the intensity measure. For $n=2$, and $B_1=B_2=B$, we have $\mu^{(2)}(B^2)=\mathbb{E}[\Phi(B)^2]$, and thus

$$\text{var}(\Phi(B)) = \mu^{(2)}(B^2) - \Lambda(B)^2 \tag{2.14}$$

In Chapter 5 and Chapter 6, the moment measures are generally used to calculate the meta distribution, the ϵ-stable region, which will be defined later.

2.2 Queuing theory

The real systems are subjected to temporal traffic variations and sources generate packets according to some stochastic processes [28]. Markov process is a special stochastic process in which the state of the system in the future is independent of the past history of the system but dependent only on the present. In our work, we model the traffic by discrete Markov process and later modelling the traffic and the strategies by discrete Markov decision process. This section provides the definitions and basic tools related to Discrete-Time Markov chains (DTMC), and the matrix-analytical method for Markov chains used in next chapters, which are reproduced mostly from [35] unless otherwise mentioned.

2.2.1 Discrete-time Markov chains

Definition 2.4. *Consider a discrete time stochastic process $X_0, X_1, \cdots$ with discrete (i.e. finite or countable) state space, $\mathscr{I}_c = \{i_0, i_1, i_2, \cdots\}$. If*

$$\mathbb{P}\{X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \cdots, X_0 = i_0\} = \mathbb{P}\{X_{t+1} = i_{t+1} | X_t = i_t\} \tag{2.15}$$

holds for any time t, and states $i_{t+1}, i_t, i_{t-1}, \cdots, i_0$, then X_t is said to be a discrete-time Markov chain. If further we have

$$\mathbb{P}\{X_{t+m+1} = j | X_{t+m} = i\} = \mathbb{P}\{X_{t+1} = j | X_t = i\},\ \forall (i,j) \in \mathscr{I}^2, \forall m \geq 0, \tag{2.16}$$

then the Markov chain is time-homogeneous or stationary.

Generally we say

$$p_{i,j} = \mathbb{P}\{X_{t+1} = j | X_t = i\}, \tag{2.17}$$

and the basic properties of the elements of the transition probability of a Markov chain are

$$0 \leq p_{i,j} \leq 1,\ \forall (i,j) \in \mathscr{I}^2 \tag{2.18}$$

$$\sum_{j \in \mathscr{I}} p_{ij} = 1. \tag{2.19}$$

DTMC has many useful proprieties that make it tractable.

Property 2.5 (Irreducible Chain)**.** *A Markov chain in which every state can be reached from every other state is called an irreducible Markov chain.*

Property 2.6 (Absorbing Chain)**.** *A Markov chain is said to be an absorbing chain if at least for one of its states i, $\mathbb{P}\{X_{t+1} = i | X_t = i\} = p_{i,i} = 1$.*

Property 2.7 (Recurrent Markov Chain)**.** *A Markov chain is said to be recurrent if all the states of the chain are recurrent. A state i is positive recurrent (or non-null persistent) if the expected return time is finite, i.e., $\mathbb{E}[\inf\{t > 1 : X_t = i\}] < \infty$.*

Property 2.8 (Ergodic Markov Chain)**.** *An irreducible Markov chain is said to be ergodic if all its states are aperiodic and positive recurrent.*

2.2.2 Matrix-analytical method for DTMC

In the queueing theory, Kendall's notation is the standard system used to describe and classify a queueing node [44]. The queueing models use three factors written $A/S/c$. Specifically, A denotes the distribution of inter-arrival time, S gives service time distribution (time between service start and completion), c is the number of service channels opens at the node. Some common notations of arrival process and the service time distribution is shown in Table 2.1 and Table 2.2.

Symbol	**Name**	**Description**
M	Markovian or memories	Exponential inter-arrival time [45]
M^X	Batch Markov	A generalization Markovian arrival process by allowing dependent inter-arrival time, correlated batch sizes [46]
D	Degenerate distribution	A deterministic or fixed inter-arrival time [47]
Geo	Geometric distribution	Geometric inter-arrival time [48, 49]
PH	Phase-type distribution	Phase-type distribution constructed by a convolution or mixture of exponential distributions [12, 50]
G (GI)	General distribution	General independent arrival process [51]

Table 2.1 – The arrival process.

Symbol	**Name**	**Description**
M	Markovian	Exponential service time [45]
D	Degenerate distribution	A deterministic or fixed service time [52]
G (GI)	General distribution	Independent service time [53]
MMPP	Markov modulated Poisson process	Exponential service time distributions, where the rate parameter is controlled by a Markov chain [54]

Table 2.2 – The service time distribution.

The matrix-analytical method (MAM) is a technique to compute the stationary probability distribution of a Markov chain which has a repeating structure (after some point) and a state space which grows unboundedly in no more than one dimension [35]. The matrix-analytical method is most suited to three classes of Markov chains: *i*) those with the GI/M/1 structure [55] *ii*) those with the M/G/1 structure [35], *iii*) and those with the Quasi-Birth-and-Death (QBD) structure which actually embodies the combined properties of both the GI/M/1 and M/G/1 structures. Since the QBD DTMC is widely used in the following chapters, we mainly characterize this structure.

The transition matrix $\mathbf{P}$ of QBD DTMC has the following structure [35]:

$$\mathbf{P} = \begin{bmatrix} \mathbf{B} & \mathbf{C} & & & & \\ \mathbf{E} & \mathbf{A_1} & \mathbf{A_0} & & & \\ & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & & \\ & & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_0} & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \tag{2.20}$$

where $\mathbf{B} \in \mathbb{R}$, $\mathbf{C} \in \mathbb{R}^{1\times n}$, $\mathbf{E} \in \mathbb{R}^{n\times 1}$, $\mathbf{A_0} \in \mathbb{R}^{n\times n}$, $\mathbf{A_1} \in \mathbb{R}^{n\times n}$ and $\mathbf{A_2} \in \mathbb{R}^{n\times n}$ are sub-stochastic metrix that capture the transitions between queue levels. Specifically, the sub-matrix $\mathbf{A_0}$ capture the event where a new task arrives; $\mathbf{A_1}$ capture the event the number of tasks in the queue unchanged; $\mathbf{A_2}$ capture a service completion occurs and no new task arrive, which decreases the number of tasks in the queue by one. Boundary vectors $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{E}$ captures the transition from idle-to-idle, idle to level and from level to idle, respectively.

If $\mathbf{P}$ is irreducible and positive recurrent then the stationary distribution is given by the solution

$$\mathbf{xP} = \mathbf{x}, \ \ \mathbf{xe} = 1 \tag{2.21}$$

where $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots]$ is the row vector that contains the steady-state probabilities and $\mathbf{e}$ represents a vector of suitable dimension with all values equal to 1.

The calculation of stationary distribution with QBD structure often requires the introduction of two important matrices, i.e., $\mathbf{R}$ and $\mathbf{G}$, which are given as

$$\mathbf{R} = \mathbf{A_0} + \mathbf{RA_1} + \mathbf{R}^2\mathbf{A_2} \tag{2.22}$$

and

$$\mathbf{G} = \mathbf{A_2} + \mathbf{A_1G} + \mathbf{A_0G}^2 \tag{2.23}$$

$\mathbf{R}$ and $\mathbf{G}$ are the minimal non-negative solutions of (2.22) and (2.23), respectively. Further, a relationship between $\mathbf{R}$ and $\mathbf{G}$ are

$$\mathbf{R} = \mathbf{A_0}(\mathbf{I} - \mathbf{A}_1 - \mathbf{A_0G})^{-1} \tag{2.24}$$

$$\mathbf{G} = (\mathbf{I} - \mathbf{A}_1 - \mathbf{RA_2})^{-1}\mathbf{A_2} \tag{2.25}$$

Noted that $\mathbf{R}$ can be computed using cyclic reduction method, invariant subspace method or logarithmic reduction method, which are detailed in [35]. Once $\mathbf{R}$ is obtained, $\mathbf{x}_0$ and x_1 and therefore iteratively all the x_i can be solved according to

$$\mathbf{x}_{i+1} = \mathbf{x}_i\mathbf{R}, \ i \geq 1 \tag{2.26}$$

and

$$\begin{pmatrix} \mathbf{x}_0 & \mathbf{x}_1 \end{pmatrix} \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{E} & \mathbf{A}_1 + \mathbf{RA}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{2.27}$$

In the following chapters, the traffic at each BS is usually modeled by the QBD DTMC, and matrix-analytical method is used to calculate the stationary distribution of the Markov chain.

2.3 Reinforcement learning

Reinforcement learning is a subclass of machine learning approach that learns online to maximize a long-term reward without any *priori* information. The methodology is to discover which actions yield the most valuable reward by trying them. Five main elements identify an RL system: the agent that can perceive the environment states and can take actions that affect the state; the environment in which the agent lives and interacts; the policy, which is a mapping from states to actions; and a reward signal that quantifies the quality of the actions. Fig. 2.1 diagrams the agent–environment interaction.

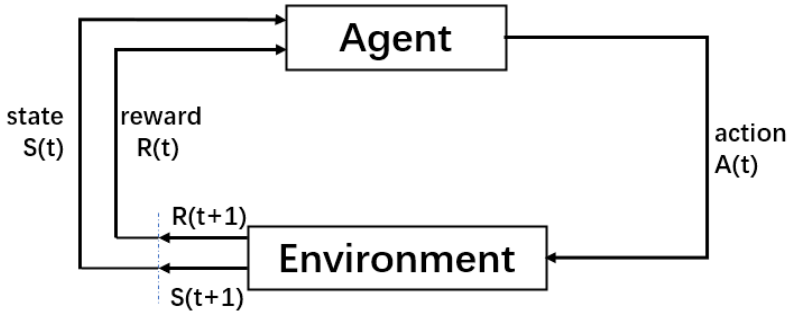


Figure 2.1 – The agent–environment interaction in reinforcement learning [1].

Definition 2.5 (Reinforcement learning [1])**.** *At each time t, the agent receives some representation of environment state $S(t) \in \mathscr{S}$ and a numerical reward $R(t) \in \mathscr{R}$, where $\mathscr{S}$ and $\mathscr{R}$ are the state space and reward space, respectively. On that basis the agent selects an action $A(t) \in \mathscr{A}$ to collect a new reward $R(t+1)$ at state $S(t+1)$ at time $t+1$, where $\mathscr{A}$ is the action space. The objective is to develop the RL policy $\pi\colon \mathscr{S} \to \mathscr{A}$ that maximizes the long term return in state S.*

RL are typically modeled using an Markov decision process (MDP) framework. The MDP provides a formal model to design and analyze RL problems as well as a rigorous way to design algorithms that can perform optimal decision making in sequential scenarios. We introduce the MDP first.

2.3.1 Markov decision process

An MDP is made of four components: a set of states, a set of actions, a transition kernel, i.e., the stochastic law of states transition given an action and a reward. Given any state and action s and a, the probability of each possible pair of next state and reward, s', r is denoted as [37]

$$p(s', r|s, a) \triangleq \mathbb{P}\left[S(t+1) = s', R(t+1) = r | S(t) = s, A(t) = a\right] \tag{2.28}$$

the expected value for state-action pairs is

$$r(s, a) \triangleq \mathbb{E}[R(t+1)|S(t) = s, A(t) = a] = \sum_{(s',r)\in\mathscr{S}\times\mathscr{R}} r\, p(s', r|s, a) \tag{2.29}$$

the state-transition probabilities is

$$p(s'|s,a) \triangleq \mathbb{P}\left[S(t+1)=s'|S(t)=s, A(t)=a\right] = \sum_{r\in\mathscr{R}} p(s',r|s,a) \tag{2.30}$$

and the expected rewards for state-action-next-state triples is

$$r(s,a,s') \triangleq \mathbb{E}\left[R(t+1)|A(t)=a, S(t+1)=s', S(t)=s\right] = \frac{\sum_{r\in\mathscr{R}} r\, p(s',r|s,a)}{p(s'|s,a)} \tag{2.31}$$

A *policy* defines how the agent behaves in a given state and specifies the best sequence of actions to get the maximum reward on the long run. The policy maps states to actions, i.e. $\pi : \mathscr{S} \rightarrow \mathscr{A}$. The policy can be *deterministic* or *stochastic*. A *deterministic* policy is a single or a set of deterministic actions when an agent encounters state $S(t)$. A *stochastic* policy is defined as a conditional probability measure of $A(t)$ given $S(t)$, i.e., $\mathbb{P}_{A(t)|S(t)}(a|s)$, which is usually simply denoted as $\pi(a|s)$.

A state-value function, $v_\pi(s)$, is a measure of the overall expected return assuming that the agent is in state s and follows a policy π. An action-value function, $q_\pi(s,a)$, also called Q-Value (where Q is abbreviation from the word Quality), is a measure of the overall expected return assuming that the agent is in state s, takes an action a, and follows a policy π.

Definition 2.6 (State-value function)**.** *The state-value function is the expected return starting from s, and thereafter following policy* π

$$v_\pi(s) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \eta^t R(t+1)|S(0)=s\right] \tag{2.32}$$

where $\eta < 1$ is the discount factor emphasizing more immediate rewards, $\mathbb{E}_\pi[\cdot]$ denotes the expected value of a random variable given that the agent follows policy π.

Definition 2.7 (Action-value function)**.** *The action-value function is the expected return starting from s, taking the action a, and thereafter following policy* π

$$q_\pi(s,a) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \eta^t R(t+1)|S(0)=s, A(0)=a\right] \tag{2.33}$$

A remarkable property of the value function is that it follows a recursive relation that is widely known as the Bellman equation [56, 37].

2.3.2 Bellman equation for a single agent

For any policy π and any state, one can prove the following consistency condition holds between the value of s and the value of its possible successor states [37]

$$v_\pi(s) = \sum_{a\in\mathscr{A}} \pi(a|s) \sum_{(s',r)\in\mathscr{S}\times\mathscr{R}} p(s',r|s,a)[r+\eta v_\pi(s')] \tag{2.34}$$

The expression in (2.34) has to be seen as the expectation of the random variable $R(t+1)+\eta v_\pi(S(t+1))$ over the joint distribution $\mathbb{P}_{A(t)|S(t)}\mathbb{P}_{S(t+1)R(t+1)|S(t)A(t)}$. The Bellman equation

links the state-value function at the current state with the next state-value function averaged over all possible states and rewards knowing the current state and the policy π.

Similarly, the action-value function can be expressed as

$$q_\pi(s,a) = \sum_{(s',r)\in\mathscr{S}\times\mathscr{R}} p(s',r|s,a)[r+\eta v_\pi(s')] \tag{2.35}$$

The relationship between $v_\pi(s)$ and $q_\pi(s,a)$ is

$$v_\pi(s) = \sum_{a\in\mathscr{A}} \pi(a|s) q_\pi(s,a) \tag{2.36}$$

The optimal state-value and action-state value functions are obtained by maximizing $v_\pi(s)$ and $q_\pi(s,a)$ over the policies, that is [37]

$$v^*(s) = \max_{\pi\in\Psi} v_\pi(s), \quad \forall s\in\mathscr{S}, \tag{2.37}$$

and

$$q^*(s,a) = \max_{\pi\in\Psi} q_\pi(s,a), \forall s\in\mathscr{S}, a\in\mathscr{A}, \tag{2.38}$$

where Ψ is the set of all stationary polices, i.e., policies that do not evolve with time.

Moreover, (2.37) can be written w.r.t. (2.38) as $v^*(s) = \max_{a\in\mathscr{A}} q^*(s,a)$, the optimal state value function also obeys to the Bellman recursion [37]

$$v^*(s) = \max_a \sum_{(s',r)\in\mathscr{S}\times\mathscr{R}} p(s',r|s,a)\left[r+v^*(s')\right] \tag{2.39}$$

Besides, by substituting $v^*(s')$ in (2.39) with the maximum over the actions of $q^*(s',a')$, we can obtain the optimal bellman equation of action-state value function:

$$q^*(s,a) = \sum_{(s',r)\in\mathscr{S}\times\mathscr{R}} p(s',r|s,a)\left[r+\eta\max_{a'\in\mathscr{A}} q^*(s',a')\right] \tag{2.40}$$

A policy π is defined to be better than or equal to a policy π' if its expected return is greater than or equal to that of π' for all states. In other words, $\pi \geq \pi'$ if and only if $v_\pi(s) \geq v_{\pi'}(s)$ for all $s\in\mathscr{S}$. The optimal strategy is labeled as π^*, which is corresponding to the optimal action-value function $q^*(\cdot)$.

2.3.3 Q-learning and SARSA algorithm

Reinforcement learning enables an agent to learn timely in a *trial-and-error manner* as time goes by $t = 1,2,\cdots$ without the need for external supervision [1]. There are two main approaches of temporal difference learning, namely, off-policy and on-policy, that differ among themselves in the way in which the action-value in (2.40) is updated. A popular off-policy approach is Q-learning [57], and a popular on-policy approach is the current State, current Reward, next State and next Action (SARSA) [58]. In chapter 6, both Q-learning and SARSA are used to investigate the optimal strategy, we introduce as following.

The basic idea of Q-learning algorithm is to build a new estimate from an old estimate, which is updated by an incremental difference between a target and the old estimate. This can be formalized as follows:

$$\underbrace{q_t(S(t),A(t))}_{\text{new estimate}} \leftarrow \underbrace{q_t(S(t),A(t))}_{\text{old estimate}} + \alpha_t \left[\underbrace{T_{t+1}}_{\text{target}} - \underbrace{q_t(S(t),A(t))}_{\text{old estimate}} \right], \tag{2.41}$$

where $0 < \alpha_t < 1$ is the learning rate. The learning rate tells us how much we want to explore something new and how much we want to exploit the current choice. Higher learning rate α_t increases responsive to the difference between the target and old estimate. The learning rate should satisfy the following conditions: $\sum_{t=0}^{\infty} \alpha_t = \infty$ and $\sum_{t=0}^{\infty} \alpha_t^2 < \infty$. Note that α can also be taken as a constant less than one, and hence not satisfying the conditions above. However in practice, learning tasks occur over a finite time horizon hence the conditions are satisfied.

In the Q-learning algorithm, the target in (2.41) is equal to:

$$T_{t+1} = R(t+1) + \eta \max_{a' \in \mathscr{A}} q_t(S(t+1), a'), \tag{2.42}$$

which is the algorithmic from of the Bellman equation in (2.40). When the algorithm has converged, T_{t+1} should be equal to $q_t(S(t),A(t))$, nullifying the difference term in (2.41).

To choose an action in an arbitrary state, either exploitation or exploration methods can be used. Exploitation selects the best-known (greedy) action that maximizes the Q-values as follows:

$$a^* = \arg\max_{a \in \mathscr{A}} q_t(S(t), a) \tag{2.43}$$

Exploration selects a random action so that its Q-value can be updated in order to discover better actions in a dynamic and stochastic operating environment as time progresses.

To balance between exploitation and exploration, the ϵ-greedy policy is a widely used technique. It consists in randomly choosing an action with probability ϵ and the action that maximize the current action-value at time t with probability $1-\epsilon$. ϵ can be kept constant or may vary during the learning in order to explore more at the beginning. Mathematically, the ϵ-greedy policy can be expressed as

$$a^* = \begin{cases} \arg\max_{a \in \mathscr{A}} q_t(S(t), a) & \text{with probability } 1-\epsilon \\ a \in \mathscr{A} & \text{with probability } \epsilon \end{cases} \tag{2.44}$$

Above all, Algorithm 1 shows the classical Q-learning algorithm embedded in the agent.

Algorithm 1 Classical Q-learning algorithm

(1) **begin procedure**
(2) Observe current state $S(t)$
(3) Select action $A(t)$ using (2.43) or ϵ-greedy policy in (2.44)
(4) Receive immediate reward $R(t+1)$
(5) Update Q-value $q_t(S(t),A(t))$ using (2.41)
(6) **end procedure**

In SARSA, the updating the Q-value depends on the agent's current state $S(t)$, the action $A(t)$ chosen by the agent, the reward $R(t)$ received by the agent for choosing this action, the state $S(t+1)$ entered by the agent after taking this action, and finally the next action A(t+1) chosen by the agent in the new state $S(t+1)$. Thus one iteration of SARSA is a quintet $(S(t), A(t), R(t), S(t+1), A(t+1))$. Q-learning updates the estimate of the best action-value function based on the maximum reward of available actions, while SARSA learns the Q-value associated with the adoption of a policy, it follows a rule, e.g., ϵ-policy. In Fig. 2.2, we illustrate the updating rules.

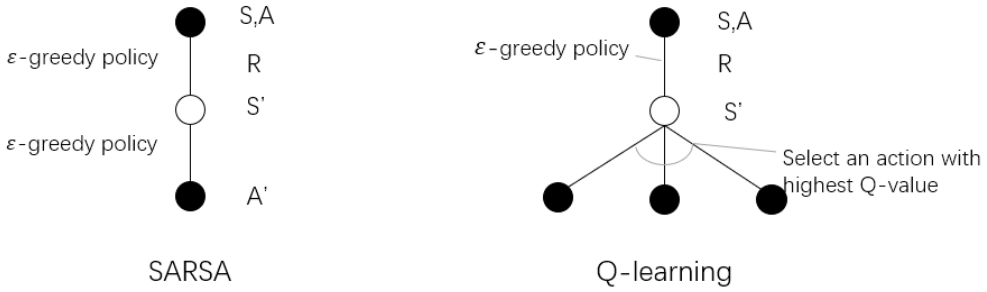


Figure 2.2 – The difference between Q-learning and SARSA.

Besides, there are also some classical RL algorithms, such as policy-based algorithm (e.g., Policy Gradient), Actor-Critic, Proximal policy optimizations. Some advantages and limitations of the most common RL algorithms are listed below in Table 2.3.

ML Approach	**Advantages**	**Limitations**
Q-learning	(i) Learn directly the optimal policy (ii) Less computation cost (iii) Relatively fast (iv) Efficient for offline learning	(i) Use of biased samples (ii) High per-sample variance (iii) Computationally expensive
SARSA	(i) Fast (ii) Efficient for online learning datasets	(i) Learns a near-optimal policy while exploring (ii) Not very efficient for offline learning
Policy Gradient	(i) Capable of finding best stochastic policy (ii) Effective for high dimensionality datasets	(i) Slow convergence (ii) High variance
Actor Critic	(i) Reduces variance with respect to pure policy methods (ii) More sample efficient than other RL methods (iii) Guaranteed convergence	(i) Must be stochastic

Table 2.3 – Advantages and limitations of RL methods [2].

2.4 Conclusion

In this chapter, after describing a brief history of stochastic geometry and its application to analyze network performance, we presented general definitions and notations of the spatial point processes that will be used in this thesis. We outlined the Poisson point process and its properties. Then, we presented the discrete time Markov chain as well as Matrix analytical method to solve the stationary distribution of DTMC, which will be used to model the traffic in the subsequent chapters. In the end, we introduced the reinforcement learning model as well as the classical algorithms. In the next chapter, we present a state of art of approaches that deal with the performance analysis in PPP networks. In particular, we review the recent techniques that have been proposed to model the spatio-temporal cellular networks.

Chapter 3

State of the art

3.1 Stochastic geometry modeling

Stochastic geometry has become a necessary theoretical tool for analyzing and characterizing large-scale wireless systems in the last decades [42]. The key idea, in the application of the SG to wireless network analysis, is to model the locations of BSs and user equipments (UEs) as a realization of a class of point processes. Instead of deterministic locations of BSs and users on a regular grid with a small number of user and BSs, the stochastic geometry gives general analytical models that catch the randomness all cellular network's realizations. Hence, general analysis for cellular networks should be based on the probabilistic spatial distribution of BSs rather than on deterministic networks realization. We are interested in the performance of a randomly selected user or the average performance of all users.

3.1.1 PPP cellular networks model

The cellular networks were mostly assumed to be spatially deployed according to an idealized hexagonal grid. In the regular hexagonal networks, the analysis can be achieved for a *fixed* user with a small number of interfering BSs [59, 60, 61]. However, hexagonal networks are highly idealized and may be inaccurate for the heterogeneous and ad hoc deployments, where the cell dimension varying considerably due to differences in the transmission power, the tower height, and the user density [27].

Motivated by its tractability, attempts to promote SG to model cellular networks can be traced back to the late 90's [62]. The studies presented in [4, 63] revealed that cellular networks deviate from the idealized hexagonal grid structure and follow an irregular topology. Instead of assuming the deterministic positions for BSs on a regular grid, the locations are modeled randomly, that take into account the changes of point positions in each realization. The results show that the coverage probability experienced by a typical user in a real network is upper bounded by the coverage probability of an ideal hexagonal grid network and lower bounded by the coverage of random networks [4].

We illustrate an example of PPP cellular networks in Fig. 3.1. The locations of the BSs are modeled as a homogeneous PPP of density λ. The cell boundaries are shown by black lines

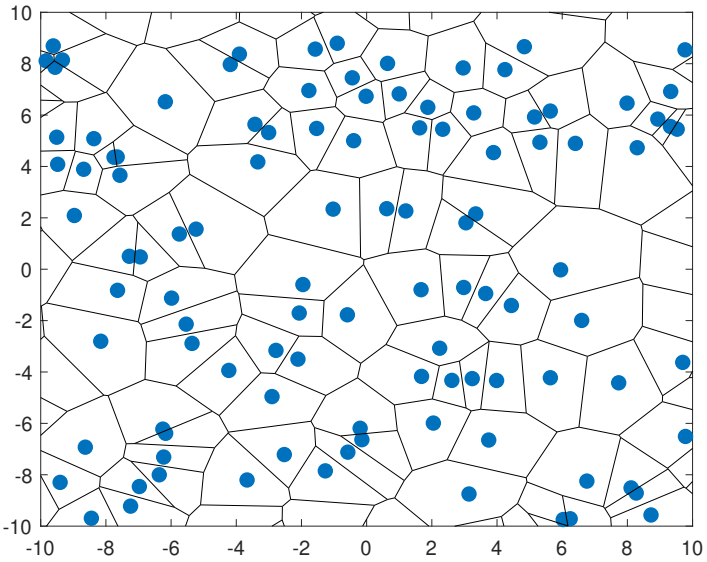


Figure 3.1 – A realization of an homogeneous Poisson process observed in a finite window.

and form a Voronoi tessellation. The UEs can be scattered according to some independent homogeneous point process with a different density, and they communicate with the nearest BS while all other BS act as interferers. Each user is associated with the nearest BS. These BS and UE models will be used in Chapter 4, 5, and 6.

It is worth noting that practical cellular network deployments are likely to exhibit some interactions among the locations of the BSs, which include spatial inhibition [64], i.e., repulsion, and spatial aggregation [65], i.e., clustering. Some point processes, like cluster, hard-core, Cox, and Gibbs processes can be used to study these characteristics in cellular networks [42, 64, 65]. However, the non-PPP model increases the complexity of analysis and have less tractability compared with PPP models. Especially, when SG is merged with more complex communication models, for example, the load awareness model, the analysis may intractable. Thus in this thesis, we still use PPP to model the locations of BSs.

3.1.2 Propagation model

When a signal is transmitted with the power P_{tx}, the received power P_{rx} at a distance r from the transmitter is given by [42]

$$P_{\text{rx}} = P_{\text{tx}} \times G_{\text{tx}} \times G_{\text{rx}} \times H \times L(r) \tag{3.1}$$

Note the in (3.1), H and $L(r)$ are dimensionless and characterize the propagation channel. The quantities G_{tx} and G_{rx} are the gains of the transmit and receive antennas, respectively, and normalized as 1 in the following chapters.

The random variable H is the fading channel gain. The fading is described by the probability distribution of the amplitude of the equivalent baseband complex waveform that is actually transmitted over the wireless channel [42]. If the link is subject to the Rayleigh fading, then the distribution of H is the square of a Rayleigh-distributed random variable, and it is therefore exponential. Without loss of generality, the expected value of H can be set to unity, thus its probability density function is

$$f_H(x) = \exp(-x),\ x > 0 \tag{3.2}$$

$L(r)$ is called the path loss, where r is the transmitter-receiver distance. A widely used model has the form

$$L(r) = r^{-\alpha} \tag{3.3}$$

where α is called the path loss exponent, with value is normally in the range from 2 to 4, where 2 is for the propagation in free-space and 4 is for relatively lossy environments. In some environments, such as buildings, stadiums and other indoor environments, the path loss exponent can reach values in the range of 4 to 6 [42].

Assume that the serving BS is denoted as b_0, the interfering BSs are modeled as a PPP Φ. The distance from the serving BS and the interfering BSs are denoted r_0 and $\{r_1, r_2, \cdots\}$, respectively. All BSs are identical distributed and transmit with the same power P, the fading coefficients from the BS to the user are i.i.d. random variables $(H_i)_{i\geq 1}$. Then the interference power at the receiver is given by

$$I = \sum_{i \in \Phi / b_0} P H_i L(r_i) \tag{3.4}$$

Besides, the user equipment always has a certain level of additive and constant noise power, which is denoted by σ^2. The SIR and SINR are defined by

$$\text{SIR} = \frac{P H_0 L(r_0)}{I}, \quad \text{SINR} = \frac{P H_0 L(r_0)}{\sigma^2 + I} \tag{3.5}$$

The distributions of the SIR or SINR drive the performance analysis of cellular networks. In the next section, we introduced the most important performance metrics which we widely used in the next chapters: the coverage probability and the meta distribution.

3.1.3 Coverage probability and meta distribution

The coverage probability evaluates the probability of the successful communication, which occurs if the SINR is larger than a threshold. The meta distribution aims to evaluate fine-grained information on the distribution of the SINR, which aims at capturing variability of SINR at particular points on an area.

Definition 3.1 (Coverage probability [4])**.** *The coverage probability P_c is defined as the probability that the typical user can reach a SINR threshold $\theta \in \mathscr{R}^+$*

$$P_c(\theta) = \mathbb{P}(\text{SINR} > \theta) \tag{3.6}$$

P_c can also be interpreted as the transmit success probability, and conditioned on Φ, it becomes

$$P_c(\theta) = \mathbb{E}_\Phi\left[\mathbb{P}(\text{SINR} > \theta|\Phi)\right] \tag{3.7}$$

Definition 3.2 (Meta distribution [66])**.** *The meta distribution of the SINR is a two-parameter distribution function*

$$\bar{F}(\theta, u) \triangleq \mathbb{P}\left(\mathbb{P}(SINR > \theta|\Phi) > u\right),\ \theta \in \mathscr{R}^+,\ u \in [0,1]. \tag{3.8}$$

We have $\bar{F}(0,u) = 1$ for $u < 1$, $\lim_{\theta\to\infty} \bar{F}(\theta,u) = 0$ for $u > 0$, $\bar{F}(\theta,1) = 0$, and $\bar{F}(\theta,0) = 1$. Due to the ergodicity of the point process, $\bar{F}(\theta,u)$ can be interpreted as the fraction of links in each realization that achieve an SINR of θ with probability at least u. Note that $\bar{F}(\theta,u)$ is the complementary cumulative distribution function (CCDF) of the transmit success probability for a given θ.

The relationship between coverage probability in (3.6) and meta distribution is

$$P_c(\theta) = \int_0^1 \bar{F}(\theta,u)\,du = \lim_{u\to 1}\int_0^u \bar{F}(\theta,x)\,dx \tag{3.9}$$

Example 3.1 ([4])**.** *The coverage probability has an elegant form when the link to the serving BS at distance r_0 has Rayleigh fading. Note that the probability density function (PDF) of r_0 is $f_{r_0}(x) = 2\pi\lambda x e^{-\lambda\pi x^2}$ (seen in (2.4)), the coverage probability can be derived as follows.*

$$P_c(\theta) = \mathbb{P}(SINR > \theta) = \int_0^\infty \mathbb{P}\left(H_0 > \frac{\theta}{PL(r_0)}\left(\sigma^2 + I\right)\right) f_{r_0}(x)\,dx \tag{3.10}$$

$$= \int_0^\infty \exp\left(-\frac{\theta}{PL(r_0)}\sigma^2\right)\mathbb{E}_I\left[\exp\left(-\frac{\theta}{PL(r_0)}I\right)\right] f_{r_0}(x)\,dx \tag{3.11}$$

For any random variable X, its Laplace transform is

$$\mathscr{L}_X(s) = \mathbb{E}\exp(-sX),\ s \geq 0 \tag{3.12}$$

Note that from (3.11), the coverage probability is given in terms of the Laplace transform of the interference power I. Let $s = \frac{\theta}{PL(r_0)}$,

$$\mathscr{L}_I(s) = \mathbb{E}_{x,H_i}\left[\exp\left(-s\sum_{i\in\Phi\setminus b_0} H_i L(r_i)\right)\right]$$

$$\overset{(a)}{=} \exp\left(-2\pi\lambda\int_{r_0}^\infty \left(1 - \mathbb{E}_H\left[\exp(-sHL(v))\right]\right) v\,dv\right). \tag{3.13}$$

where (a) follows from i*) the i.i.d. distribution of interference channel H_i and its independence from the point process Φ; and* ii*) the PGFL of the PPP.*

If the interference links also experience Rayleigh fading, (3.13) can be further expressed as

$$\mathscr{L}_I(s) = \exp\left(-2\pi\lambda\int_{r_0}^\infty \left(1 - \int_0^\infty \left[\exp(-shL(v))\exp(-h)\,dh\right]\right) v\,dv\right) \tag{3.14}$$

$$= \exp\left(-2\pi\lambda\int_{r_0}^\infty \left(1 - \frac{1}{1+sL(r_0)}\right) v\,dv\right) \tag{3.15}$$

Then the coverage probability has the following expression

$$P_c(\theta) = \int_0^\infty \exp\left(-\frac{\theta}{PL(r_0)}\sigma^2\right) \mathscr{L}_I\left(\frac{\theta}{PL(r_0)}\right) f_{r_0}(x)\Big) dx \tag{3.16}$$

$$= \pi\lambda \int_0^\infty e^{-\pi\lambda v(1+\rho(\theta,\alpha))-\theta\sigma^2 v^{\alpha/2}} dv \tag{3.17}$$

where $\rho(\theta,\alpha) = \theta^{\alpha/2}\int_{\theta^{2/\alpha}}^\infty \frac{1}{1+u^{\alpha/2}} du$.

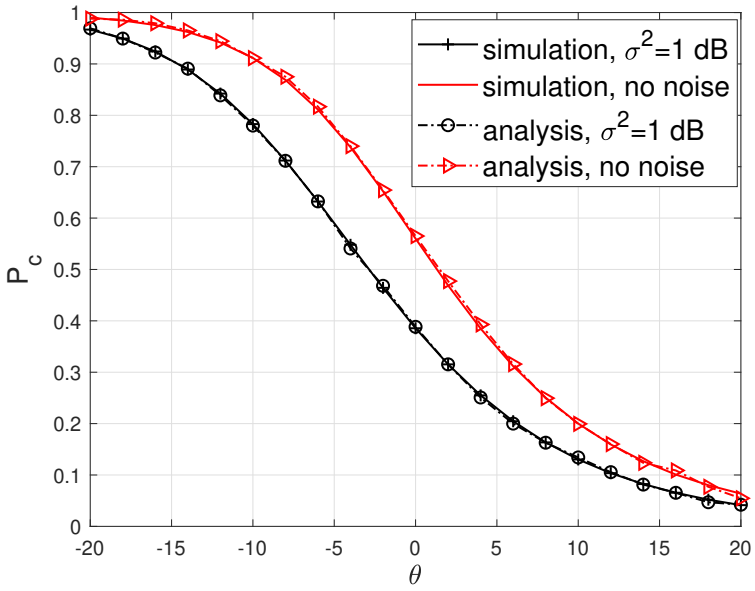


Figure 3.2 – The coverage probability $P_c(\theta)$ versus θ, with $\alpha = 4$, and $\lambda = 0.25$.

Fig. 3.2 plots the coverage probability in (3.17) where the desired and interference links both experience Rayleigh fading. To illustrate the impact of the noise, we consider $\sigma^2 = 1$ *dB and* $\sigma^2 = 0$ *dB. We see the coverage probability decreases when* θ *and the noise* σ^2 *increase.*

Example 3.2 ([66])**.** *A classical form of the meta distribution can be obtained when the link to the serving BS at distance* r_0 *is Rayleigh distributed. We neglect the noise at the user equipments, so that the SINR is replaced by the SIR for purpose of analysis. We start by defining the conditional SIR distribution given the BS point process*

$$P_s(\theta) \triangleq \mathbb{P}(SIR > \theta|\Phi) \tag{3.18}$$

The quantity of interest is the meta distribution of the SIR, which is the distribution of P_s*:*

$$\bar{F}(\theta,u) \triangleq \mathbb{P}\left(P_s(\theta) > u\right),\ \theta \in \mathbb{R}^+,\ u \in [0,1] \tag{3.19}$$

While a directed calculation of the CCDF of (3.18) seems infeasible, we shall see that the moment of $P_s(\theta)$ *can be expressed in closed-form, which allows the derivation of an exact analytical*

expression. Assuming that the transmit power P is normalized, the bth moment of the $P_s(\theta)$ is given by [41, Lemma 6.3.3]:

$$M_b = \mathbb{E}_\Phi\left[\mathbb{P}\left(H_0 > \frac{\theta}{PL(r_0)} I\right)^b\right] = \left\{1 + 2\int_0^1 \left[1 - \frac{1}{(1+\theta r^\alpha)^b}\right] r^{-3} dr\right\}^{-1} \tag{3.20}$$

Using the Gil-Pelaze inversion theorem, and assigning b in (3.20) as $b = iw$. we obtain an exact integral expression for the meta distribution

$$\bar{F}(\theta, u) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{u^{-iw} M_{iw}\right\} dw \tag{3.21}$$

$$= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{\frac{u^{-iw}}{1 + 2\int_0^1 \left[1 - \frac{1}{(1+\theta r^\alpha)^b}\right] r^{-3} dr}\right\} dw,\ i = \sqrt{-1}. \tag{3.22}$$

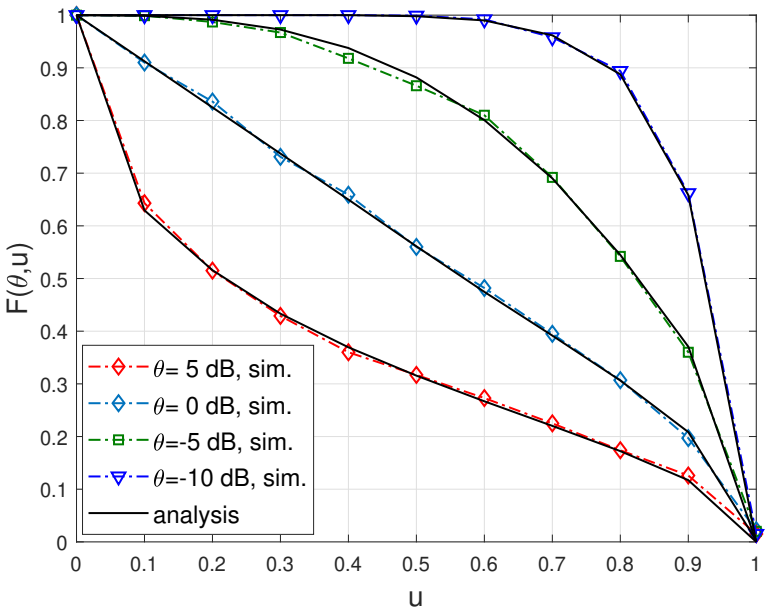


Figure 3.3 – Meta distribution $\bar{F}(\theta, u)$ with $\theta \in [-10, -5, 0, 5]$ dB, with $\alpha = 4$, and $\lambda = 0.25$.

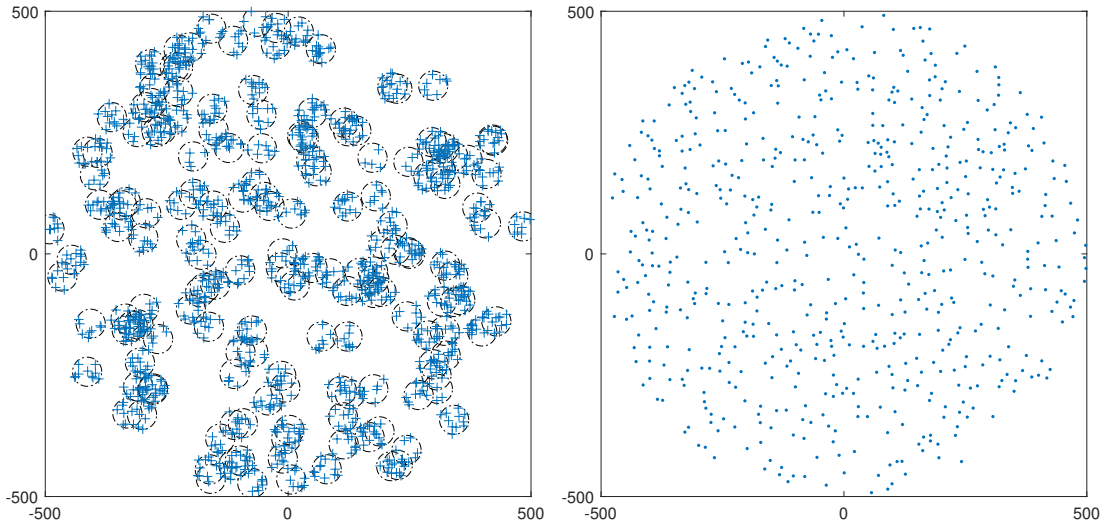
Fig. 3.3 plots the meta distribution $\bar{F}(\theta, u)$ in (3.22) with respect to u and for four values of θ, and when the desired link and interference links experience Rayleigh fading. Fig. 3.3 shows that 60% of links achieve an SIR of -5 dB with 80% of reliability. Moreover, if u is fixed, the value of $\bar{F}(\theta, u)$ increases as θ decreases.

3.1.4 Some advanced research approaches

The previous section detailed the coverage probability and meta-distribution of a simple model based on a single layer, where all BSs have the same parameters and are operating at a

full load, referring to the fundamental works [4, 5]. However, in reality, the network designer needs to make model assumptions based on specific scenarios, such as multiple antenna setups, the effect of irregular distribution of BSs, the choice of serving BSs, and the effect of load, etc. These complex models naturally affect the expression of the SINR distribution. In recent years, a large amount of research has been devoted to studying scenarios closer to real-world network deployments, and we summarize some of them below.

Location model or node type (transmitter, receiver) The spatial distribution of the networks nodes can be categorized into three types: independent [4, 67], repulsive [68, 69, 70], and clustered [71, 72, 73]. Additionally, a point process can be a mix of the above three types. A snapshots of the aforementioned clusters and repulsive point processes are given in Fig. 3.4.



(a) Doubly Poisson cluster process

(b) Matern hard-core process of type II

Figure 3.4 – Snapshots of cluster and repulsive point processes over a circle with radius of 500 m.

Various cell association strategies The most common strategy is the *nearest-neighbor cell association*, where the users are associated to the closet BS [4, 74, 75]. Furthermore, considering system models incorporating various propagation, the association policy can be *the smallest path loss cell association* [76, 77]. Also, in some environments, the interferers may be closer to the typical user than the serving BS. The user is served by the BS that provides the best SINR instantaneously, namely *the maximum-SINR association policy* [78, 79, 75].

Downlink-uplink analysis In downlink wireless networks, the analysis generally focuses on the received SINR at the level of the typical UE served by one or more BSs [4, 78, 24]. However, with the growing interest in symmetric traffic applications, the uplink performance analysis

is becoming increasingly crucial [80, 81, 82]. The main difference from the downlink model is that power controls are widely used in uplink cellular networks, where each user adjusts its transmission power to partially/locally invert the effect of the path loss [80, 81]. On the other hand, the dependency in the location of concurrent uplink users needs to be considered [83]. The paradigm of decoupled uplink-downlink access are investigated in [84, 85], where different association policies are considered for uplink and downlink. The typical user will not necessarily access to the same BS in both directions.

Propagation models Wireless communications are typically impaired by various effects. For example, there are fluctuations in the received signal power due to shadows, multiple copies of the same transmitted signal received by the receiver due to multiple propagation paths, and transmission loss problems due to the distance from the receiver to the transmitter. The study of the shadowing can be found in [79, 86, 22] and representative examples of the impact of the path loss functions in [87, 88, 89]. The detailed survey can be founded in [90, 91].

3.1.5 Summary

Section 3.1 has introduced how the Poisson point process facilitates the performance modeling and the analysis of large-scale wireless networks. However, most of the literature relies heavily on the assumption that BSs transmit concurrently all the time, which translated to a fully load (or full buffer) scenario, resulting in pessimistic estimates of the coverage and the average rate. Although, this might be justified for macrocells in peak traffic hours, this is not applicable for the real system who is subjected to temporal traffic variations. In the next section, we investigate the spatio-temporal modeling which can capture the various traffic loads.

3.2 Spatio-temporal modeling

In this section, we introduce the spatio-temporal model that we will use in the next sections and examine the advantages and weaknesses of the literature dealing with this problem. In particular, the spatial domain of dynamic systems is captured by appropriate PP modeling of the nodes, while temporal variations are captured by temporal arrival and service processes. Besides, we introduce the concept of ϵ-stable regions, which will be studied in the next chapters.

3.2.1 SINR model with traffic-aware

A generalized model of dynamic SINR in the cellular networks can be defined as below [92]. Considering the typical user located at origin o, the interference of the typical user at time instant t is

$$I(t) = \sum_{i \in \Phi \setminus b_0} \beta_{i,t} H_{i,t} L(r_i), \tag{3.23}$$

and the SINR of the typical user is

$$\mathrm{SINR}(t) = \frac{H_{0,t} L(r_0)}{\sigma^2 + I(t)}, \tag{3.24}$$

where $L(\cdot)$ is the path loss function defined in (3.3). The BSs locations are modeled by a PPP Φ of density λ: b_0 is the serving BS under a given association strategy, $H_{0,t}$ and $H_{i,t}$ are the fading coefficients of the serving BS and interfering BS i at time slot t, respectively, σ^2 is the noise power, and $\beta_{i,t}$ is the state indicator of the transmitter located at x which equals 1 or 0 when the transmitter is on or off, respectively.

Queue model We considered a time-slotted network through the manuscript. Each time slot has an equal and small time interval duration τ. We assume that each transmitter holds an independent buffer to restore the backlogged packets. The transmission during the time is asynchronous: the new packets actually arrive at time slot t , but they are first considered in the time slot $(t+1)$. The queue at a transmitter is modeled as

$$B(t+1) = [B(t) - R(t)]^+ + X(t) \tag{3.25}$$

where $R(t) \in \{0,1\}$ is the service process, only one packet can be transmitted per time slot, $X(t) \in \mathbb{N}$ is the arrival process, $B(t)$ is the length of the queue at the beginning of the time slot t. The operator $[v]^+$ stands $\max(0, v)$.

The buffer status $B(t)$ at each transmitter depends on the packet arrival and service processes. $B(t)$ impacts the activation of the transmitters since the transmitter can only transmit when the buffer is not empty, i.e., $B(t) > 0$. In turn, $B(t)$ impacts the mutual interference $I(t)$. The buffer statuses of the transmitters are interdependent, leading to interacting queues.

Service process The service process depends on the dynamic SINR (3.24). One can assume that when the SINR is larger than a given threshold θ, the transmission succeeds, otherwise the transmission fails. Then, the transmit success probability at time slot t can be defined as

$$p_t = \mathbb{P}(\mathrm{SINR}(t) > \theta) \tag{3.26}$$

The queues are spatially coupled since the mutual interference $I(t)$ directly affects the $\mathrm{SINR}(t)$, the transmit success probability and the service processes of the queues. On the other hand, the queues are temporally coupled since the current buffer status $B(t)$ is affected by the previous service process $R(t)$. Due to the random nature of channel fading and aggregate interference, the service process is dynamic.

Example 3.3. *An example of the correlation between the queues at two transmitters is depicted in Fig. 3.5. Transmitter T1, which serves receiver R1, has a longer transmission link and a shorter interference link than transmitter T2, which serves receiver R2. Conversely, the channel disparity between the two communication links results in different packet service rates. Compared to T2, T1 suffers from more path loss and interference. Correspondingly, given a similar traffic load, T1 tends to vacate its queue more slowly and thus remains active more frequently than T2. On the*

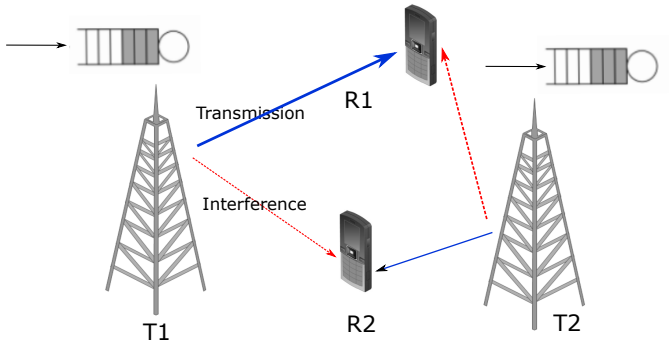


Figure 3.5 – Illustration of the interacting queues.

other hand, the queue lengths of T1 and T2 determine the activation of T1 and T2. In particular, if both transmitters are busy, their transmissions will cause mutual interference, which slows down the departure process. If one of the transmitters has an empty queue, the other receiver's interference is null and hence enjoys a favorable condition. Due to the interacting queues, transmissions in a large-scale system inherently experience spatial and temporal correlation.

3.2.2 Spatio-temporal modeling approach

The previous efforts have typically considered one aspect of the traffic: (1) *Abstraction based on queueing theory*, which evaluates scheduling algorithms and ignores the interaction between the traffic and the SINR statistics, and hence with networks geometry [93, 94, 95]; (2) *Abstraction based on stochastic geometry*, which usually does not consider the temporal arrival process of the traffic and focuses on the throughout in full buffered networks, i.e., each link always has a packet to sent [4, 80, 68]. This last assumption provide a pessimistic view of the aggregate interference as well as some other performance metrics. Moreover, no insights regarding the packet delays can be obtained since the queueing dynamics are ignored.

Existing works on spatio-temporal dynamic models, based on queuing theory and stochastic geometry, can be divided into two categories. The first category is the simultaneous spatio-temporal arrival models of traffic, also called high mobility networks [6, 7, 8]. The second category is the static Poisson networks, where the location of the receiver is fixed during the time evolution [9, 75, 10].

In high mobility networks, the traffic is usually presented by a three-dimensional PPP, where the location of the nodes is modeled by a two-dimensional PPP and the traffic arrivals are modeled by a one-dimensional PPP [6, 7, 8]. These studies characterize the total traffic in each cell. For example, the authors in [6] analyze the stability of high mobility networks. The work in [7] models users as homogeneous spatio-temporal PPPs, and then the traffic arriving at each BS is described by the amount of data available to all users, within the coverage of the relevant BS. The nodes location in the subsequent time slots are considered independent from the location in the current time slot. The work in [8] models the traffic in an uplink single-cell network where the traffic arrives to the cell as a Poisson process. Such models of high mobility networks have two main drawbacks. First, since users are not explicitly characterized in these

models, many network operations, such as user scheduling per cell and traffic offloading between cells, are not well captured. Second, the measures of each individual user, such as throughput and latency, are not well defined.

Compared with high-mobility networks, the static Poisson networks, where the locations of transmitters and receivers are fixed during the time evolution, are challenging to analyze, because the inherent correlations of the interference and signal levels persist among different time slots, due to the static locations of the nodes [9]. A simple way to remove the correlations of the interference is to consider that the interferers are activated independently at each time slot. Thus, the interfering BSs is a randomly thinned PPP from the original PPP, according to a thinning factor [11, 10]. The coverage probability and the delay of the typical UE for heterogeneous cellular networks are analyzed in [11] and [10]. The effects of the thinning factor on the performance metrics of interest are also studied. However, these works ignore the interaction between the queues of different BSs.

The works in [12, 13] pushed further the analysis of the interaction between the dynamicity of the queues and the topology of the network. A traffic-aware spatio-temporal model for IoT devices supported by cellular uplink connectivity has been developed in [12]. A quite complete transmission scheme, i.e., back-off and transmission power, has been proposed using Markov chains whose evolution depends on the queue state of the devices. Thanks to this model, authors studied the tradeoff between the scalability of the network, i.e., supporting as much as possible a high number of devices, and its stability, i.e., the queues are not diverging. However, the performance analysis in this work is based on the first moment measure, i.e., the coverage probability: the variability of the SINR is not captured at particular points in the area (which will be studied in our works). In [13], a novel spatio-temporal mathematical framework is provided to analyze the preamble transmission success probability of a cellular IoT network, where the number of accumulated packets in the queues is approximated by a Poisson distribution. The relationship between transmit success probability in each time slot and the length of the remaining blocked packets is described by an iterative algorithm. However, the theoretical findings of this article can be verified when the number of time slots is relatively low, and this approximation will have a deviation over time. Fig. 3.6 shows the CDF of the number of accumulated packets obtained by simulation and Poisson approximation in [13, Theorem 2], for the 1st, the 3rd, the 5th and the 15th time slot. It can be observed that after 3 time slots, the assumption that the number of accumulated packets follow a Poisson law becomes less accurate.

Chapter 4 is devoted to develop a comprehensive approach to handle the interaction between the coverage probability and the queueing state evolution, using discrete time Markov chain. A simple model is considered, but contrarily to the state of the art, closed-form expressions are given that make the bridge between the coverage probability and the fraction of active base stations under conditional stable state. Besides, to the best of our knowledge, all the works mentioned before studied the coverage probability with infinite queue lengths. However, the packet loss probability is also an important performance measure needed for the design of telecommunication networks. The quantity of interest is the probability of a new packet is dropped when the buffer has a finite size. We also develop a tractable mathematical model to analyze the coverage probability and packet loss probability in a downlink cellular

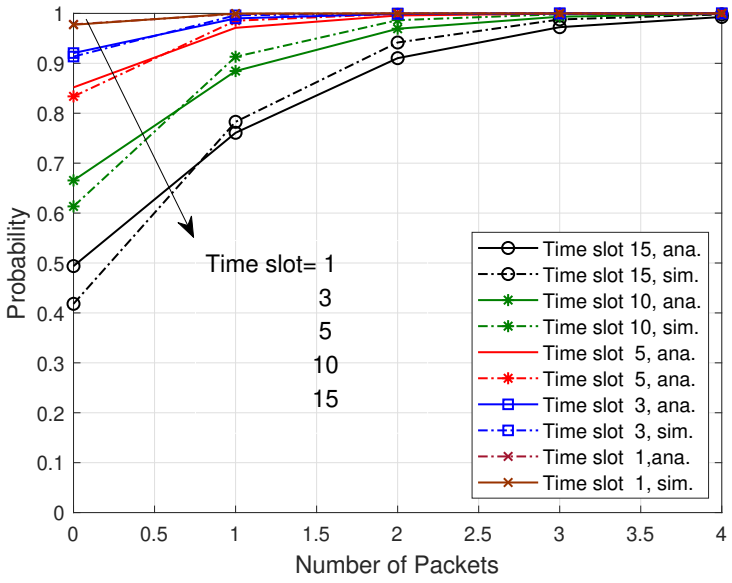


Figure 3.6 – CDFs of the number of accumulated packets between Poisson, analytical and simulation.

network considering the buffer restriction.

It is worth mentioning that there are also some performance metrics which have been studied for spatio-temporal modeling. For example, the SINR gain is used to quantify the impact of the target model relative to the reference model on the SINR distribution [96]. The joint success probability is defined as the probability that k correlated time transmissions succeed. The joint success probability can refer to temporal, spatial, or spatio-temporal transmission events [97, 98, 99]. The interference correlation coefficient evaluates the correlation degree of interference at two locations or time slots [100].

3.2.3 ϵ-stable region approach

A review of the literature shows that the spatio-temporal traffic modeling has been particularly exploited to evaluate important metrics of interest: the ϵ-stable region [29]. Unlike the stable region, that is based on the first moment measure, the ϵ-stable region relies on the moment generating function of the SINR. The ϵ-stable region refers to the set of arrival rates such that the proportion of unstable queues in the network is below ϵ. The characterization of the ϵ-stable region relies on the use of meta-distribution defined in (3.8).

Definition 3.3 (ϵ-stable region [29])**.** *Let ξ be the arrival rate of the traffic. For any $\epsilon \in [0,1]$, the ϵ-stability region $\mathscr{S}_\epsilon$ is defined as*

$$\mathscr{S}_\epsilon = \left\{ \xi \in \mathbb{R}^+ : \mathbb{P}\left[\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{P}(SINR(t) > \theta | \Phi) \leq \xi \right] \leq \epsilon \right\} \tag{3.27}$$

$\mathbb{P}\left[\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{P}(\text{SINR}(t)>\theta|\Phi)\le\xi\right]$ is the probability that the queue at the typical transmitter is unstable. This probability is obtained by averaging over the point process. The ϵ-stable region is the set of arrival rates such that the proportion of unstable queues in the network is not larger than ϵ.

We define ξ_c as $\xi_c = \sup\mathscr{S}_\epsilon$. Deriving the ϵ-stable region $\mathscr{S}_\epsilon$ is equivalent to obtaining the critical arrival rate ξ_c. The network is ϵ-stable if and only if the arrival rate is lower than the maximal one, i.e., $\xi \le \xi_c$.

Example 3.4. *Let us consider a simple system with only the typical user, i.e., there is no other users in the network. The success probability of the typical transmitter is* $p_s \triangleq \mathbb{P}(SNR>\theta) = \mathbb{P}(\frac{h_0 r_0^{-\alpha}}{\sigma^2}>\theta)$. *We assume that* $h_0 \sim \exp(1)$ *is the small-scale Rayleigh fading. By averaging over* h_0, *the success probability of the typical transmitter is* $\exp(-\sigma^2\theta r_0^\alpha)$.

Loynes' theorem [101] mentioned that for a point-to-point system with random arrival and departure processes, the stable region requires that the service rate be larger than the arrival rate. If the distance of the desired link r_0 *is fixed, the stability condition follows*

$$\xi \le \xi_0 \triangleq \exp(-\sigma^2\theta r_0^\alpha) \tag{3.28}$$

If r_0 *is random and follows the probability density function* $f_{r_0}(x) = 2\pi\lambda r e^{-\lambda\pi r^2}$, *the stable region has the following expression*

$$\begin{aligned}\mathscr{S}_\epsilon &= \left\{\xi\in\mathbb{R}^+ : \mathbb{P}\left[\exp(-\sigma^2\theta r_0^\alpha)<\xi\right]\le\epsilon\right\}\\ &= \left\{\xi\in\mathbb{R}^+ : \int_{\left(\frac{\ln\xi}{\sigma^2\theta}\right)^{\frac{1}{\alpha}}}^{\infty} 2\pi\lambda r e^{-\lambda\pi r^2} dr \le \epsilon\right\} \qquad (3.29)\\ &= \left\{\xi\in\mathbb{R}^+ : \exp\left(-\lambda\pi\left(\frac{\ln\xi}{\sigma^2\theta}\right)^{\frac{2}{\alpha}}\right)\le\epsilon\right\} \qquad (3.30)\end{aligned}$$

Fig 3.7 plot the the critical arrival rate ξ_c *in (3.30) where the small-scale fading is Rayleigh and the PDF of the distance from the receiver to the transmitter follows* $f_{r_0}(x) = 2\pi\lambda r e^{-\lambda\pi r^2}$. *The critical arrival rate, such that the probability of the queue is unstable, is below* 40% *is* $\xi_c = 0.5$ *packet/slot when* $\sigma^2 = 0.5$, *and* $\xi_c = 0.25$ *packet/slot when* $\sigma^2 = 1$.

Example 3.4 provided a noise-limited case where only the typical link is considered and the interference is ignored. However, the traffic conditions are more complicated in a large-scale network with multiple queues, since the service rate depends on the state of all the transmitters in the network. Then, sufficient and necessary conditions for system stability have been introduced and studied in [29], and meta-stability in [8], where the network appears stable for possibly a long time and then suddenly exhibits instability. Remarkably, the stochastic geometry and queueing theory have been merged to give sufficient and necessary conditions for the stability of interacting queues [29]. However, this work has considered a peer-to-peer network with constant link distances. And, the bounds are not very tight especially under some network configurations. On the other hand, the stability and meta-stability of uplink random access networks considering the data traffic have been studied in [8]. The analysis in

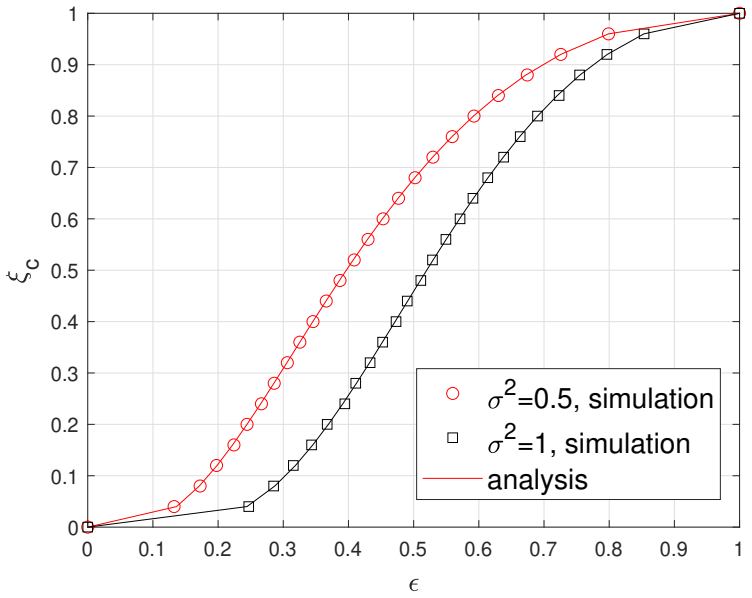


Figure 3.7 – The ϵ-stable region of the typical link, without interference.

this work was based on double-stochastic networks, i.e., space-time Poisson call arrivals. A single cell network has been considered in this work, thereby ignoring the interaction between queues of different BSs.

In Chapter 5, we provided the closed-form of the upper and lower bounds of ϵ-stable region in the case of random link distances. Besides, we provide an approximate definition of the ϵ-stable region and derive the closed-form accordingly. In particular, the interaction between transmit success probability and queueing state evolution is handled thanks to a discrete time Markov chain (DTMC). The result reveals that the proposed approximation is tighter than the bounds and easy to evaluate.

3.2.4 Summary

The introduction of queueing theory in a stochastic geometry approach allows to assess important network performance measures such as the average delay or stability. The analysis remains however challenging due to the complex interaction between the packet arrival rate process and the service rate depending on the coverage probability that in turn depends on the interference in the networks and all the queue states of transmitters. To ensure tractability of the results, a simple transmission strategy where all BSs are active and transmitting when the buffer is not empty is discussed in Chapter 4 and 5. However, even considering such a simple transmission strategy, complex mathematical techniques are widely used to obtain closed-form performance metrics. For example, the PGFL theorem are widely applied to calculated the expectation of interference. The moment measures are generally used to calculate the meta distribution and ϵ-stable region.

On the other hand, in the dynamic downlink cellular networks, a question naturally arises whether BSs can make the best decision based on the dynamic network environment, e.g., to give the maximum stable region at the minimum transmitting cost with respect to the channel condition? However, it is non-trivial to derive closed-form strategies and the stable region as well accordingly. In order to pursue adaptive transmission strategies according to dynamic environment, we applied the Reinforcement learning in the Chapter 6.

3.3 RL applied in traffic-aware systems

While stochastic geometry provides a powerful model-driven approach that aims at evaluating performance metrics based on the conventional probabilistic models [42], the reinforcement learning is a class of learning processes in which an agent can periodically make decisions, observe the results, and then automatically adjust its strategy to achieve the optimal policy [102]. A RL model can be abstracted as agents interacting with their environment, performing actions and learning through the *trial-and-error* method, to achieve long-term goals. It can be efficiently applied to a wireless communication system when: *i*) The mathematical modeling of the environment is too complex to be implemented in an agent; *ii*) An accurate mapping between the network features and its performance is needed by the agent; *iii*) The desired outcome of the learning can be described as a scalar reward.

We start with an example of how RL can be applied in a point-to-point communication system to handle the temporal traffic. The model is abstracted from [14, 15]. Then, the approaches related to the application of the RL in the wireless network, especially in cellular networks, are summarized. In Chapter 6, the SG and RL are jointly used to tackle the dynamic cellular network, allowing to obtain the optimistic transmission strategy which can maximize the long-term returns accordingly.

3.3.1 An example of traffic management using RL

Let us consider a point to point communication system over a block fading channel. The channel gain is assumed to be constant on a slot of time duration Δt and changes from one slot to another one according to the distribution $P_{H(t+1)|H(t)}(h'|h)$, with $H \in \mathscr{H}$ and where $\mathscr{H}$ is assumed to be a finite countable set. At each time slot, a certain number of packets is generated by the transmitter and stored in a buffer, waiting for their transmission. The transmitter aims at sending the packets at lowest power consumption with an average delay cost constraint.

At each time slot t, $X(t) \in \mathscr{X}$ new packets are generated and stored in the buffer before being transmitted. $\{X(t)\}$ are i.i.d. random variables with Bernoulli distribution with intensity ξ, the arrival rate. $R(t)$ packets are successfully transmitted and removed from the buffer with some success probability $f_s(h)$. This success probability increases with a better channel state. When a packet is not received successfully, the packet is kept in the buffer for a later retransmission via automatic repeat request (ARQ) control. The buffer state $B(t) \in \mathscr{B} = \{0, 1, \cdots, B_{\max}\}$ represents the number of packets stored in the queue at time t and $B_{\max}$ is the maximal buffer size. The state evolution can be described by a Markov chain as in (3.25).

According, the *state space* can be defined as the space containing the channel state, the buffer state, and the new packet arrival state, i.e., $\mathscr{S} = \mathscr{H} \times \mathscr{B} \times \mathscr{X}$.

At each time slot, the transmitter has to decide whether to transmit a packet, unless the buffer is empty. The *action space* is then described as $\mathscr{A} = \{0, 1\}$, where 0 stands for not transmitting, and 1 for transmitting a packet.

One may be interested to transmit packets with the minimal transmission cost while limiting the waiting time in the buffer. In that case, the *long-term return* can be expressed w.r.t. two cost functions, i.e., the transmit cost and the waiting time cost functions [14], c and w. The non-negative waiting time cost, which is applicable when the buffer is not empty, i.e., when $B(t) > 0$, is $w(\cdot) : \mathscr{A} \rightarrow \mathbb{R}_+$. The non-negative transmission cost depend on the action and the current channel state $H(t)$ by the transmission cost function $c(\cdot, \cdot) : \mathscr{A} \times \mathscr{H} \rightarrow \mathbb{R}_+$.

Policy The transmission scheduling policy consists in mapping the system state to an action at each time slot t. Hence a desired policy should solve an optimization problem. From the cost functions defined previously, the expected discounted power and waiting time costs, given an initial state $s_0 \triangleq S(0)$, are defined as:

$$C_\pi(s_0) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \eta^t c(A(t), S(t)) | S(0) = s_0 \right] \tag{3.31}$$

$$W_\pi(s_0) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \eta^t w(A(t), S(t)) | S(0) = s_0 \right] \tag{3.32}$$

with $\eta \in [0, 1]$ the discounting coefficient. The expectation is taken over the distribution of the policy and the dynamic of the underlying MDP. The problem to find the minimal power consumption while limiting the waiting time cost can be formally described as [14]

$$\min_{\pi \in \Phi} C_\pi(s_0) \ s.t. W_\pi(s_0) \leq \delta \forall s_0 \in \mathscr{S}. \tag{3.33}$$

Note that minimizing the waiting time cost under a total power budget, as studied in [15], leads to an equivalent strategy. The problem relies to a constrained optimisation problem with unknown dynamics. One can combine the power and waiting time cost function c and w in (3.31) and (3.32) in a dual Lagrangian expression such that $l(a, s; \lambda) = c(a, s) + \lambda w(a, s)$. One can apply the Q-learning algorithm detailed in Section 2.3.3 by replacing R in (2.42) by the Lagrangian cost to obtain the optimal policy.

Experiments We illustrate the performance of Algorithm 1 via numerical example in a video transmission application, as in [14]. We assume that *i*) the time is divided into slots of size 0.5ms, so that at each time slot exactly one packet can be transmitted, *ii*) the block fading channel is modeled by a three-state Markov chain, i.e., $\mathscr{H} = \{h_1, h_2, h_3\}$ with the transmission probability matrix

$$A^c = \begin{pmatrix} 0.85 & 0.15 & 0 \\ 0.15 & 0.7 & 0.15 \\ 0 & 0.15 & 0.85 \end{pmatrix} \tag{3.34}$$

The success probability $f_s(h)$ are given by $f_s(1,h_0)=0.3$, $f_s(1,h_1)=0.5$, and $f_s(1,h_3)=0.95$. We assume that the user has a buffer of size $B=50$, and the average arrival rate is $\xi=0.25$. Let the transmission cost be $c(h_0,a)=1.7a$, $c(h_1,a)=1.7a$ and $c(h_2,a)=0.2a$ for all $a\in\mathscr{A}$, and $w(1)=0$, $w(0)=0.6$. Fig 3.8 illustrates the average performance, i.e., (3.33) with Lagrange factor $\lambda=2$, of ϵ-greedy methods mentioned in (2.44). The greedy method, i.e., $\epsilon=0$, performs significantly worse in the long run because it often gets stuck performing locally optimal actions. The ϵ-greedy method eventually performs better because it continues to explore, and to improve the chance of recognizing the optimal action. The $\epsilon=1/t$ method finds the optimal action earlier, and eventually performed as $\epsilon=0.1$ method on both performance measures.

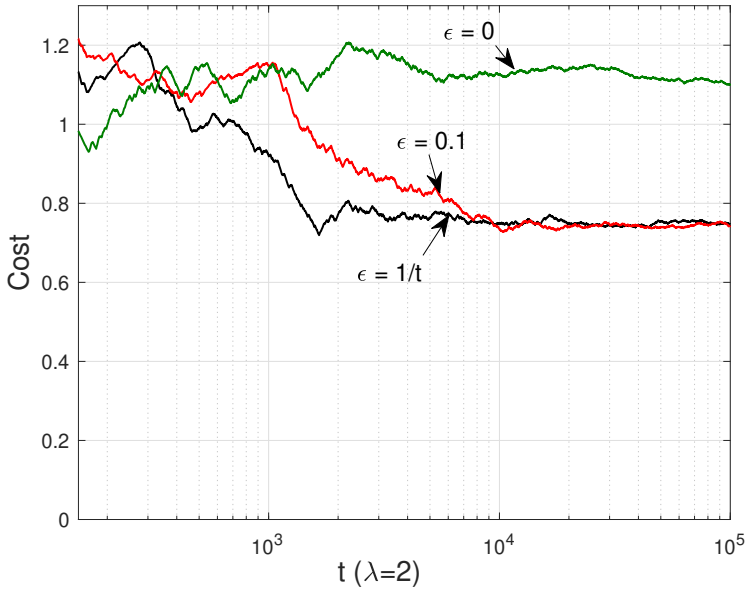


Figure 3.8 – The Lagrangian cost estimated by Q-learning algorithm (Algorithm 1) under different ϵ-greedy factors.

Two extensions of this point-to-point example are as below: *i*) The above Markov decision process can be generalized from the binary action space to the general case of a finite action space, i.e., $\mathscr{A}=\{a_0,a_1,\cdots,a_M\}$, hence the multi-power consumption levels can be considered rather than simply turn on and off the transmitter [103]; *ii*) An alternative definition of waiting cost can be $w(a,s)=\beta\mathbb{1}\{b(t+1)>B_{\max}\}+(b(t)-r(t))^+$, where the first term represents the cost to be in overflow with β constant, the second term is the holding cost, i.e., the cost for keeping $b-r$ in the buffer if the transmission is successful. Hence the buffer waiting cost depends on the buffer state [15]; *iii*) Q-learning does not assume any knowledge of the dynamics of the underlying MDP and therefore slows down the convergence time. On the other hand, the exploration time component is the basis of Q-learning, and when there are a large number of state and action combinations, the convergence time can be too long. However, in wireless communication, some dynamics may not be completely unknown. The authors in [104, 15] use the concept of post-decision states to base the actions taken on states that consider only

known dynamics.

3.3.2 Approach that combines RL and SG

Reinforcement learning tracks optimistic policies for wireless networks in uncertain and stochastic environments by providing a Markov decision process framework [105, 3]. In the last decade, the reinforcement learning has been combined with various complex network models taking into account dynamic spectrum access, user association, caching and offloading, to cite a few [106, 107, 108]. However, since the nature of the problems studied with stochastic and reinforcement learning are so fundamentally different, it is rare to find common ground where the strength of these tools can be jointly leveraged. On one hand, the RL is a learning process in which an agent can periodically make decisions, observe the results, and then automatically adjust its strategy to achieve the optimal policy [37, 38]. On the other hand, the basic premise of the SG is to provide a unified mathematical framework to model large-scale random networks. The key performance metrics, such as interference, coverage, and rate, can be characterized as exact and tractable expressions [27, 4].

In chapter 6, we concretely demonstrate that these two mathematical tools can be jointly applied to a class of problems. The spatial locations determine how the BSs interfere, and the temporal traffic dynamic affects the queue evolution and thus forms an MDP. The agent tries to achieve a goal despite uncertainty about its environment. Different from the traditional reinforcement learning application systems, the main difficulties to leverage the SG and RL together are listed below:

- The traditional RL algorithms are applied in discrete state and action spaces. Most of the researches studying traffic load consider the SINR or the SNR as a finite countable set of values [14, 15]. However, the SINR modeled by the SG are continuous random variables, which will lead to continuous state spaces and action spaces. To overcome this problem, we discretized the SINR in the subsequent model construction in Chapter 6.

- The dimensionality problem usually arises in large-scale networks. Since the Q-learning does not assume any knowledge about the dynamic of the underlying MDP, the exploration time part, which is fundamental in Q-learning, slows down the convergence time due to the large number of combination of states and actions. To solve the computationally expensive problem with the large action space, multi-agent deep RL method is proposed in [16, 106]. Instead of using a table for updating the action-state value function, as in classical Q-learning algorithm in Section 2.3.3, one may use deep reinforcement learning searching for approximating the action-state values with a suitable function $q_\theta : \mathscr{S} \times A \to \mathbb{R}$ with the vector parameter θ. On the other hand, the authors in [104, 15] use the concept of the post-decision states, introducing an *a priori* knowledge about the environment's dynamics to develop more efficient learning algorithms.

- Last but not least, the fact that UEs and BSs are generated according to some stochastic processes will lead to a random dimension of state spaces. To overcome this problem, we first fixed the point processes to obtain the optimistic strategies and performance associated with a PPP realization since we considered a static Poisson network. Then,

we take the expectation over the point processes to obtain the system-level average performance.

The references that are most closely related to this part of our works are [14, 15]. In [14], the problem is subject to an average delay constraint for a point-to-point communication system. Specifically, the optimization problem is to minimize an average transmission cost devising the transmission scheduling policies according to the channel states. In [15], the energy-efficient point-to-point transmission of delay-sensitive data over a fading channel is studied. The authors formulate the stochastic optimization problem as an MDP and solve it online using reinforcement learning. However, both studies consider point-to-point problems, thus the interference from other transmitters are not taken into account.

On the other side, the RL has been applied to cellular networks to solve resource allocation or rate adaptation problem, e.g., [16, 17]. In [16], a distributed optimization method based on a multi-agent RL is developed to solve user association and resource allocation. In [17], the RL is applied for inter-cell power control and rate adaptation in the downlink of a radio access network. However, the performance analysis in these researches are based on a regular cellular network. That means, only a small finite number of interfering BSs are considered, without considering the effect of the random distribution of aggregated interference.

In the last two years, a few works have considered joint stochastic geometry as well as reinforcement learning tools to solve the problem of resource allocation or user association in stochastic networks. In [18], the authors consider a model for a NOMA-based downlink Fog-RAN to meet the requirements of high spectral efficiency and huge device connectivity. Stochastic geometry is used to capture the impact of the random distribution of users. From the user perspective, two types of user association are available: Fog-computing-based access points and remote radio heads. RL techniques are used to develop user association algorithms to maximize the total spectral efficiency in the network.

In [20], the authors present a multi-operator (OP) sharing problem for small cell network deployments, with a focus on the user scheduling problem. The optimal problem is to maximize the social welfare of the network, i.e., how to share resources to the small cells in the network to maximize the overall weighted sum rate. However, the analysis is based on a single time slot of the overall system, thus the traffic is unaware in this system. Similarly, [19] considered the resource allocation problem to maximize the D2D users total throughput while keep the interference to the cellular user under limits. In addition, the paper [38] investigates how to deploy unmanned aerial vehicles (UAVs) in three-dimensional space to maximize network utility. In this work, the UAV is modeled using a binary Poisson process, and the RL algorithm is used to solve the deployment problem. However, the works mentioned above concentrated the user allocation problem or resource allocation problems, none of them consider a stochastic network with traffic-aware.

In Chapter 4 and 5, mathematical frameworks based on the SG and queueing theory are provided to analyze the coverage probability as well as the ϵ-stable region. In order to keep mathematically tractability, these works are based on a basic communication strategy where BS keeps transmitting when the buffer is not empty. Since the optimal transmission scheduling problem is naturally an MDP due to the uncertainty in the packet service process and the Markov property of the queue evolution, we advance a new approach to analyzing the system

with the adaptive transmission by applying reinforcement learning in Chapter 6. In particular, the optimal transmission scheduling problem is formulated as a constrained Markov decision process (CMDP), where the constraint is the waiting cost. The numerical results show that the RL-based policy can hold the same stable region as a greedy algorithm but with a lower transmitting cost.

3.4 Conclusion

This chapter first describes how Poisson point processes facilitate the modeling and analysis of large-scale wireless networks. We introduce the propagation model, including the small-scale fading and the large-scale fading. Based on this, we discuss two important metrics, namely the coverage probability and the meta-distribution. The introduction of queueing theory in a stochastic geometry approach allows assessing important network performance measures, such as the average delay or the stability. However, the analysis remains challenging due to the complex interaction between the packet arrival rate and the service rate, depending on the coverage probability, which in turn depends on the interference and on all the queue states. As we proposed in Section 3.2.2, the existing literature either ignores the interaction between the queues or analyzes it by approximation. A tractable mathematical framework to analyze the transient and stationary metrics is still lacking, which will be studied in Chapter 4. In Chapter 5, we will fully characterized the ϵ-stable region.

On the other hand, a question naturally arises in the traffic-aware networks whether BSs can make the best decision based on the dynamic network environment, e.g., to give the maximum stable region at the minimum transmitting cost, depending on the channel condition. However, it is non-trivial to derive closed-form strategies, and the stable region accordingly. In Chapter 6, we propose a reinforcement learning framework to compute the optimal transmission policy. We formulate a constrained optimization problem to minimize the long-term transmission cost with delay constraints. The problem is then a constrained Markov decision process due to the dynamic evolution of the queue. Based on reinforcement learning and stochastic geometry tools, we analyze the stable region of the network based on different transmission schemes and show that there is a tradeoff between the stable probability and the transmission cost depending on the traffic intensity.

Chapter 4

Coverage Analysis in Dynamic Downlink Cellular Networks

4.1 Introduction

Stochastic geometry is an effective theoretical tool to model the locations of base stations and user equipments by considering a real deployment, as realizations of a class of random point processes, as discussed in Section 3.1. However, the majority of the existing literature heavily relies on the assumption that each transmitter's buffer is never empty, which does not characterize a random traffic [4, 68].

To analyze the impact of the traffic in large-scale networks, we introduced the SINR model with traffic-aware in Section 3.2.1. As discussed, the main difficulty in random traffic characterization comes from the correlation between the buffer states of different transmitters, that leads to interacting queues. In recent years, some works have considered wireless systems with spatio-temporal models, as detailed in Section 3.2.2. However, the exciting literature either ignores the interaction between the queues, e.g., [78, 10], or studies the network stability by providing untight bounds [9]. Moreover, existing literature usually does not take into account the buffer size. A tractable mathematical framework is still lacking to characterize the SINR distribution and meta distribution in large-scale networks with traffic-aware.

In this chapter, we study the coverage analysis of dynamic downlink cellular networks, considering two application scenarios: infinite buffer size and finite buffer size [1]. We propose tractable mathematical models to analyze the coverage probability while considering queue dynamics. We use different DTMC (introduced in Section 2.2) to handle the interaction between the coverage probability and the queueing state evolution, for different application scenarios. A simple model is considered, but contrarily to the state of the art, closed-form expressions are given that link the coverage probability to the fraction of active BSs, under the conditional stable state. Besides, we analyze the average queue delay in the infinite buffer case, and the packet loss probability in the finite buffer case.

[1] These works led to publications [C1], [C2], see § 1.3.

4.2 Network model and assumptions

We consider the system model given in Section 3.1.1. The BSs are spatially distributed in $\mathbb{R}^2$, following an homogeneous PPP $\Phi = \{x_i\}_{i\in\mathbb{N}}$ with intensity λ. The UEs density is assumed to be high enough such that each BS has at least one user associated to it. Besides, each UE is served by its nearest BS. In the network, all BSs are assumed to transmit with constant normalized power in the same bandwidth.

We adopt the block fading propagation model introduced in Section 3.1.2. The channel between any pair of transmitter and receiver is assumed to be i.i.d. and *quasi-static*, i.e., the channel is constant in one transmission time slot and varies independently between the time slots. Specifically, the small-scale fading is modeled by the Rayleigh distribution, and the large-scale fading follows the power law in (3.3).

The time is divided into very short equal intervals where only one packet arrives or leaves from the BS queue at a time. We assume that the packet size is fixed and that it takes exactly one time slot to be transmitted. Each BS maintains an independent buffer to store the generated packets. The packet arrival process at each transmitter is assumed to be a Bernoulli process with intensity $\xi \in [0,1]$ at each time slot. Contrarily to the arrival process, the departure process cannot be fixed *a priori*. It is characterized according to the time-dependent SINR distribution, as shown in Section 3.2.1. If the received SINR exceeds the threshold θ, the packet is transmitted successfully and removed from the queue. Otherwise, the transmission fails, and the packet remains in the queue waiting for retransmission in the next time slot, until it is successfully received. There is no limit to the number of possible retransmissions in this manuscript. However, in practice, the number of needed retransmissions remains low when the system is stable as we will see later. The buffer length at each BS can be infinite and finite, that will be dealt with in Section 4.4 and 4.5, respectively.

The realization of the point process Φ is conditioned on a full activity of the BS at position x_0. Then, the relevant probability measure is the reduced Palm probability, denoted as $\mathbb{P}^{x_0}$. Furthermore, we define Φ_t to be the set of BSs that are transmitting in the time slot $t \in \mathbb{N}$.

4.2.1 SINR model

The received SINR at time slot t experienced by the typical UE is

$$\gamma_t = \frac{H_{x_0,t}\,\|x_0\|^{-\alpha}}{\sigma^2 + \sum\limits_{x\in\Phi\setminus x_0} H_{x,t}\,\|x\|^{-\alpha}\,\mathbb{1}(x\in\Phi_t)} \tag{4.1}$$

where $H_{x_0,t}$ and $H_{x,t}$ are the exponential channel gains between the typical UE and its desired BS located at x_0, and with the interfering BS located at x at time slot t, respectively. α is the path loss exponent, and σ^2 denotes the power of additive white Gaussian noise. $\mathbb{1}(\cdot)$ is the indicator function, which is equal to 1 or 0 when the transmitter is on or off, respectively.

Moreover, we note q_t as

$$q_t = \mathbb{P}_{\mathbb{1}(x\in\Phi_t)}(\mathbb{1}(x\in\Phi_t) = 1) \tag{4.2}$$

which can be seen as *i*) the fraction of active interfering BS at time slot t; *ii*) the probability that a randomly chosen BS is active at time slot t.

4.3 Dynamic coverage probability

Considering that the typical UE receives data at time slot t, i.e., its associated BS in x_0 is always active, the dynamic coverage probability is defined as [9]

$$p_t \triangleq \mathbb{P}^{x_0}[\gamma_t > \theta],\ \theta \in \mathbb{R}^+ \tag{4.3}$$

Theorem 4.1. *The dynamic coverage probability has the following expression*

$$p_t = 2\pi\lambda \int_0^\infty e^{-\sigma^2\theta r^\alpha} e^{-\pi\lambda r^2(1+q_t\rho(\alpha,\theta))} r\,dr \tag{4.4}$$

where $\rho(\alpha,\theta) = \int_1^\infty [1+u^{\frac{\alpha}{2}}\theta^{-1}]^{-1} du$.

Proof. See Appendix A.1. □

Theorem 4.1 quantifies how the coverage probability behaves at a given time slot and depends on the traffic. It illustrates that the queues states affect the coverage via the parameter q_t. As q_t depends on the time t, the coverage probability is time depending. As q_t decreases, there are fewer active interferers in the network, and therefore, the total interference decreases, and p_t increases. The description of q_t depends on the specific queueing model, i.e., the finite or infinite buffer size. In the particular case of the interference-limited network, Theorem 4.1 takes the following form.

Corollary 4.1. *In an interference-limited network, i.e.* $\sigma^2 \to 0$, *we have*

$$p_t = \left[1 + \int_1^\infty \frac{q_t}{1+u^{\frac{\alpha}{2}}\theta^{-1}} du\right]^{-1} \tag{4.5}$$

and for a path loss exponent $\alpha = 4$, *we have*

$$p_t = \left[1 + q_t\sqrt{\theta}\tan^{-1}(\sqrt{\theta})\right]^{-1} \tag{4.6}$$

At the end of this section, we introduce the concept of network stability. The network is called stable if the number of active transmitters converges to a limit regardless of the network initial condition [8].

Definition 4.1. *Let* $\{\Phi_t\}_{t=0,1,\cdots}$ *be the series of the point process along the time. Moreover, let* $\{p_t\}_{t=0,1,\cdots}$ *be the stable coverage probability of the network* $\{\Phi_t\}$ *at time* t. *Under the conditions where the limits of such series exist, stable coverage probability is*

$$p = \lim_{t\to\infty} p_t, \tag{4.7}$$

and $q = \mathbb{P}_{\mathbb{1}(x\in\widetilde{\Phi})}(\mathbb{1}(x \in \widetilde{\Phi}) = 1)$, *where* $\tilde{\Phi}$ *is the limit of the PPP series which represents the point process where the activity of base stations does not evolve with time.*

Lemma 4.1. *Considering* $\lim_{t\to\infty} q_t = q$*, and let* $\rho(\alpha,\theta) = \int_1^\infty [1+u^{\frac{\alpha}{2}}\theta^{-1}]^{-1} du$*, the stable coverage probability is a function of* q *and can be expressed as following*

$$p = 2\pi\lambda \int_0^\infty e^{-\sigma^2\theta r^\alpha} e^{-\pi\lambda r^2(1+q\rho(\alpha,\theta))} r\mathrm{d}r \tag{4.8}$$

Proof. Considering $\lim_{t\to\infty} q_t = q$, we have

$$\begin{aligned} p &= \lim_{t\to\infty} 2\pi\lambda \int_0^\infty \exp\left(-\sigma^2\theta r^\alpha\right) \exp\left(-\pi\lambda r^2\left(1+q_t\rho(\alpha,\theta)\right)\right) r\mathrm{d}r \\ &\overset{a}{=} 2\pi\lambda \int_0^\infty \exp\left(-\sigma^2\theta r^\alpha\right) \exp\left(-\pi\lambda r^2\left(1+\lim_{t\to\infty} q_t\rho(\alpha,\theta)\right)\right) r\mathrm{d}r \\ &= 2\pi\lambda \int_0^\infty \exp\left(-\sigma^2\theta r^\alpha\right) \exp\left(-\pi\lambda r^2\left(1+q\rho(\alpha,\theta)\right)\right) r\mathrm{d}r \end{aligned} \tag{4.9}$$

where (a) follows the fact that p_t is non-negative and continuous. Let $A(r) = e^{-\sigma^2\theta r^\alpha}$, and $g = A(r)e^{-\pi\lambda r^2}$, we have $\left|A(r)e^{-\pi\lambda(1+q_t\rho(\alpha,\theta))}\right| \le g$, $\forall\theta\in\mathbb{R}^+, t\in\mathbb{N}$. Since g is integrable, by the dominant convergence theorem [109], the result follows. □

4.4 Coverage analysis with infinite buffer

In this section, we analyze the coverage probability and queue delay in downlink cellular networks when the buffer at each BS has infinite size. Particularly, we develop a comprehensive approach to handle the interaction between the coverage probability and the queueing state evolution, thanks to a DTMC. We also denote the explicit upper and lower bounds on the dynamic coverage probability. At the end, we characterize the queue delay of a randomly chosen BS when the DTMC works at the stationary regime.

4.4.1 Stable coverage probability

Theorem 5. *Under the prescribed system assumption, the stable coverage probability with infinite buffer size is given by*

$$p(\theta,\xi) = 2\pi\lambda \int_0^\infty e^{-\sigma^2\theta r^\alpha} e^{-x\lambda r^2(1+q\rho(\alpha,\theta))} r\mathrm{d}r \tag{4.10}$$

and the active probability of a randomly chosen BS is

$$q = \begin{cases} \xi/p, & \text{if } p > \xi, \\ 1, & \text{if } p \le \xi. \end{cases} \tag{4.11}$$

Proof. We consider the scenario where the arrived packets at each BS are stored in a buffer, with an infinite size, until their successful transmission. The number of packets in the queue of a randomly chosen BS is modeled as a birth and death process that can be represented with the DTMC with infinite states, given in Fig. 4.1.

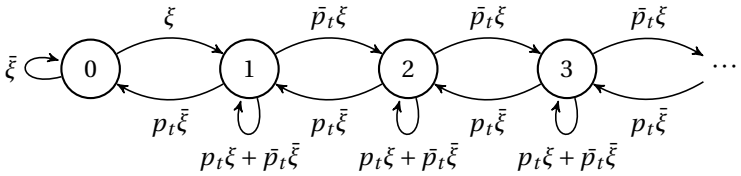


Figure 4.1 – DTMC model of typical BS with infinite buffer length.

In Fig. 4.1, $\bar{a} = 1 - a$ with $a \in \{p_t, \xi\}$, where p_t is the successful transmission rate and ξ the arrival rate of the packets. Each state is the number of packets in the queue at a given time slot. The state 0 represents the empty buffer event. When the buffer is in this state, the transmitter remains silent. The number of packets in the queue can be characterized by the stationary distribution of the DTMC in Fig. 4.1. The transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} \bar{\xi} & \xi & 0 & 0 & 0 & \cdots \\ p\bar{\xi} & \bar{p}\bar{\xi} + p\xi & \bar{p}\xi & 0 & 0 & \cdots \\ 0 & p\bar{\xi} & \bar{p}\bar{\xi} + p\xi & \bar{p}\xi & 0 & \cdots \\ 0 & 0 & p\bar{\xi} & \bar{p}\bar{\xi} + p\xi & \bar{p}\xi & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{4.12}$$

For a stationary Markov chains, we have [35]

$$\boldsymbol{x}\mathbf{P} = \boldsymbol{x}, \ \boldsymbol{x}\boldsymbol{e}^{\mathrm{T}} = 1 \tag{4.13}$$

where $\boldsymbol{x} = [x_0, x_1, x_2, \cdots, x_i, \cdots]$ is the row vector that contains the stable-state probabilities, in which x_i denotes the probability of being in the state with i packets. i.e., $x_i = \mathbb{P}[X = i]$, and $\boldsymbol{e}^{\mathrm{T}}$ is a column vector of ones with the proper length, i.e., infinite length. After algebraic manipulation, the final expression when DTMC works at stable state is (see Appendix A.2)

$$x_0 = \frac{p - \xi}{p}, \forall \ p > \xi \tag{4.14}$$

$$x_i = R^i \frac{x_0}{\bar{p}}, \text{ where } R = \frac{\xi\bar{p}}{\bar{\xi}p}, \ \forall \ i \in [1, +\infty) \tag{4.15}$$

Since x_0 is the probability of having an empty buffer and therefore causing the BS to be inactive, the activity probability of the typical BS is

$$q = 1 - x_0 = \frac{\xi}{p}, \forall \ p > \xi \tag{4.16}$$

If $p \leq \xi$, then all states are transient and BS are always active, i.e. $q = 1$.By substituting (4.4) and (4.14)-(4.16), we obtain the result. □

According to (4.10) and (4.11), the interdependence between q and p shows the relationship between the queue and the stochastic geometry in the analysis. According to the relative

values of p and ξ, a randomly chosen BS has a probability of ξ/p to be active if its arrival rate is less than the departure rate, and it is always active in the opposite case. The computation of the probability q is performed dynamically with Algorithm 2. It is important to note that, when $p \leq \xi$, $q = 1$ and all the buffer lengths grow up to infinity, the DTMC are not stable. The stable coverage probability (4.10) can however be defined but at the cost of infinite queue lengths or dropped packets.

Algorithm 2 Iterative algorithm for computation of p and q of Theorem 5.

Initialize $q_1 \in (\xi, 1)$, $q_0 = 0$, $i = 0$, $\epsilon \ll 1$
while $|q_{i+1} - q_i| \geq \epsilon$ **do**
  $i \leftarrow i + 1$, $q \leftarrow q_i$, $p \leftarrow p(\theta, \xi)$ (4.10)
  if $p > \xi$ **then**
    $q_{i+1} \leftarrow \xi/p$
  else
    $q_{i+1} \leftarrow 1$
    break
  end if
end while
Return $q \leftarrow q_{i+1}$ and $p \leftarrow p(\theta, \xi)$

4.4.2 Upper and lower bounds

Simulation results will show that this stability behaves between two extreme cases that are summarized in the next lemma.

Lemma 4.2. *Considering the depicted downlink cellular network, the coverage probability can be bounded as follows*

$$p_{\mathrm{l}} \leq p \leq p_{\mathrm{u}} \tag{4.17}$$

where

$$p_{\mathrm{u}} = 2\pi\lambda \int_0^\infty \exp(-\sigma^2\theta r^\alpha) e^{-\pi\lambda r^2 (1+\xi\rho(\alpha,\theta))} r \mathrm{d}r \tag{4.18}$$

and

$$p_{\mathrm{l}} = 2\pi\lambda \int_0^\infty \exp(-\sigma^2\theta r^\alpha) e^{-\pi\lambda r^2 (1+\rho(\alpha,\theta))} r \mathrm{d}r \tag{4.19}$$

Proof. A favorable system is considered for the upper bound [29], if the transmission of a packet fails, this packet is dropped instead of being re-transmitted. The interfering transmitter is then active with probability ξ, i.e., the packet arrival rate in the system, or we can say that the fraction of the active interfering transmitters remains constant in time. Substituting q_t in (4.4) by its minimum value $q_t = \xi$, the upper bound is obtained. In the lower bound case, the highest interference situation is obtained when all BSs are always active [29]. This corresponds to $q_t = 1$, which also gives the lowest value of the function in (4.4). □

It is worth to mention that the lower bound (4.19) is the coverage probability given in [4], and the upper bound (4.18) is the coverage probability given in [4] with a BS density thinned by a factor ξ. With Theorem 5, the stability condition in Lemma 4.2 is ensured at a cost of infinite buffer lengths if $p_{\mathrm{l}} \leq \xi$. Moreover, the bounds in Lemma 4.2 reduce to a simpler form, when the network is considered as interference-limited.

Corollary 4.2. *In an interference-limited network, i.e.* $\sigma^2 \to 0$,

$$p_{\mathrm{u}} = \left[1 + \xi \rho(\alpha, \theta)\right]^{-1} \tag{4.20}$$

$$p_{\mathrm{l}} = \left[1 + \rho(\alpha, \theta)\right]^{-1} \tag{4.21}$$

4.4.3 Queue delay analysis

The queue delay measures the delay between the time when a packet arrives at the queue and the time when it starts to be served, i.e., when it is transmitted.

Definition 4.2 (Queue delay [35])**.** *Considering a first in first out (FIFO) system, let W be the queue delay, i.e., the number of time slots for a randomly chosen packet spent in the queue before being transmitted. The average queue delay is given as* $\mathbb{E}[W] = \sum_{w=1}^{\infty} w\mathbb{P}(W = w)$.

Lemma 4.3. *Considering the FIFO system, with the service rate p and the arrival rate ξ, the average queue delay has the following expression*

$$\mathbb{E}[W] = \frac{\xi(1-\xi)}{p(p-\xi)}, \ \forall p > \xi \tag{4.22}$$

Proof. The result can be obtained from the FIFO system analysis in [35], and we detail the proof in the following. Let W be the queue delay, then

$$\mathbb{P}(W = 0) = \frac{p-\xi}{p}, \forall \ p > \xi \tag{4.23}$$

$$\mathbb{P}(W = w) = \sum_{i=1}^{w} x_i \binom{i-1}{w-1} p^i (1-p)^{w-i}, \ \forall w \geq 1, \ p > \xi \tag{4.24}$$

where x_i is given in (4.15), which is the probability of having i packets in the queue when the DTMC in Fig, 4.1 converges. The arguments of (4.24) are as follows: if we have i packets in the queue, an arriving packet will wait w time slots if in the first $i-1$ time slots exactly $i-1$ packets are completed and the service completion of the ith packet occurs at time slot w. Thus the summation over i from 1 to w. Considering the service rate as in (4.10), $\forall w \geq 1, p > \xi$, we have

$$\mathbb{P}(W = w) = \sum_{i=1}^{w} R^i \frac{x_0}{\bar{p}} \binom{i-1}{w-1} p^i (1-p)^{w-i} \tag{4.25}$$

$$= \frac{p-\xi}{p} (1-p)^{w-1} \frac{\xi}{1-\xi} \left(\frac{1}{1-\xi}\right)^{w-1} \tag{4.26}$$

$$= \frac{\xi(p-\xi)}{p} (1-\xi)^{-w} (1-p)^{w-1} \tag{4.27}$$

The mean queue delay is then

$$\mathbb{E}(W) = \sum_{w=0}^{\infty} w\mathbb{P}(W = w) = \frac{\xi(1-\xi)}{p(p-\xi)} \tag{4.28}$$

that ends the proof. □

4.4.4 Numerical results

In this section, we evaluate the performance of dynamic downlink cellular networks under different traffic intensity. BS positions are generated using a PPP with density $\lambda = 0.25$. Each UE is associated to its nearest BS. The packets arrive to each BS according to the Bernoulli process with parameter ξ and the service is then geometric with the parameter p_t at each time slot. For each network realization, the queues are let to evolve up to the convergence, i.e., the number of active transmitters does not evolve with time, then a new network realization is drawn and the process repeats.

Fig. 4.2 plots the coverage probability expressed in Theorem 4.1 w.r.t. the threshold θ. Moreover, two initialization states are considered: the full load case, i.e. the lower-bound in Lemma 4.2, and the light traffic initialization case, i.e. the upper-bound in Lemma 4.2. The time evolution of the coverage probability when the number of time slots increases is illustrated thanks to the arrows in Fig. 4.2. Whatever the initialization state is, the coverage probability converges to the stable coverage probability, corresponding to the stable distribution of the DTMC when $p \geq \xi$.

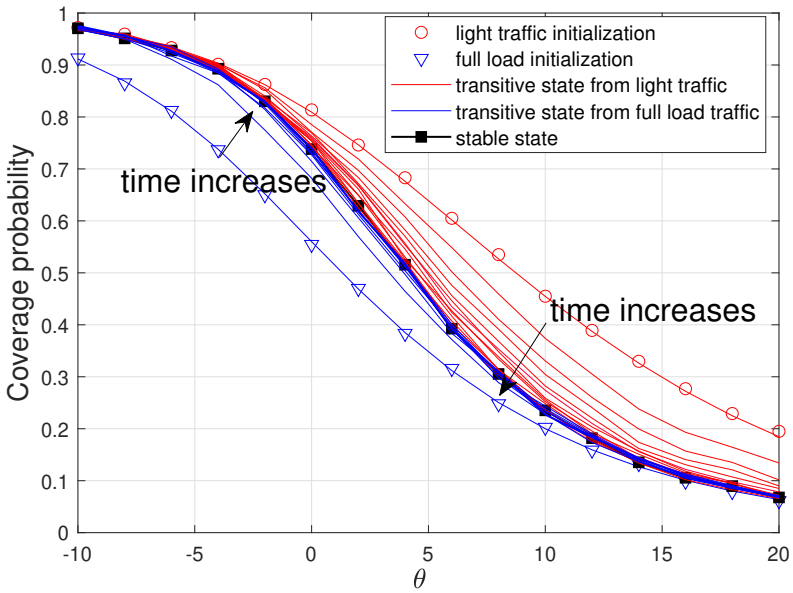


Figure 4.2 – Comparison of dynamic coverage probability with two initialization, $\xi = 0.3$, $\sigma^2 = 0$, $\lambda = 0.25$ and $\alpha = 4$.

Fig. 4.3 compares the analytical results in Theorem 5 with the Monte Carlo simulations under two network densities, $\lambda = 0.05$ and $\lambda = 0.2$. The average arrival rate is set to $\xi = 0.3$ and $\sigma^2 = 0.1$. The results corroborate the good match between simulations and analytical expressions. Moreover, we observe that the region between upper and lower bounds reduces when λ decreases. Indeed, as the density becomes lower, the interference level at the typical user decreases also and hence the upper bound is close to the lower bound.

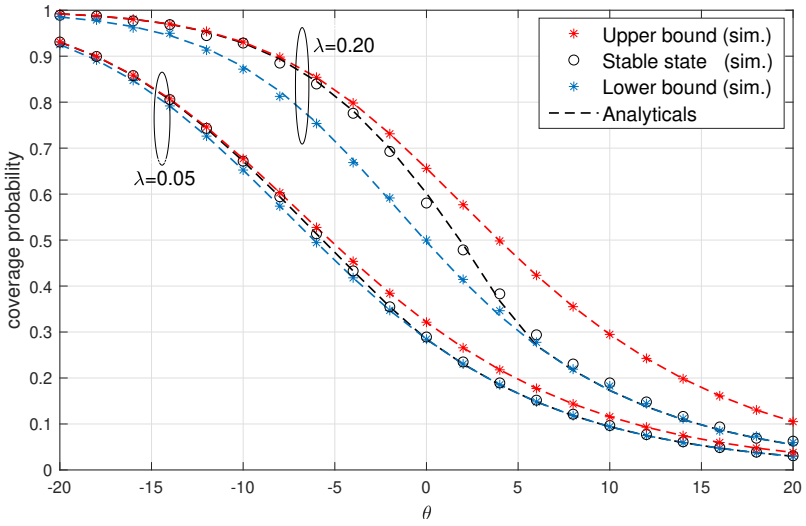


Figure 4.3 – Comparison of Monte Carlo simulation and analytically iterative algorithm of coverage probability at stable state.

Fig. 4.4 show the average queue delay for different packet arrival rates. As shown in the figure, when the network is stable, the average queue delay remains only a few time slots. Once the network is unstable, the average queue delay grows to infinity. In addition, to ensure the stability of the queue, i.e., the average queue length does not grow to infinity, the critical SINR threshold should decrease as the traffic intensity increases. For example, when $\xi = 0.8$, the network remains stable at $\theta < -10$ dB, while when $\xi = 0.3$, $\theta < 4$ is required to ensure a stable queue.

4.5 Coverage analysis with finite buffer

In this section, we proposed a tractable mathematical model to analyze the coverage probability and packet loss probability in the downlink cellular networks, considering a buffer restriction. Particularly, we derive the closed-form expression of the coverage probability that depends on the activity probability of a randomly chosen BS which is related to the buffer length. We also characterize the packet loss probability of a randomly chosen BS when the network becomes stable, i.e., the DTMC works at a stationary regime.

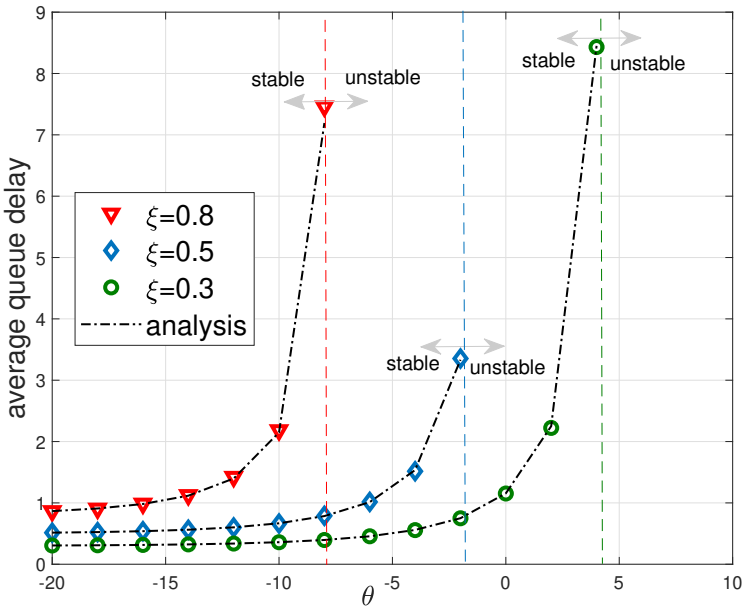


Figure 4.4 – Average queue delay $\mathbb{E}[W]$ versus θ, $\alpha = 4$, $\lambda = 0.25$, $\xi \in \{0.8, 0.5, 0.3\}$ packet/slot.

Theorem 6. *The stable coverage probability with a finite buffer restriction B is given by the fixed-point equation*

$$p(\theta,\xi,B) = 2\pi\lambda \int_0^\infty e^{-\sigma^2\theta r^\alpha} e^{-\pi\lambda r^2\left(1+\frac{(R^{B+2}-R)\rho(\alpha,\theta)}{R^{B+2}-R+(R-1)\bar{p}(\theta,\xi,B)}\right)} r\,dr$$

where $\bar{p}(\theta,\xi,B) = 1 - p(\theta,\xi,B)$, and $R = \frac{\xi\bar{p}(\theta,\xi,B)}{\bar{\xi}p(\theta,\xi,B)}$.

Proof. We consider the scenario where the arrived packets at each BS are stored in a buffer with finite size, until their successful transmission. When a packet arrives and the buffer is full, this new arrival packet is dropped. The buffer length restriction B are the same for all BS and it is the maximal number of packets the buffer can contained. The number of packets in the queues of a randomly chosen BS is modeled as a birth and death process that can be represented with the DTMC in Fig. 4.5 with $B+2$ states.

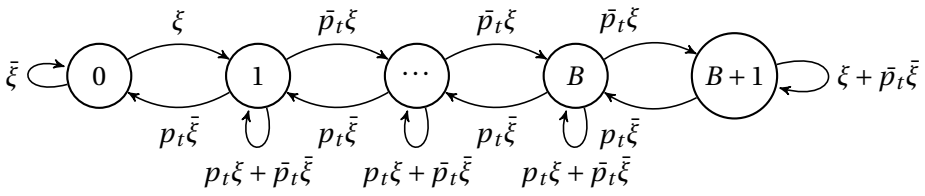


Figure 4.5 – DTMC model of the typical BS with a finite buffer length.

In Fig. 4.5, the state 0 represents the empty buffer event. When the buffer is in this state, the transmitter remains silent. When the queue is in the state B, it means that the buffer is full and hence any new arriving packet is dropped, i.e., the state $B+1$ is reached. The number of packets in the queue can be characterized by the stationary distribution of the previous DTMC. The transition probability matrix of size $(B+2)\times(B+2)$ is given by (4.12) in the finite case, i.e.,

$$\mathbf{P}=\begin{bmatrix} \bar{\xi} & \xi & 0 & 0 & \cdots & 0 \\ p\bar{\xi} & \bar{p}\bar{\xi}+p\xi & \bar{p}\xi & 0 & \cdots & 0 \\ 0 & p\bar{\xi} & \bar{p}\bar{\xi}+p\xi & \bar{p}\xi & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & p\bar{\xi} & \bar{p}\bar{\xi}+p\xi & \bar{p}\xi \\ 0 & 0 & 0 & 0 & p\bar{\xi} & \bar{p}\bar{\xi}+\xi \end{bmatrix} \tag{4.29}$$

With $\boldsymbol{x}=[x_0,x_1,x_2,\cdots,x_B,x_{B+1}]$, the solution of (4.13) is

$$\begin{cases} x_0 = x_0\bar{\xi}+x_1\bar{\xi}p \\ x_1 = x_0\xi+x_1(\bar{\xi}\bar{p}+p\xi)+x_2\bar{\xi}p \\ x_i = x_{i-1}\xi\bar{p}+x_i(\bar{\xi}\bar{p}+p\xi)+x_{i+1}\bar{\xi}p,\ 2\le i\le B \\ x_{B+1} = x_B\xi\bar{p}+x_{B+1}(\bar{p}+p\xi) \end{cases} \tag{4.30}$$

Then

$$x_i=\frac{x_0}{\bar{p}}\left(\frac{\xi\bar{p}}{\bar{\xi}p}\right)^i,\ 1\le i\le B \tag{4.31}$$

$$x_{B+1}=\left(\frac{\xi\bar{p}}{\bar{\xi}p}\right)^B\frac{\xi}{\bar{\xi}p}x_0 \tag{4.32}$$

After normalization, we obtain

$$x_0=\left[1+\xi R^B(\bar{\xi}p)^{-1}+(\bar{p})^{-1}\sum_{i=1}^{B}R^i\right]^{-1} \tag{4.33}$$

Combining (4.33) and (4.7), and the condition that $q=1-x_0$, we have

$$q=1-\left[1+\xi R^B(\bar{\xi}p)^{-1}+\sum_{i=1}^{B}R^i(\bar{p})^{-1}\right]^{-1} \tag{4.34}$$

$$p=2\pi\lambda\int_0^\infty e^{-\sigma^2\theta r^\alpha}e^{-\pi\lambda r^2(1+q\rho(\alpha,\theta))}r\,\mathrm{d}r \tag{4.35}$$

where $\rho(\alpha,\theta)=\int_1^\infty[1+u^{\frac{\alpha}{2}}\theta^{-1}]^{-1}du$, and $R=\frac{\xi\bar{p}}{\bar{\xi}p}$. According to (4.34) and (4.35), the interdependence between q and p shows the relationship between the queue and the stochastic geometry in the analysis. □

Corollary 4.3. *In an interference-limited network, i.e., $\sigma^2 \to 0$, we have*

$$p(\theta,\xi,B) = \left[1+\Upsilon\int_1^{\infty}\frac{1}{1+u^{\frac{\alpha}{2}}\theta^{-1}}du\right]^{-1} \tag{4.36}$$

where $\Upsilon = 1+(1-R^{B+2})^{-1}(1-R)^{-1}(1-p(\theta,\xi,B))$. *When the path loss exponent* $\alpha = 4$, *the stable coverage probability can be further simplified to*

$$p(\theta,\xi,B) = \left[1+\Upsilon\sqrt{\theta}(\frac{\pi}{2}-\arctan\sqrt{\theta})\right]^{-1} \tag{4.37}$$

For the sake of simplicity, we use p instead of $p(\theta,\xi,B)$ in the rest of this section. The fix point equation expressed in Theorem 6 can be iteratively solved using Algorithm 3.

Algorithm 3 Iterative algorithm to compute p of Theorem 6.

Initialize $q_1 \in (\xi,1)$, $q_0 = 0$, $i = 0$, $\epsilon \ll 1$
while $|q_{i+1}-q_i| \geq \epsilon$ **do**
  $i \leftarrow i+1$, $q \leftarrow q_i$, $p \leftarrow p$ in (4.35)
  if $|q_{i+1}-q_i| \geq \epsilon$ **then**
    $q_{i+1} \leftarrow q$ in (4.34)
    break
  end if
end while
Return $q \leftarrow q_{i+1}$ and p

4.5.1 Packet loss probability

The packet loss probability is the probability that a new packet is dropped when it meets the maximum queue length situation. This probability is given by the following lemma.

Lemma 4.4. *The packet loss probability at a randomly chosen BS with finite buffer length restriction B is given by*

$$p_{loss} = \frac{R^{B+2}-R^{B+1}}{(R-1)\bar{p}+R^{B+2}-R} \tag{4.38}$$

where p is the stable coverage probability given by Theorem 6 and $R = \frac{\xi\bar{p}}{\bar{\xi}p}$.

Proof. p_{loss} is the probability to be in the state $B+1$ and is obtained using (4.32) and (4.33). □

4.5.2 Numerical results

Fig. 4.6 compares the analytical result in Theorem 6 evaluated with Algorithm 3, with the Monte-Carlo simulations under different arrival rate ξ and buffer restriction B. The analytical results are shown with solid lines and simulations with marks. The results corroborate the

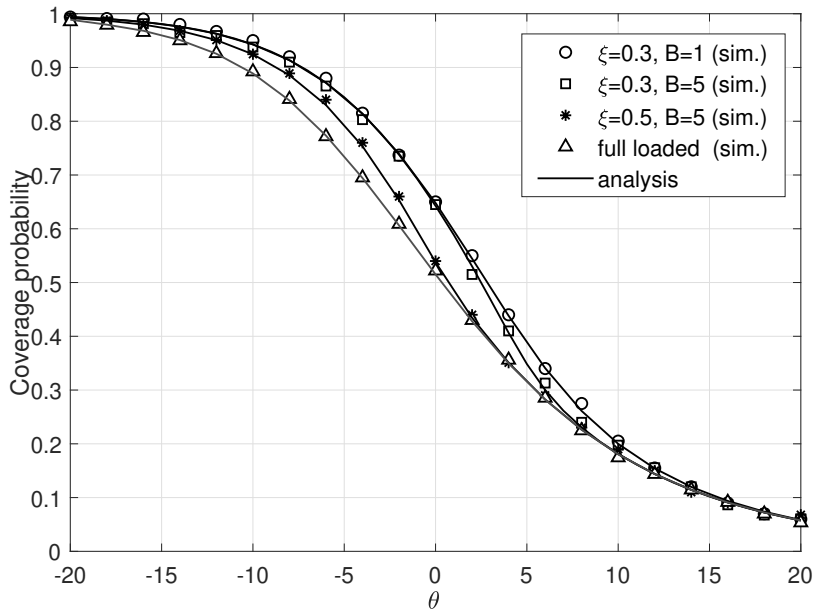


Figure 4.6 – Coverage probability at a stable state with different buffer restrictions B and arrival rates ξ.

good match between simulations and analytical expressions. We observe that the full load assumption is pessimistic w.r.t. the coverage probability. When the buffer length is kept constant as $B = 5$, we observe that the higher arrival rate leads to a lower stable coverage probability on a large range of coverage threshold θ. Indeed, when ξ increases, the queues are more solicited, then the BSs often have a packet to transmit, and hence they generate interference and the coverage probability at typical UE decreases.

A more surprising result is that the coverage probability is inversely related to the buffer size, i.e., increasing the buffer size B decreases the coverage probability. For example, fixing the arrival rate as $\xi = 0.3$, we observed that when B changes from 1 to 5, the coverage probability slightly degrades in the threshold range $[2, 14]$dB. This is because a BS with a large B drops fewer packets at a given arrival rate than a BS with a small buffer size. Therefore, when $B = 5$, the activity probability of a randomly chosen BS is more significant, which implies an increase in interference and a decrease in coverage probability. However, the constrained queue size implies a large packet loss since the randomly selected queue has a high probability of becoming full when B is low.

Fig. 4.7 plots the coverage probability and the packet loss probability in Lemma 4.4 for different arrival rates ξ and the queue lengths B. When fixing the queue length $B = 5$, we observe that the coverage probability improves when the arrival rate decreases, i.e., from $\xi = 0.3$ to $\xi = 0.1$. It comes from the fact that a lower arrival rate leads to fewer packets in the queue, which reduces the BSs' active duration and decreases the interference to the typical UE. Moreover, since fewer packets are generated, the packet loss probability is correspondingly

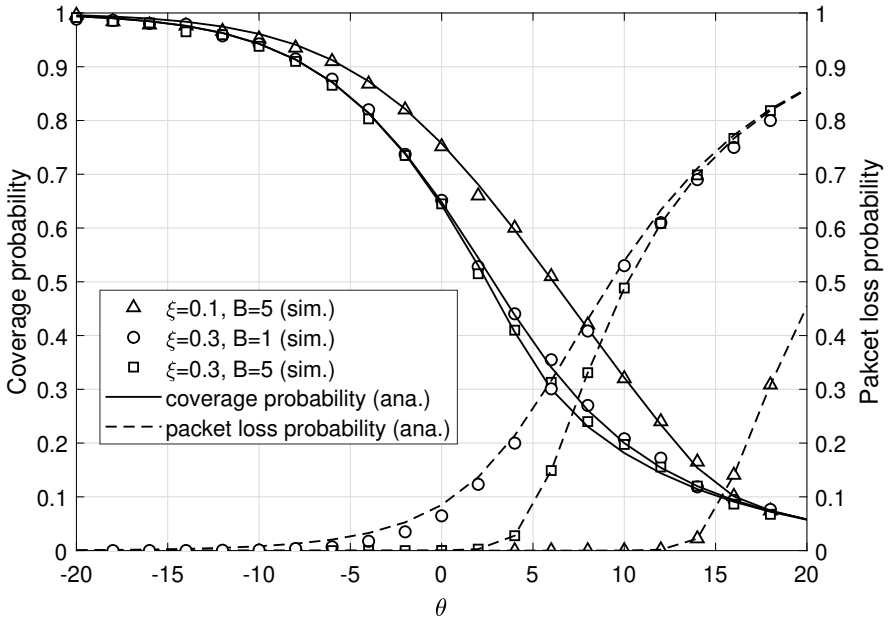


Figure 4.7 – Packet loss probability and coverage probability at stable state, $\lambda = 0.25$, $\sigma^2 = -10$ dB, $\alpha = 4$.

reduced for a given SINR threshold. When fixing the arrival rate at $\xi = 0.3$, we observe that the packet loss decreases significantly when B increases for a low to a medium range of values θ.

4.6 Conclusion

This chapter proposed tractable mathematical models to analyze the coverage probability in a dynamic traffic randomly deployed downlink cellular network. The queue evolution at each transmitter has been handled with a DTMC and a Bernoulli distribution for the packet arrival. The interaction between the coverage probability and the queue state evolution has been captured in closed-form. The explicit bounds on the dynamic coverage probability, the queue delay performances and the packet loss probability have been analysed.

Chapter 5

Analysis of ϵ-stable Region

5.1 Introduction

In this chapter, we use the system model of Chapter 4 with infinite buffer size assumption. Unlike the analysis in Section 4.4, where the SINR analysis provided limited information about the performance seen by a specific user, in this chapter, on the other hand, we characterize the ϵ-stable region in a large-scale dynamic downlink cellular network, with multi-cells and random link distances. We provide closed-form expression of the upper and lower bounds of ϵ-stable region by considering the modified systems and Markov inequality. Moreover, we propose an alternative definition of the ϵ-stable region and derive accordingly a tight approximation of the critical arrival rate that is unavailable in literature. In particular, the DTMC model in Section 4.4 is used to handle the interaction between the transmit success probability and the queue state evolution, to obtain the tight approximation of the critical arrival rate, contrary to the bounds provided in literature where the interaction between queues is not considered. Our result reveals that the proposed approximation is tighter than the bounds [1].

5.2 Transmit success probability

In this chapter, we ignore the noise at the UE and consider an interference-limited system for the sake of simplicity (the noise-limited system has been discussed in Example 3.4). Therefore, the SINR in (4.1) is replaced by the SIR. To obtain more fine-grained information, we start by defining the transmit success probability of the typical user at the time t, which is the conditional SIR distribution given the BS point process at the time slot t.

By applying the Slivnyak's theorem, it is sufficient to focus on the SIR of a typical UE at the origin rather than considering each UE in the PPP. The transmit success probability at the

[1] These works led to publications [C3], see § 1.3.

typical UE is

$$\mu_t \triangleq \mathbb{P}^{x_0}\left(\mathrm{SIR}_t > \theta | \Phi\right) = \mathbb{P}\left\{ \frac{H_{x_0,t} \|x_0\|^{-\alpha}}{\sum\limits_{x \in \Phi \setminus x_0} H_{x,t} \|x\|^{-\alpha} \mathbb{1}(x \in \Phi_t)} > \theta \Big| \Phi \right\} \tag{5.1}$$

Lemma 5.1. *The transmit success probability experienced by the typical UE at time t is*

$$\mu_t = \prod_{x \in \Phi \setminus x_0} \left(\frac{q_t}{1 + \theta \|x_0\|^{\alpha} \|x\|^{-\alpha}} + 1 - q_t \right) \tag{5.2}$$

Proof. The steps of this proof are similar to the steps in Appendix A.1. □

In Section 4.4, we analyse the first moment of μ_t, i.e., $\mathbb{E}[\mu_t]$. It can be seen as the average of the coverage probabilities of all the UEs in the network due to the ergodic setting. These metrics provide limited information. For example, given the packet arrival rate as $\xi = 0.4$, if the stable coverage probability $p = 0.5$ (that is, the SINR at typical UE exceeding the threshold θ when the time goes to infinity, i.e., p=$\lim\limits_{t\to\infty} \mathbb{E}[\mu_t]$), the network is said to be stable since $p > \xi$. However, it could be that half of the users which has a stable coverage probability of 0.8 and another half of users that have a stable coverage probability of 0.2. The other extreme case is that all the users have a stable coverage probability of 0.5. Clearly, the user experience will be quite different in the two cases, but the stable coverage probability p does not capture the difference.

Before we analyze the distribution of μ_t and further ϵ-stable region, we address two issues: *i*) due to the random packet arrival and retransmission of failed deliveries, the active state $\mathbb{1}(x \in \Phi_t)$ at each transmitter varies over time, and *ii*) there may exist common interfering BSs seen by the same UE from one time slot to another, which introduce temporal correlation inside the queues. It is non-trivial to obtain the closed-form of ϵ-stable region defined in (3.27) since the transmit success probability (5.1) is time-dependent.

5.3 Bounds of ϵ-stable region

5.3.1 Lower bounds

To derive the lower bound, we consider the full load system where all the BSs keep transmitting all the time. The full load system leads to the highest interference level, and the lowest transmit success probability in (5.1), i.e., $q_t = 1,\ \forall t \in \mathbb{N}$. By deriving the ϵ-stable region for the full load system, we get the lower bounds for the original system to be ϵ-stable. Noted that ξ_c^l is defined as $\xi_c^l \le \xi_c$.

Lemma 5.2. *The bth moment of the transmit success probability in full load system is given by*

$$M_b = \frac{1}{{}_2F_1(b, -\frac{2}{\alpha}; 1 - \frac{2}{\alpha}, -\theta)}. \tag{5.3}$$

Proof. Given the BS process Φ, the transmit success probability is

$$\mu_l = \mathbb{E}_{\{H_{x_0}\},\{H_x\}}\left[\mathbb{P}\left(\frac{H_{x_0}\|x_0\|^{-\alpha}}{\sum_{x\in\Phi\setminus x_0} H_x\|x\|^{-\alpha}} \geq \theta \middle| \Phi\right)\right] \tag{5.4}$$

$$= \prod_{x\in\Phi\setminus x_0}\left(\frac{1}{1+\theta\|x_0\|^{\alpha}\|x\|^{-\alpha}}\right) \tag{5.5}$$

The bth moment follows

$$M_b = \mathbb{E}\left[\prod_{x\in\Phi\setminus x_0}\left(\frac{1}{1+\theta\|x_0\|^{\alpha}\|x\|^{-\alpha}}\right)^b\right] \tag{5.6}$$

Instead of calculating this expectation in two steps as usual (first condition on $\|x_0\|$ then take the expectation with respect to it), we use the PGFL of the RDP from (2.12). Since (5.5) depends on the BS locations only through the relative distances, we can directly apply the PGFL of the RDP and can obtain (5.3).

$$M_b = \left[1+\int_1^{\infty}\left[1-\left(\frac{1}{1+\theta v^{-\frac{\alpha}{2}}}\right)^b\right]dv\right]^{-1} \tag{5.7}$$

which can be expressed as (5.3), the further detail see in Appendix A.3. □

Using the Gil-Pelaez inversion theorem, we obtain an exact integral expression for the critical arrival rate ξ_c^l for the full load case.

Theorem 5.1. *Considering the depicted downlink cellular network, the critical arrival rate ξ_c can be bounded as follows*

$$\xi_c \geq \xi_c^l = \sup\left\{\xi\in[0,1]: \frac{1}{2}-\frac{1}{\pi}\times\int_0^{\infty}\frac{1}{w}\mathrm{Im}\left\{\frac{\xi^{-iw}}{1+\int_1^{\infty}\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-b}\right]dv}\right\}dw \leq \epsilon\right\} \tag{5.8}$$

Proof. Seen in Appendix A.4. □

Remark 5.1. *When the path loss exponent $\alpha = 4$, the lower bound of the critical arrival rate can be simplified as*

$$\xi_c^l = \sup\left\{\xi\in[0,1]: \frac{1}{2}-\int_0^{\infty}\frac{1}{\pi w}\mathrm{Im}\left\{\xi^{-iw}{}_2F_1(b,-\frac{1}{2};\frac{1}{2},-\theta)\right\}dw \leq \epsilon\right\} \tag{5.9}$$

The following lemma gives a lower bound of ϵ-stable region by using Markov inequality, as in [5]. This bound is weaker compared with Theorem 5.1 but easier to evaluate.

Lemma 5.3. *Considering the depicted downlink cellular network, the critical arrival rate ξ_c can be bounded as follows*

$$\xi_c \geq \tilde{\xi}_c^l = \max_{n\in\mathbb{N}^+}\left[(1-\epsilon)\left[1+\int_1^\infty\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-n}\right]dv\right]\right]^{-\frac{1}{n}} \tag{5.10}$$

When the path loss exponent $\alpha = 4$, $\tilde{\xi}_c^l$ can be further simplified to

$$\tilde{\xi}_c^l = \max_{n\in\mathbb{N}^+}\left[(1-\epsilon)\left[{}_2F_1(-\frac{1}{2},-n;\frac{1}{2};-\theta)\right]\right]^{-\frac{1}{n}} \tag{5.11}$$

Proof. See Appendix A.5. □

The proof indicates that $\tilde{\xi}_c^l$ is a weaker bound than the one given by Theorem 5.1, because $\xi_c \geq \xi_c^l > \tilde{\xi}_c^l$.

Remark 5.2. *When $\epsilon \to 0$, the critical arrival rate approaches to 0; when $\epsilon \to 1$, the critical arrival rate approaches to 1:*

$$\lim_{\epsilon\to 0}\xi_c^l = 0, \quad \lim_{\epsilon\to 1}\xi_c^l = 1, \quad \forall\theta \geq 0 \tag{5.12}$$

5.3.2 Upper bounds

In order to derive upper bounds for the ϵ-stable region, we consider a favorable system, as described in Lemma 4.2. The upper bound strategy is: If the transmission of a packet fails, this packet is dropped instead of being re-transmitted. The interfering BS just serves the packet at each time slot and then it is active with probability ξ. Thus, the interference at typical UE is always lower than the original system at each time slot. By deriving the ϵ-stable region for the favorable system, we get the upper bounds ξ_c^u for the original system to be ϵ-stable, such as $\xi_c \leq \xi_c^u$.

Theorem 5.2. *Considering the depicted downlink cellular network, the critical arrival rate ξ_c can be bounded as follows*

$$\xi_c \leq \xi_c^u = \sup\left\{\xi\in[0,1]: \frac{1}{2}-\frac{1}{\pi}\times\int_0^\infty\frac{1}{w}\mathrm{Im}\left\{\frac{\xi^{-iw}}{1+\int_1^\infty\left[1-\left(1-\frac{\xi\theta}{\theta+v^{\alpha/2}}\right)^{iw}\right]dv}\right\}dw \leq \epsilon\right\} \tag{5.13}$$

Proof. In this favorable system, we have $\mathbb{1}(x\in\Phi_t) = \xi,\ \forall t\in\mathbb{N},\ \forall x\in\mathbb{R}^2$. This corresponds to $q_t = \xi$, which leads to the lowest interference and the lowest transmit success probability in (5.1). The detailed proof is in Appendix A.6. □

Corollary 5.1. *Given a slotted system with transmitters distributed as a PPP and per-link with random distance, for all $n>0$, a upper bound of ξ_c is $\xi_c<\xi_c^u<\tilde{\xi}_c^u$, where $\tilde{\xi}_c^u$ is the solution of the fixed-point equation*

$$\tilde{\xi}_c^u=\epsilon^{\frac{1}{n}}\left[1+\int_1^{\infty}\left[1-\left(1-\frac{\tilde{\xi}_c^u\theta}{\theta+v^{\frac{\alpha}{2}}}\right)^{-n}\right]dv\right]^{\frac{1}{n}},\ \forall n\in\mathbb{N}^+ \tag{5.14}$$

Proof. Seen in Appendix A.7. □

Remark 5.3. *When the SIR receiving threshold $\theta\to 0$, for all $\epsilon\geq 0$, the critical arrival rate approaches to 1. This indicates that the lower bound and upper bound are tight for small θ.*

Proof. In the original system, where the queues interact with each other, when $\theta\to 0$, a transmission is almost surely successful if it is scheduled. Therefore, the serving processes of the packets at different transmitters can be approximated as independent, and the critical arrival rates below 1 are intuitively reasonable. On the other side, a transmission is almost surely failed when $\theta\to\infty$. Therefore, the critical arrival rate of a ϵ-stable network tends to 0. A detailed proof is approved in Appendix A.8. □

5.4 Approximation

In the previous section, we derived various bounds of the $\epsilon-$stable region. However, these bounds are tight in some network configurations, e.g., at small SIR threshold θ value. In contrast, they may be loose in other network configurations, as we will show later in the numerical results. Since describing the distribution of ϵ-stable region is not easy, as we mentioned in Section 5.2, we consider a modified ϵ-stable region definition in this section. This definition is motivated by Fig. 4.2, where two different initialization states were considered to get the stable coverage probability. Under this definition, we ignore the transient values of the initial period of μ_t and describe the ϵ-stable region only when time goes to infinity. Compared to (3.27), the new region is simpler to handle and yields some tractable expressions.

Definition 5.1. *A modified definition of the ϵ-stable region, instead of (3.27), is given by*

$$\mathscr{S}_\epsilon=\left\{\xi\in[0,1]:\mathbb{P}\left\{\lim_{t\to\infty}\mu_t\leq\xi\right\}\leq\epsilon\right\} \tag{5.15}$$

We define μ as $\mu=\lim_{t\to\infty}\mu_t$. Following the same argument as the one in Lemma 4.7, μ has the following expression

$$\mu=\prod_{x\in\Phi\setminus x_0}\left(\frac{q}{1+\theta\|x_0\|^{\alpha}\|x\|^{-\alpha}}+1-q\right) \tag{5.16}$$

As we discussed in § 4.4, a randomly chosen BS has a probability ξ/μ to be active if its arrival rate is less than the departure rate, and it is always active in the opposite case. That is

$$q=\begin{cases}\xi/\mu, & \text{if } \mu>\xi,\\ 1, & \text{if } \mu\leq\xi.\end{cases} \tag{5.17}$$

It is important to note when $\mu < \xi$, then $q = 1$ and all the queue lengths and average queue delays grow up to infinity, corresponding to an unstable network. In the following, we present the approximation of ϵ-stable region in downlink cellular networks.

Theorem 5.3. *Considering the dynamic downlink cellular network introduced above and the definition in* (5.15)*, the approximated critical arrival rate of ϵ-stable region can be characterized as follows*

$$\tilde{\xi}_c = \sup\left\{\xi \in [0,1] : \frac{1}{2} - \frac{1}{\pi} \times \int_0^\infty \frac{1}{w} \mathrm{Im}\left\{\frac{\xi^{-iw}}{1 + \int_1^\infty \left[1 - \left(1 - \frac{\mathbb{E}[q]\theta}{\theta + v^{\alpha/2}}\right)^{iw}\right] dv}\right\} dw \le \epsilon\right\} \tag{5.18}$$

where

$$\mathbb{E}[q] = \begin{cases} \frac{\xi}{1-\theta\xi\rho(\theta,\alpha)}, & \text{if } \frac{1}{1+\theta\rho(\theta,\alpha)} > \xi \\ 1, & \text{if } \frac{1}{1+\theta\rho(\theta,\alpha)} \le \xi \end{cases}$$

and $\rho(\alpha,\theta) = \int_1^\infty [\theta + u^{\frac{\alpha}{2}}]^{-1} du$.

Proof. See Appendix A.9. □

The expression in (5.18) quantifies how the key features of a dynamic network, i.e., interference, SIR threshold and packet arrival rate, affect the distribution of the ϵ-stable region. Several remarks regarding Theorem 5.3 are in order.

Remark 5.4. *The upper and lower bound of the critical arrival rate in Theorem 5.1 and Theorem 5.2 corresponds to $\mathbb{E}[q] = \xi$ and $\mathbb{E}[q] = 1$ in Theorem 5.3, respectively.*

Remark 5.5. *When the SIR threshold $\theta \to 0$, for all $\epsilon \ge 0$, the critical arrival rate approaches to 1. Letting $\theta \to 0$, Theorem 5.3 becomes*

$$\begin{aligned} \lim_{\theta\to 0} \tilde{\xi}_c &= \sup\left\{\xi \in [0,1] : \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{\xi^{-iw}\right\} dw \le \epsilon\right\} \\ &= \sup\left\{\xi \in [0,1] : \frac{1}{2} + \frac{1}{\pi} \times \frac{\pi}{2}\mathrm{sgn}(\ln\xi) \le \epsilon\right\} \\ &= 1 \end{aligned} \tag{5.19}$$

since $\mathrm{sgn}(\ln\xi) = -1, \forall\ \xi \in (0,1)$. *Similar conclusion can be drawn for the upper bound ξ_c^u and lower bound ξ_c^l.*

Remark 5.5 illuminates that a transmission attempt is almost surely successful when $\theta \to 0$, thus the admissible critical arrival rate approaches 1.

5.5 Numerical results

In this paragraph, we validate the accuracy of our analysis through simulations, and explore the impact of the traffic condition on the network performance, from several aspects. Unless otherwise mentioned, the following parameters are used throughout this paragraph: the path loss exponent is $\alpha = 4$, the BS density is $\lambda = 0.25$, and the packet arrival rate is $\xi \in [0,1]$ packet/slot.

Three simulation scenarios are considered:

(i) The original system described in section 4.2. For each network realization, the queues are let to evolve up to the convergence, i.e., when the number of active transmitters stabilizes and does not evolve with time. Then a new network realization is drawn again and the process repeats;

(ii) The full load case, where all BSs keep transmitting all the time, leading to the lower bound ξ_c^l described in Theorem 5.1;

(iii) The favorable system, where a randomly chosen BS is active with probability ξ, leading to the upper bound ξ_c^u described in Theorem 5.2.

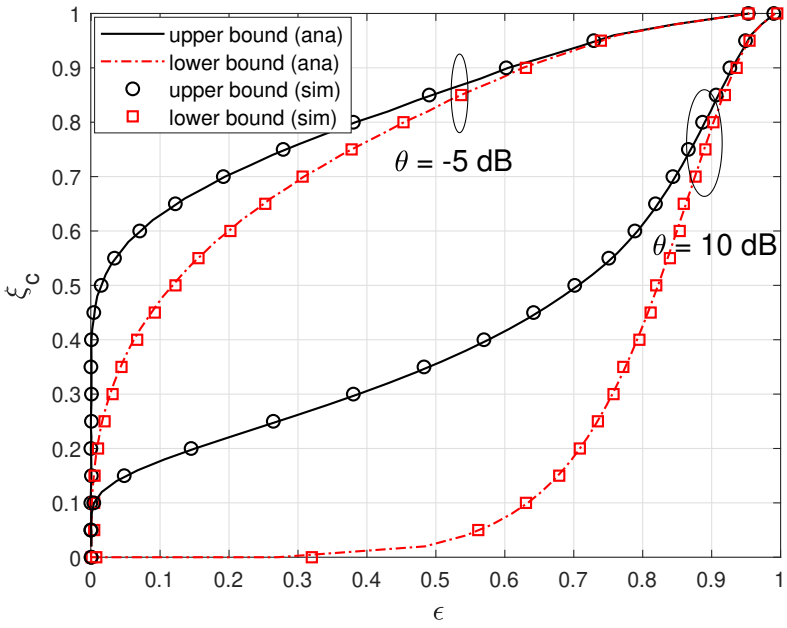


Figure 5.1 – Upper and lower bounds of the ϵ-stable region.

Fig. 5.1 compares the analytical results with the Monte Carlo simulations under two different SIR thresholds, $\theta = 0$ dB and $\theta = 5$ dB. The network is ϵ-stable if and only if the average arrival rate $\xi \leq \xi_c$ at each BS. The figure shows a good match between simulations and

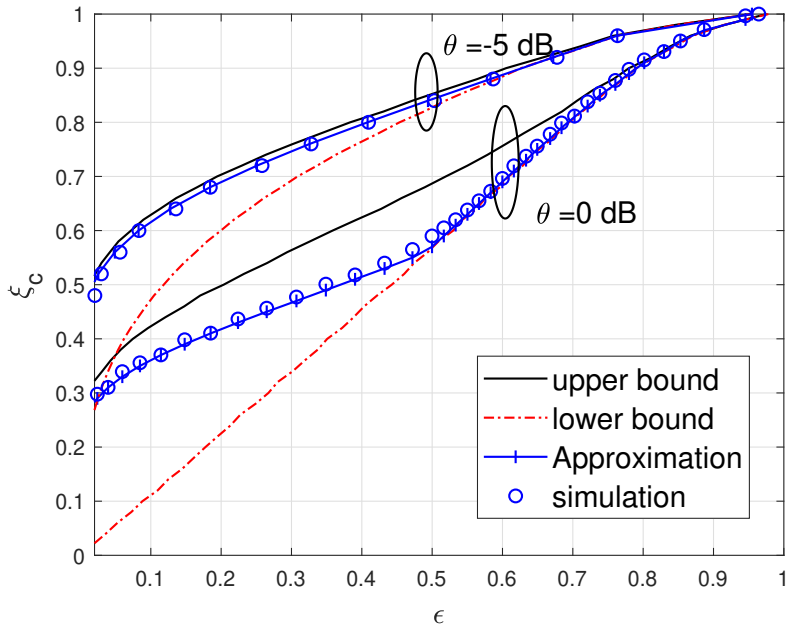


Figure 5.2 – The approximation and bounds of the ϵ-stable region, $\theta \in \{-5, 0\}$ dB.

analytical expressions. Moreover, we observe that the region between upper and lower bounds reduces when ϵ increases, the bounds converge to 1 when $\epsilon = 1$, as expected. Moreover, the lower and upper bounds decrease when θ increases. Indeed, as θ increases, a transmission has a higher chance to fail when scheduled. Hence, the possible arrival rates, i.e., those for which the network is stable, decrease.

Fig. 5.2 focuses on the upper and lower bounds, i.e., ξ_c^u and ξ_c^l respectively, as well as the critical arrival rate $\tilde{\xi}_c$ derived in Theorem 5.3. The critical arrival rate obtained by simulation is based on Definition 3.27 and it is compared to the expression in Theorem 5.3 which is based on (5.15). We can observe that the critical rate lies between our upper and lower bounds, and Theorem 5.3 reveals to be a good approximation of the true critical rate, as confirmed by the simulations. This observation implies that the transient phase present in Definition 3.27, but not in (5.15), has a negligible effect on the critical arrival rate. Moreover, it is observed that the critical arrival rate $\tilde{\xi}_c$ is close to the upper bound ξ_c^u when θ is relatively small, i.e., $\theta = -5$ dB. This is because decreasing θ will increase the opportunity of a successful transmission. Thus the active probability of the typical BS in the real case is much closer to the active probability in the favorable system.

5.6 Conclusion

This chapter has proposed a complete characterization of the ϵ-stable region in a dynamic downlink cellular network. We derived the closed-form expression of the critical arrival rate,

the rate at which the proportion of unstable queues are under a certain threshold. The upper and lower bounds of the ϵ-stable region have also been derived. These results allow to quickly evaluate the proportion of queues, in average, that are in outage when the network deployment is modeled with a PPP and when the network traffic is modeled with a DTMC. In the next chapter, we consider the problem of reinforcement learning-based transmission policies considering the channel state information, queueing state and packet arrival state in dynamic downlink cellular networks.

Chapter 6

RL based Transmission Policies in Dynamic Cellular Networks

6.1 Introduction

The characterization of the stable regions by considering the resource allocation is non-trivial to obtain, because of the dependence between the geometry and the dynamics of the networks and the allocation strategy. However, the dynamic nature of the network considered in this thesis leads itself perfectly to description by a Markovian decision process for which reinforcement learning strategies can be proposed.

In this chapter, we provide transmission policies considering the channel state information, the queue states, and the aggregate interference in dynamic downlink cellular networks. Large scale networks with multi-cells and random link distances are considered. We studied a constrained optimization problem to minimize the long-term transmission cost with delay constraints. The problem is formulated with an infinite horizon Markov decision process, and solving it online using reinforcement learning. First,We proposed algorithms based on Q-learning and SARSA to train the formulated RL model. Then, we analyze the stable region of both greedy policy and RL-based policy. The greedy policy provides an upper bound of stable probability compared with the RL-based policy. We show that there exists a trade-off between the stable probability and the transmitting costs which depends on the traffic intensity. The numerical results reveal that the RL-based policies hold the same stable region compared to the greedy policy however with a lower transmission cost [1].

6.2 RL problem formulation

In this chapter, we come back to the system model introduced in Section 4.2 with infinite buffer assumption. We consider a more flexible transmission policy based on reinforcement learning for the BS located in 0-cell, while the other BSs works with greedy policy, i.e., keep transmitting when buffer is not empty.

[1] Part of results is submitted to publication [C4], see § 1.3.

As discussed in Section 2.3, RL is based on the interaction of an agent with an unknown environment. The decision is made through a process of trial and error. In each state $s \in \mathscr{S}$, the agent selects an action from the set of possible actions $\mathscr{A}$. The choice of an action is dictated by the policy defined by the distribution $\pi(a|s)$, which is updated using a learning process. The interactions between the agent and the environment continue until the agent has learned the policy that maximizes its cumulative reward over the long term [1].

6.2.1 Formulation as a constrained Markov decision process

In this section, we formulate the wireless transmission management problem as a constrained Markov decision process (MDP). We defined the state space, the action space as well as the policy.

State space The state $S(t) \in \mathscr{S}$ at the time slot t is defined by the vector $[S_c(t), S_b(t), S_y(t)]$, where

- $S_b \in \mathscr{B} = \{0,1\}$ indicates whether the buffer is empty or not;
- $S_y \in \mathscr{Y} = \{0,1\}$ is 1 if a new packet arrives in the buffer and 0 otherwise, where $\mathbb{P}(S_y = 1) = \xi,\ \forall t \in \mathbb{N}$;
- $S_c \in \mathscr{C}$ represents the state of the channel where $S_c = i$, if $\gamma_t \in [\theta_i, \theta_{i+1})$, $i \in [1, M]$. Noted that $\theta_0 = 0$ and $\theta_{M+1} = +\infty$.

Note that the state space $\mathscr{S} = \mathscr{B} \times \mathscr{Y} \times \mathscr{C}$.

Action space At each time slot, the BS in the 0-cell has to decided whether to transmit a packet or not, unless the buffer is empty, in which case it remains silent. The action space is $\mathscr{A} = \{0,1\}$, where 0 stands for not transmitting, and 1 for transmitting. If a packet is transmitted, i.e., $A(t) = 1$, the packet is removed from the buffer if it is well received, with a certain probability $f_s(1, S_c(t))$, which characterizes the quality of the communication:

$$f_s : \mathscr{A} \times \mathscr{C} \rightarrow [0,1] \tag{6.1}$$

where $f_s(\cdot,\cdot)$ is increasing in the SIR state, i.e., a better SIR level leads to a higher success probability, and $f_s(0, S_c(t)) = 0$,

$$f_s(1, S_c(t)) = \begin{cases} 0, & \text{if } \gamma_t < \theta_1 \\ f_m, & \text{if } \theta_m < \gamma_t < \theta_{m+1}, m \in [1, M-1] \\ f_M, & \text{if } \gamma_t > \theta_M \end{cases} \tag{6.2}$$

We are interested to transmit packets with the minimal transmission cost while limiting the waiting time in the buffer. In that case, the agent's objective function is modeled by two cost functions, i.e., the transmit cost and the delay cost functions:

1. The transmission cost is a non-increasing function with the SIR, i.e., it does not cost more to transmit in good conditions rather than in bad ones. Let $f(\cdot,\cdot):\mathscr{A}\times\mathscr{C}\to\mathbb{R}^+$ be a non-increasing function that depends on the action and channel states such that

$$C(A(t),S(t))=\begin{cases} f\left(A(t),S_c(t)\right), \text{ if } A(t)=1 \\ 0, \text{ if } A(t)=0 \end{cases} \tag{6.3}$$

2. The non-negative delay cost, which is applicable when the buffer is not empty, i.e., when $B(t)>0$, $W(\cdot,\cdot):\mathscr{A}\times\mathscr{S}\to\mathbb{R}^+$

$$W(A(t),S(t))=\begin{cases} 0, \text{ if } A(t)=1, \text{ or } A(t)=0 \text{ and } S_b(t)=0 \\ w, \text{ if } A(t)=0 \text{ and } S_b(t)>0 \end{cases} \tag{6.4}$$

Policy The transmission scheduling policy is a map between the system states and the actions. It may be deterministic or be defined in the sense of distributions. Based on the cost functions defined in (6.3) and (6.4), we define two *long-term* costs, knowing an initial state $S(0)=s$ and a realization of the network Φ as

$$C_\pi(s)=\mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\eta^t C(A(t),S(t))\Big|S(0)=s,\Phi\right] \tag{6.5}$$

$$W_\pi(s)=\mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\eta^t W(A(t),S(t))\Big|S(0)=s,\Phi\right] \tag{6.6}$$

where the expectation is taken over the distribution of the policy and the dynamic of the underlying MDP, and $\eta\in(0,1)$ is the discounting factor. As η approaches 1, the agent cares more about future rewards than the lower value of η. As η approaches 0, it means that the policy emphasized the short-term gain. We choose to use the discounted criterion because it ensure the existence of a stationary policy straightforwardly [110, 15].

The RL problem consists in finding the policy π which minimizes the average transmission cost under the delay cost constraint, i.e.,

$$\min_{\pi\in\Psi} C_\pi(s) \quad \text{s.t.} \quad W_\pi(s)\le\delta,\ \forall s\in\mathscr{S} \tag{6.7}$$

Formally, we denote the collection of probability distribution on subsets of $\mathscr{A}$, then the policy is a map function

$$u:\mathscr{B}\times\mathscr{Y}\times\mathscr{C}\to\mathscr{P}(\mathscr{A}) \tag{6.8}$$

For an infinite horizon MDP, the only case of interest is the existence of an optimal stationary policy. A general policy is defined by a single decision rule $u=u_t(\cdot,\cdot,\cdot)$, $\forall t=1,2,\cdots$, and a stationary policy is defined by $\pi=(u_t(\cdot,\cdot,\cdot)=u(\cdot,\cdot,\cdot))$, which means invariant in time. Besides, we denote the set of all policies by $\mathscr{U}$, and the set of stationary policies by Ψ.

6.2.2 The Lagrangian approach

Considering the structure of the constrained optimal transmission policy, we reformulate the constrained MDP as a parameterized unconstrained MDP using the Lagrange multiplier approach. For each Lagrangian multiplier λ, the instantaneous Lagrangian cost, $L : \mathbb{N} \times \mathscr{S} \rightarrow \mathscr{L} \subset \mathbb{R}$, at time t is defined as

$$L(t,\lambda) = C(A(t),S(t)) + \lambda W(A(t),S(t)) \tag{6.9}$$

From [111], solving the constrained MDP problem is equivalent to solve the unconstrained MDP and its Lagrangian dual problem. We present this in the following:

Theorem 6.1. *The optimal value of the constrained MDP can be formulated as*

$$L_{\pi^*}(s,\lambda^*) = \min_{\pi\in\Psi}\max_{\lambda\geq 0}(L_\pi(s,\lambda) - \lambda\delta) = \max_{\lambda\geq 0}\min_{\pi\in\Psi}(L_\pi(s,\lambda) - \lambda\delta) \tag{6.10}$$

where

$$L_\pi(s,\lambda) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\eta^t L(t,\lambda)\Big| S(0) = s,\Phi\right] = C_\pi(s) + \lambda W_\pi(s) \tag{6.11}$$

and a policy π^* is optimal for the constrained MDP if and only if

$$L_{\pi^*}(s,\lambda^*) = \max_{\lambda\geq 0}(L_{\pi^*}(s,\lambda) - \lambda\delta) \tag{6.12}$$

Proof. The detailed proof can be found in [111, Chapter 3]. Noted that the discounted cost is this book is defined with the normalization constant $(1-\eta)$. However the techniques are the same for both cases, and one could retrieve one from the other by multiplying or dividing the immediate cost by this factor, as $\eta \neq 1$. The main idea of the proof comes from three aspect: (i) the constrained optimal problem is equivalent to solving a non-constrained sup-inf problem; (ii) the inf and the sup can be interchanged under suitable conditions by invoking a saddle point theorem: the inf in the inf-sup problem is in fact achieved by some policy which is optimal for constrained optimal problem; (iii) under the Slater conditions, the sup-inf is also obtained as max-min, this comes from the fact that the objective function and the inequality constrained function are convex with respect to the stationary policy π and the set of stationary policies is a closed convex polytope. □

The dual problem, i.e., the max-min problem, is more familiar, since it involves first minimization with respect to the policies, and only then maximizing with respect to λ. For each fixed λ we are faced with a standard non-constrained problem of a controlled Markov chain, and we can therefore obtain the minimization through well-known dynamic programming techniques. For a fixed λ, the rightmost minimization in (6.10) is equivalent to solving the following dynamic programming equation:

$$L_{\pi^*}(s,\lambda) = \min_{a\in\mathscr{A}}\sum_{(s',l)\in\mathscr{S}\times\mathscr{L}} p(s',l|s,a)\left[L(0,\lambda) + L_{\pi^*}(s',\lambda)\right] \tag{6.13}$$

where $L_{\pi^*}(s,\lambda): \mathscr{S} \to \mathbb{R}$ is the optimal state-value function given λ and Φ.

Similarly, the state-action function $q_\pi(s,\lambda)$ is the expected long-term cost starting from the state s, taking the action a, and following policy π, which can be expressed as:

$$q_\pi(s,\lambda) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \eta^t L(t,\lambda) \middle| S(0)=s, A(0)=a, \Phi\right] \tag{6.14}$$

Lemma 6.1. *Given Φ and λ, the optimal policy π^*_λ can be obtained by*

$$\pi^*_\lambda = \arg\min_{a\in\mathscr{A}} q^*_\lambda(s,a), \forall s \in \mathscr{S} \tag{6.15}$$

where $q^*_\lambda(s,a)$ satisfying

$$q^*_\lambda(s,a) = \sum_{s'\in\mathscr{S}} \sum_{l\in\mathscr{L}} p(s',l|s,a)\left[l + \min_{a'} q^*_\lambda(s',a')\right] \tag{6.16}$$

Proof. The proof is based on the Bellman equation introduced in Section 2.3.2 and detailed in Appendix A.10. □

In practice, the transition probabilities, i.e., $p(s',l|s,a)$, are unknown a *priori*. Consequently, π^* and $q^*_\lambda(s,a)$ cannot be computed using value iteration, instead, they must be learned online based on experience. In the next, we consider two classical RL learning methods to obtain the optimal $q^*_\lambda(s,a)$ in (6.16).

6.3 Learning the optimal policy

6.3.1 Q-learning and SARSA

As we introduced in section 2.3.3, Q-learning and SARSA are two iterative algorithms which make it possible to converge towards the optimal stationary policy for unconstrained MDP, in the sense of the cost function defined above. To do this, the value of the state-action function is updated from an incremental difference between the objective and the previous estimate of the state-action function, that is:

$$\underbrace{q_{t+1}(S(t+1),A(t+1))}_{\text{new estimate}} \leftarrow \underbrace{q_t(S(t),A(t))}_{\text{old estimate}} + \alpha_t \left[\underbrace{T_{t+1}}_{\text{target}} - \underbrace{q_t(S(t),A(t))}_{\text{old estimate}}\right] \tag{6.17}$$

where α_t is the step-size which should satisfying $\sum_{t=0}^{\infty}(\alpha_t)^2 < \infty$ to ensure convergence [37]. In practice, it can be taken as a constant far less than one, e.g., $\alpha = 0.01$ in [14], $\alpha = 0.00025$ in [112]. On the other hand, T_{t+1} is the objective value of the algorithm and takes a slightly different form depending on whether we consider Q-learning or SARSA.

Q-learning In this case, the target value has the form:

$$T_{t+1} = L(t+1,\lambda) + \eta \min_{a' \in \mathscr{A}} q_t(S(t+1), a'), \tag{6.18}$$

To balance exploitation and exploration, the ϵ-greedy policy is used to select the action at each instant. That is, for $\epsilon \in [0,1]$

$$a^* = \begin{cases} \arg\min_{a \in \mathscr{A}} q(S(t), a), \text{ with probability } 1-\epsilon_t, \\ a \in \mathscr{A}, \text{ with probability } \epsilon_t. \end{cases} \tag{6.19}$$

where ϵ_t can be kept constant or may vary during the learning in order to explore more at the beginning, i.e., $\epsilon_t \approx 1$, and exploit more after a while, e.g., $\epsilon_t = \frac{1}{t+1}$ [56].

SARSA SARSA differs from the Q-learning by the target definition in (6.18). In SARSA, we use

$$T_{t+1} = L(t+1,\lambda) + \eta q_t(S(t+1), A(t+1)) \tag{6.20}$$

Unlike *Q-learning*, the *behaviour* and *target* policies, are both ϵ-greedy, i.e. the next action to take observing the state $S(t+1)$ is the action a' which maximizes $q_t(S(t+1), a')$ with probability $1-\epsilon$, and a random action with probability ϵ.

In Algorithm 4, we show the Q-learning and SARSA algorithm embedded in the agent to obtain $q^*_\lambda(s,a)$, where the M episodes correspond to the M realizations of the Φ process.

Algorithm 4 Q-learning (SARSA) algorithm to obtain $q^*_\lambda(s,a)$.

Initialization arrival rate ξ, BS density λ_b;
for episode = $1,2,\cdots,M$ **do**
  (1) Initialize λ, $q_\lambda(s,a) = 0, \forall\ s \in \mathscr{S}, a \in \mathscr{A}$
  (2) Initialize $S(0) = s$
  for $t = 1,2,\cdots$ **do**
    (a) Choosing actions a for s by ϵ-greedy (2.44)
    (b) Take action a, observe immediate Lagrangian cost l
    (c) Realization of channel distribution for all BSs, observe dynamic SIR
    (d) observe new state $s' = (s'_c, s'_y, s'_b)$
    (e) Choose a' for s': $\min_{a' \in \mathscr{A}} q_t(S(t+1), a')$ for Q-learning, or ϵ-greedy for SARSA;
    (f) update $q_\lambda(s,a)$ by (6.18) for Q-learning or (6.20) for SARSA;
  end for
end for

6.3.2 Optimal Lagrange multiplier

The optimal value of the Lagrange multiplier λ can be learned online using *stochastic sub-gradients*, as in [15, 104, 14]. Let λ^* be the optimal Lagrange multiplier, then it satisfies

$$\lambda^* = \arg\max_{\lambda \geq 0} \min_{\pi \in \Psi} (L_\pi(s,\lambda) - \lambda\delta) \tag{6.21}$$

Let π_λ be the optimal transmission policy corresponding to the Lagrange multiplier λ, and note that $W(\pi_\lambda) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \eta^t W(A(t), S(t)) \middle| S(0) = s, \Phi \right]$, λ^* can be estimated iteratively as

$$\lambda_{k+1} = \lambda_k + \beta_k (W(\pi_{\lambda_k}) - (1-\eta)\delta) \tag{6.22}$$

where $(1-\eta)\delta$ terms converts the discounted delay constraint δ to an average delay constraint, β_k is a time-varying learning rate. To ensure the sequence of Lagrange multipliers $(\lambda_1, \lambda_2, \cdots)$ convergence to λ^*, β_k should satisfy the following conditions [113]:

$$\beta_k \geq 0, \sum_{k=0}^{\infty} \beta_k = \infty, \text{ and } \sum_{k=0}^{\infty} (\beta_k)^2 < \infty \tag{6.23}$$

In our case, we choose $\beta_k = \frac{1}{k}$, which satisfied the convergence condition. At each update of λ, we train the stationary strategy with dynamic programming based on the current λ and obtain $W(\pi_\lambda)$ accordingly. From the dynamic programming point of view, λ is constant during minimization the *long-term* cost with respect to the policies. From the point of view of updating λ, $W(\pi_{\lambda_k})$ converges to the optimal value corresponding to the current value of λ. Thus, we can obtain a sequence of update policies and λ values. An algorithm is provided to optimize λ seen in Alg. 5. Note that λ^* depends on the geometry of the network, in particular, the position of the user w.r.t its BS.

Algorithm 5 Iteration algorithm for computation of λ^*.

Initialize $\lambda_1 > \epsilon$, $\lambda_0 = 0$, $k = 0$, $\epsilon \ll 1$
while $|\lambda_{k+1} - \lambda_k| \geq \epsilon$ **do**
  $k+1 \leftarrow k$
  $W(\pi_{\lambda_k}) = \mathbb{E}_{\pi_{\lambda_k}} \left[\sum_{t=0}^{\infty} \eta^t (W(A(t), S(t)) | S(0) = s, \Phi \right]$
  $\lambda_{k+1} \leftarrow \lambda_k + \frac{1}{k}(W(\pi_{\lambda_k}) - (1-\eta)\delta)$
end while
Return $\lambda^* \leftarrow \lambda_{k+1}$ and $W(\pi^*_\lambda)$

6.4 Stable probability analysis

In this section, we investigate the stable probability for greedy policy and RL-based policies, which is defined as

$$p_s = \mathbb{P}_\Phi \left[\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E} \left[f_s(S_c(t), A(t)) \middle| S(0) = s, \Phi \right] > \xi \right] \tag{6.24}$$

where the expectation is taken over the MDP dynamic according to a certain policy and the channel fading of desired links and interfering links, and p_s is obtained by averaging over the point process. The stable probability is the CCDF of the average transmit success probability for given θ and a given arrival rate ξ.

6.4.1 Greedy policy

This policy will be the baseline policy to which RL performance will be compared to. In the greedy algorithm, the BS is always active. We note this deterministic strategy as $\tilde{\pi}$. Given the realization of PPP, the average transmit success probability of greedy policy is

$$r_\Phi(\tilde{\pi}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}\left[f_s(1, S_c(t)) | S(0) = s, \Phi \right] \tag{6.25}$$

Correspondingly, the stable probability $\tilde{p}_s$ is

$$\tilde{p}_s = \mathbb{P}\left[r_\Phi(\tilde{\pi}) > \xi \right] \tag{6.26}$$

It is worth noting that in Chapter 5, we also calculated the stable probabilities for greedy policy. The difference is that, in the previous chapter, the transmission is successful or not depend on if the SIR exceed a given threshold; in this chapter, we divide the SIR into multiple regions where each region has a corresponding transmit success probability in order to adaptive reinforcement learning algorithm.

Special case M=2 We first discuss a special case $M = 2$ and $f_m - f_{m-1} = \Delta f$, where the SIR region is divided into three regions with increasing equal transmit success probability. To obtain the stable probability in (6.26), we give the moments of the average transmit success probability in (6.25), i.e. $M_b \triangleq \mathbb{E}[(r_\Phi(\tilde{\pi}))^b]$ as following.

Theorem 6.2 (Moments)**.** *When $M = 2$, $f_m - f_{m-1} = \Delta f$, the bth moments of the average transmit success probability for greedy policy is*

$$M_b = \Delta f^b \sum_{k=0}^{\infty} \binom{b}{k} \left[1 + \int_1^\infty \left[1 - \left(1 - \frac{q\theta_1}{\theta_1 + v^{\frac{\alpha}{2}}} \right)^k \left(1 - \frac{q\theta_2}{\theta_2 + v^{\frac{\alpha}{2}}} \right)^{b-k} \right] dv \right]^{-1}, \; b \in \mathbb{N} \tag{6.27}$$

Proof. Seen in Appendix A.11. □

Using the Gil-Pelaze inversion theorem, we obtain an exact integral expression of stable probability for greedy policy from the purely imaginary moments M_{iw}.

Theorem 6.3 (Stable probability)**.** *When $M = 2$, $f_m - f_{m-1} = \Delta f$, the stable probability for greedy policy is*

$$\tilde{p}_s = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{1}{w} \mathrm{Im}\left\{ u^{-iw} \psi_X(w) \right\} dw \tag{6.28}$$

where $\psi_X(w) = \Delta f^{iw} \sum_{k=0}^{\infty} \binom{iw}{k} \left[1 + \int_1^\infty \left[1 - \left(1 - \frac{q\theta_1}{\theta_1 + v^{\frac{\alpha}{2}}} \right)^{iw-k} \left(1 - \frac{q\theta_2}{\theta_2 + v^{\frac{\alpha}{2}}} \right)^k \right] dv \right]^{-1}$.

Proof. Let $X \triangleq \log(r_\Phi(\tilde{\pi}))$. The characteristic function of X is

$$\psi_X(w) \triangleq \mathbb{E}[e^{iwX}] = \mathbb{E}[(r_\Phi(\tilde{\pi}))^{iw}] = M_{iw}, w \in \mathscr{R} \tag{6.29}$$

Then by the Gil-Pelaze theorem, (6.28) is obtained. □

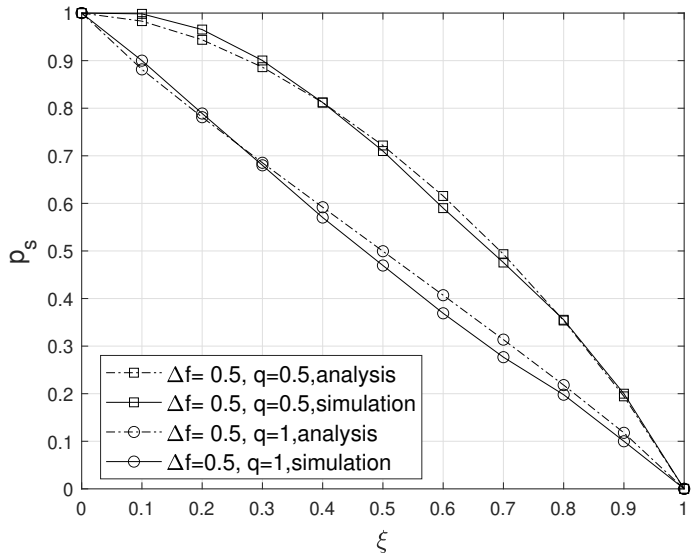


Figure 6.1 – The exact expression.

Figure 6.1 validates Theorem 6.1, based on two settings: $q = 0.5$ and $q = 1$, which means a randomly chosen interfering BS is active with probability 0.5 and 1, respectively. In the simulation, we truncated w into region [0,50], thus there is a small difference between the simulation and analysis in Figure 6.1. One drawback of theorem is that it is not easy to evaluate, and it will take a long time to converge. This is because the binomial equation $\binom{iw}{k}$ is subject to oscillatory convergence, which converges more slowly as w increases.

On the other hand, since $r_\Phi(\tilde{\pi})$ is supported on $[0,1]$, the beta distribution can be used to approximate the meta distribution of $r_\Phi(\tilde{\pi})$, thanks to [66]. Simulation results show that the approximation of the beta distribution is much simpler, and matches the simulation exactly (as shown later). The beta distribution is defined as follows:

Definition 6.1. *The probability density function (pdf) of a beta distributed random variable X with mean μ is [114]*

$$f_X(x) = \frac{x^{\frac{\mu(\beta-1)-1}{1-\mu}}(1-x)^{\beta-1}}{B(\mu\beta/(1-\mu),\beta)}, \tag{6.30}$$

where $B(\cdot,\cdot)$ is the β-function, with $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$, and $var[X] = \frac{\mu(1-\mu)^2}{\beta+1-\mu}$.

In our case, the mean and variance of $r_\Phi(\tilde{\pi})$ can be obtained according to the first moment and the second moment in (6.27), that is

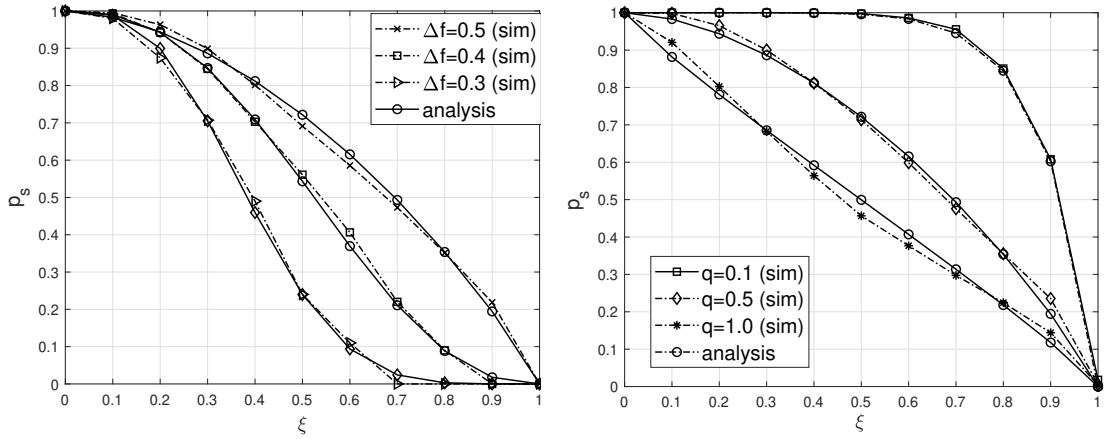
$$\mathbb{E}[r_\Phi(\tilde{\pi})] = M_1 \tag{6.31}$$

$$\mathrm{Var}[r_\Phi(\tilde{\pi})] = M_2 - M_1^2 \tag{6.32}$$

Then the beta distribution of $r_\Phi(\tilde{\pi})$ can be obtained according to Definition 6.1, the stable probability is then approximated as

$$\tilde{p}_s = \mathbb{P}\left[r_\Phi(\tilde{\pi}) > \xi\right] \approx \int_\xi^\infty f_{r_\Phi(\tilde{\pi})}(x)\,dx \tag{6.33}$$

Fig. 6.2a and 6.2b plots the stable probability $\tilde{p}_s$ vs. ξ labelled on Δf and q, respectively when $M = 2$. The analysis is based on Theorem 6.2 and Beta approximation. The stable probability $\tilde{p}_s$ is decreasing with ξ increases. For example, given the fixed ξ in Fig. 6.2a, it can be observed that $\tilde{p}_s$ is shifted to higher value when $\Delta f = 0.5$ compared with $\Delta f = 0.3$. This implies that the higher transmit success probability the larger stable probability. Besides, the active probability of interferers largely affects the stable region since a larger q results in a higher interference level and reduces the SIR and further reduces the size of the stable region. As we show in Fig. 6.2b, the lower the interfering base station activity, the higher the stable probability.



(a) The effect of Δf.

(b) The effect of q

Figure 6.2 – $\tilde{p}_s$ when $M = 2$.

General case M>2 We now consider the general case where SIR is divided into $M > 2$ regions. The moment generation function of transmit success probability can be obtained by using multinomial series.

Theorem 6.4. *When $M \in \mathbb{N}^+$, $f_m - f_{m-1} = \Delta f_m$, the moment of the transmit success probability for greedy policy is*

$$M_b = \sum_{\substack{n_1,n_2,\cdots,n_M \geq 0 \\ n_1+n_2+\cdots+n_M=b}} \frac{b!}{n_1!n_2!\cdots n_M!} \prod_{m=1}^{M} \Delta f_m^{n_m} \left[1 + 2\int_0^1 \left[1 - \prod_{m=1}^{M}\left(1 - \frac{q\theta_m}{\theta_m + v^{-\alpha}}\right)^{n_m}\right] v^{-3}\,dv\right]^{-1} \tag{6.34}$$

Proof. Seen in Appendix A.12. □

Corollary 6.1. *When $M \in \mathbb{N}^+$, $f_m - f_{m-1} = \Delta f$, the moment of the transmit success probability for greedy policy is*

$$M_b = \Delta f^b \sum_{\substack{n_1,n_2,\cdots,n_M \geq 0 \\ n_1+n_2+\cdots+n_M=b}} \frac{b!}{n_1!n_2!\cdots n_M!} \left[1 + 2\int_0^1 \left[1 - \prod_{k=1}^{M} \left(1 - \frac{q\theta_k}{\theta_k + v^{-\alpha}} \right)^{n_k} \right] v^{-3} dv \right]^{-1}, \forall b \in \mathbb{N} \tag{6.35}$$

Similarly, the mean and variance of $r_\Phi(\tilde{\pi})$ can be obtained by $\mu = M_1$ and $\text{Var} = M_2 - M_1^2$. The stable probability is then approximated using beta distribution mentioned in (6.30) and (6.33).

Fig. 6.3 plots the stable probability p_s when $M > 2$. We compared a range of SIR threshold and a range of corresponding transmit success probability f_s. It can be observed the simulations validates the proposed Beta approximation. The number of active interferers plays a significant role on the performance of stable probability.

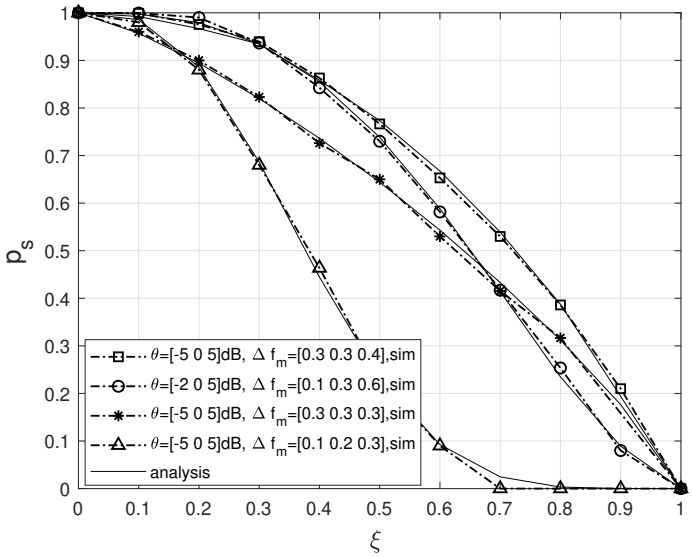


Figure 6.3 – p_s when $M > 2$.

6.4.2 RL-based policies

In this section, we investigate the stable probability of RL-based policies. Given Φ and policy π, the transmit success probability of stationary policy π is

$$r_\Phi(\pi) = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_{\pi, H_{x,t}, H_{x_0,t}} \left[\sum_{t=0}^{T} f_s(A(t), S_c(t)) | S(0) = s, \Phi \right] \tag{6.36}$$

and let ξ be the arrival rate, the stable probability p_s of RL-based policy is $p_s = \mathbb{P}[r_\Phi(\pi) > \xi]$, with respect to (6.24).

Remark 6.1. *The greedy policy provides an upper bound of the stable probability of Q-learning based policy.*

Proof. Combing (6.36), (6.24), (6.25) and (6.26), we have

$$p_s = \mathbb{P}\left[\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T} f_s(A(t), S_c(t))|S(0)=s,\Phi\right] > \xi\right] \tag{6.37}$$

$$\leq \mathbb{P}\left[\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f_s(1, S_c(t))|S(0)=s,\Phi\right] > \xi\right]$$

$$= \tilde{p}_s \tag{6.38}$$

and the proof is complete. □

Remark 6.1 implies that greedy policy will always provide better stable regions than RL-based policies. However, we find that by optimizing the Lagrange parameter λ, the RL-based policy can eventually maintain the same stable probability as the greedy policy at a lower transmission cost.

In the next, we characterize the bth moments of the transmit success probability considering the stationary policy π obtained by an RL-based algorithm.

Lemma 6.2. *Given the stationary policy π and the PPP realization Φ, the average transmit success probability is*

$$r_\Phi(\pi) = \sum_{m=1}^{M} b_m \prod_{x\in\Phi\setminus x_0}\left(\frac{q}{1+\theta_m\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right) \tag{6.39}$$

where $b_m = f_m\pi(1|s_c^m,\Phi) - f_{m-1}\pi(1|s_c^{m-1},\Phi)$.

Proof. Seen in Appendix A.13. □

Theorem 6.5. *Given the stationary policy π, the bth moment of transmit success probability is*

$$M_b = \sum_{\substack{n_1,n_2,\cdots,n_M\geq 0\\ n_1+n_2+\cdots+n_M=b}} \frac{b!}{n_1!n_2!\cdots n_M!}\left[1+\int_1^\infty\left[1-\prod_{m=1}^{M}\mathbb{E}_\Phi[b_m^{n_m}]\left(1-\frac{q\theta_m}{\theta_m+v^{\frac{\alpha}{2}}}\right)^{n_m}\right]dv\right]^{-1} \tag{6.40}$$

Proof. Seen in Appendix A.14. □

Lemma 6.3. *The average transmit success probability is the first-moment in (6.40), and has the following expression*

$$\mathbb{E}[r_\Phi] = \sum_{\substack{n_1,n_2,\cdots,n_M\geq 0\\ n_1+n_2+\cdots+n_M=1}}\left[1+\int_1^\infty\left[1-\prod_{m=1}^{M}\mathbb{E}_\Phi[b_m^{n_m}]\left(1-\frac{q\theta_m}{\theta_m+v^{\frac{\alpha}{2}}}\right)^{n_m}\right]dv\right]^{-1} \tag{6.41}$$

where $\mathbb{E}_\Phi[b_m^{n_m}] = \sum_{i=0}^{n_m}(-1)^{n_m-i}f_m^i f_{m-1}^{n_m-i}\mathbb{E}_\Phi\left[\pi(1|s_c^m,\Phi)^i\right]\mathbb{E}_\Phi\left[\pi(1|s_c^{m-1},\Phi)^{n_m-i}\right]$.

6.5 Numerical results

6.5.1 Simulation setup

Consider a PPP of density $\lambda = 0.25$ stations/km^2 on a area of 900 km^2. The transmit power is normalized to 1 for each base station and the path loss exponent is $\alpha = 4$. For each realization of the network, the state-action function of the typical agent is updated using a table for Q-learning and SARSA algorithms until convergence, i.e. when , $\forall o > 0 \left| q_{t+1}(s,a) - q_t(s,a) \right| \leq o$, $\forall (s,a) \in \mathscr{S} \times \mathscr{A}$. Then a new network realization is drawn and the process repeats. The simulation is repeated 5000 times.

Assume that the time is divided into slots with very small sizes such that, at each time slot, at most one packet can be transmitted or can arrive. The packets arrival process is modeled by the Bernoulli distribution with intensity ξ. At each time slot, the agent decides to transmit a packet or not, i.e., $A = 0$ or 1 respectively. In the simulation setup, we assume that SIR is divided in 3-regions delimited by two threshold, i.e., $\theta_1 = -1.47$ dB, $\theta_2 = 5.07$ dB. When SIR $< \theta_1$, the transmission failed; when $\theta_1 <$ SIR $< \theta_2$, a packet is transmitted successfully with probability $f_1 = 0.5$; when SIR $> \theta_2$, a packet is transmitted successfully with probability $f_2 = 1$. There exists a transmission cost during the transmission that depends on the SIR level. The transmission cost $C(A(t), S(t))$ decreases with the SIR increase. In this chapter, the non-increasing function defined in (6.3) is arbitrary chosen as: $f(a, s_1) = 1.7a$, $f(a, s_2) = 0.8a$ and $f(a, s_3) = 0.2a$ for all $a \in \mathscr{A}$, as in [14]. The choice of the cost function does not impact the behavior of algorithms, but only the absolute values. At a given time slot, if the base station does not transmit anything while the buffer is not empty, there is a delay cost defined in (6.4) $w = 0.6a$.

Parameters for RL training We consider the temporal learning approach to obtain the optimal $q^*(s,a)$, Q-learning and SARSA in Algorithm 4. Besides, we use the stochastic sub-gradient method to obtain the optimal value of the Lagrange multiplier λ as in Algorithm 5. We choose a discount factor of $\eta = 0.95$ in the optimization objective (η closer to 1 yields better performance after convergence, but lead to slower convergence). The packet arrival rate ξ is settled in a range $\xi \in [0,1]$packet/slot. The learning rate (or the step-size parameter) α_t is settled as 0.01 at each time slot, and the ϵ-greedy algorithm parameter is $\epsilon_t = 1/t$. Table 6.1 summarizes the parameters used in our MATLAB-based simulator.

6.5.2 Learning algorithm comparison

In Fig. 6.4, we compare the long-term total cost and delay cost defined in (6.7) of different policies. In each realization, the optimal policy is trained based on the framework in Section 6.2. It can be observed that RL-based algorithms largely reduces the long-term cost, especially when the traffic intensity ξ is low. Both Q-learning and SARSA can adjust the transmission policy according to the traffic intensity while greedy policy is insensitive since it remains all the time.

Fig. 6.5 compares the total cost and stable probability of the policies as a function of traffic intensity. For the same network configuration, the RL-based policy is able to maintain the same

Parameter	**Value**
BS density	0.25 BS/km^2
User density	5 UE/km^2
Channel coefficient	$H_{x,t} \sim \exp(1)$
Path loss exponent	$\alpha = 4$
Discounting factor	$\eta = 0.95$
Learning step	$\alpha_t = 0.01$
BS actions	{on,off}
SIR state threshold	$\{-1.47, 5.07\}$ dB
transmission cost	$\{1.7a, 0.8a, 0.2a\}$
delay cost	0.6(1- a)
Packet arrival rate	$\xi \in [0,1]$ packet/slot
Transmit success probability	$f_s = \{0, 0.5, 1\}$
Exploration factor	$\epsilon = 1/t$
Training length	20000 slots

Table 6.1 – Simulation parameters.

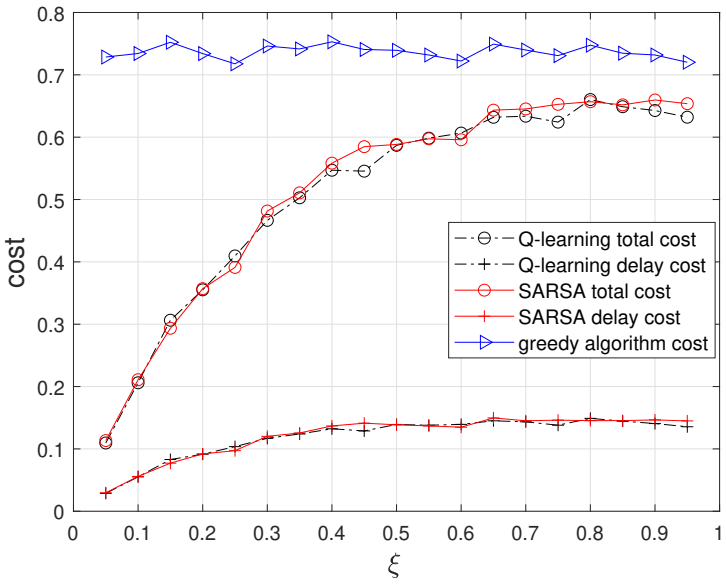


Figure 6.4 – The comparison of cost of different policies.

stability region at a lower transmission cost than the greedy policy. There is no significant difference in performance between Q-learning and SARSA algorithms that both converge to the optimal policy. We observe that there is a tradeoff between the stable probability and the total cost of the RL-based policies. Indeed, as the traffic intensity increases, the stable probability decreases and the agent tends to be more active by sending packets, which increases the transmission cost accordingly. RL-based policies allow agents to flexibly adjust the transmission policy according to traffic intensity and network configuration, while the greedy policy is only sensitive to buffer states.

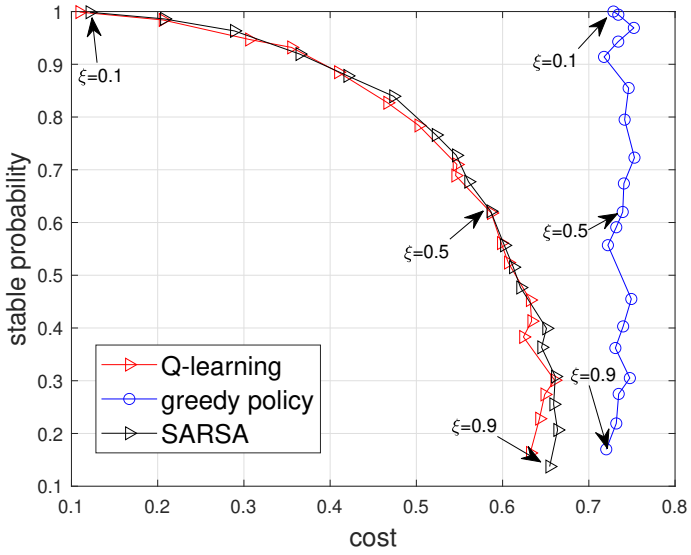


Figure 6.5 – The tradeoff between the stable probability and total cost.

Further, we analysis the policy obtained by Q-learning algorithm in average over the point process. For each PPP network, the optimal transmission policy can be obtained thanks to Q-learning and SARSA. In Fig. 6.6, we investigated the policy followed by the agent, averaged over all network realizations. The actions of the agent are either to transmit or remain silent while accounting one of the 12 states of the environment. The policy is completely described by the activity probability, which is defined as $\bar{\pi}_s = \mathbb{E}_\Phi[\pi(A=1|s, r_\Phi(s,\pi) > \xi)]$, i.e., the probability the agent be active in a state where the buffer does not diverge. We observe that the agent is more active when the SIR is good and the buffer is not empty. For example, in the state $[s_c^3, s_b=1, s_y=1]$, which corresponding to SIR $> \theta_2$, with non-empty buffer and a new arrived packet, the activity probability is 98.7%, because the success probability is high in this state. In the state $[s_c^1, s_b=0, s_y=0]$, the agent's activity probability nears to 0. The agent has a 66.7% probability of being active when it encounters a moderate level of SIR, while in the greedy policy, the agent is always active in this state.

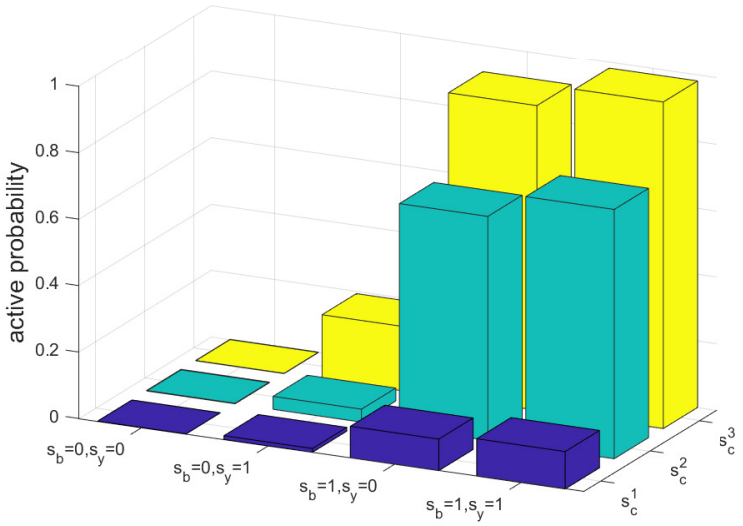


Figure 6.6 – Activity probability based on Q-learning.

6.6 Conclusion

In this chapter, we propose a reinforcement learning framework to compute the optimal transmission policy for BSs located in cell-0. We formulate a constrained optimization problem to minimize the long-term transmission cost with delay constraints. The problem is naturally a constrained Markov decision process due to the dynamic evolution of the queue.

First, we used the Lagrangian approach to transform the constrained MDP into an unconstrained MDP problem, which can be solved using dynamic programming. To obtain the optimal transmission policy, we considered two classical reinforcement learning algorithms, namely *Q*-learning and SARSA. Second, we analyze the probability of stability of greedy and RL-based transmission policies under this system, i.e., the probability that the average firing success rate is greater than the average packet arrival rate. By using stochastic geometry tools and applying Bellman's equations, we gave intrinsic insights into the performance of the RL-based transmission strategy. Finally, we validated the theoretical analysis through Monte Carlo simulations, and analyzed the performance of different transmission strategies. The simulation results show that the proposed RL-based transmission strategy is able to guarantee the same stability threshold at a lower transmission cost compared to the traditional greedy strategy.

Chapter 7

Conclusion and future works

7.1 Conclusion

This thesis deals with the study of downlink performance of dynamic stochastic cellular networks. The main issue addressed in this manuscript is the characterization of the stability region of the stochastic network when the traffic model is combined with the geometric description of the network. We first describe the stable coverage probability of the stochastic network, i.e., the coverage probability dissipated by the base station queue. Using the concept of dynamic coverage probability, we consider the interaction between queueing states in the network using a discrete Markov chain model of queues, where the service rate of typical users depends on the dynamic coverage probability. We explore the cases of finite and infinite queue sizes, respectively.

We also want to have a more detailed description of this phenomenon by answering the question "What is the proportion of unstable queues in the network?" In this case, we use the notion of ϵ-stable, which describes the set of traffic strengths in which a randomly selected queue has a divergence probability below ϵ. Finally, characterizing the stability region by considering the resource allocation is very difficult due to the dependence between the geometry and dynamics of the network and the allocation policy. To overcome this problem, we propose to use reinforcement learning algorithms to study stable probability. The dynamics of the network considered in this thesis is perfectly suited to be described by a Markovian decision process for which reinforcement learning strategies can be proposed. Thus, we study the performance when a typical base station can choose to transmit or remain silent depending on the observed network state.

In Chapter 2, we provided the mathematical tools from stochastic geometry and queueing theory which were later used in this thesis. For example, the PGFL function and Campbell's theorem are introduced to characterize the aggregate interference in the cellular network with irregular locations. The Slivnyak's theorem ensures that in a stable PPP network, evaluating the SINR of a device at the origin is sufficient to characterize the average performance of a large network. The Kendall's notation and matrix-analytical methods for discrete-time Markov chains are introduced to model the traffic conditions and queues evolution. Besides, we introduced the Markov decision process and classical reinforcement learning algorithms,

which were be further applied to investigate the transmission strategies issue.

In Chapter 3, we conducted a brief survey on spatio-temporal research approaches in large-scale networks from literature. Considering the random arrival of the packets, a recent line of research has introduced a spatio-temporal model to analyze the coverage probability from a joint spatial-temporal aspect [9, 10, 11]. To simplify the analysis, these works assume that the queueing evolution is independent and identically distributed among all the BSs. which is commonly known as *mean field approximation*.

However, in practical systems, multiple BSs sharing the same spectrum inevitably interfere with each other due to the broadcasting nature of wireless channels. As such, the queueing statues of the BSs are correlated in both time and space domains among all the BSs. The spatial distance and channel gain related interference affect the transmission success probability and the states of the interacting queues. Conversely, the buffer states in the previous time slot also affects the activation of the interfering BSs and the interference of the devices. In view of this, the interactions among the queues are dependent on both the spatial and temporal factors. The novelty of our work lies in taking into account the correlation between the interference created by all the base stations and the state of the queues at the level of the transmitters over time. Work in [12, 13] focused on the interaction between queue dynamics and network topology. However, performance analyzes focus on the coverage probability. System stability issues from the queue perspective are not well addressed, which is the focus of our work.

In Chapter 4, we characterized the SINR performance in dynamic cellular networks for different scenarios, i.e. finite buffer and infinite buffer. By leveraging proprieties from queueing theory, we analyzed the dynamic coverage probability, the stable coverage probability, the delay performance as well as the packet loss probabilities at the typical UE, and verified then with Monte Carlo simulations. In particular, the stable coverage probability of a typical user is inversely proportional to the traffic intensity, i.e. the higher intensity of the packet arrive, the lower the stable coverage. The influence of the packet buffer length to the coverage and packet loss probability is explicitly derived. We show in particular that small buffer length leads to a better coverage probability but also to a larger packet loss probability, advocating for a tradeoff between these two metrics.

One issue is that even with the same traffic arrival intensity at each BS, the spatial interaction of the queues may result in very different queue lengths after a sufficiently long period. Specifically, users located at the cell edge usually get a worse SINR than users located at the center, and thus have a lower transmission success probability and it leads to a long queue because of the retransmission scheme. In addition, the aggregated interference level is different at different BSs. For BSs that always experience high interference levels, the service rate slows down, resulting in longer queue lengths. Therefore, the coverage probability based on the average performance of the SINR is not sufficient to characterize the network. Thus we studied a more refined metric known as ϵ-stable region in Chapter 5. This concept is introduced in [29] allows us to answer the question: "What is the set of arrival rates at the desired SINR such that the proportion of unstable queues in the network is lower than ϵ?".

In Chapter 5, we provide closed-form expression of the upper and lower bounds of ϵ-stable region. Moreover, we propose an alternative definition of the ϵ-stable region and derived accordingly a tight approximation of the critical arrival rate that was unavailable in literature.

The results demonstrated that the packet critical arrival rate is close to the upper bound when the SINR threshold θ is relatively small, and is close to the lower bound when the SINR threshold θ is relatively large. This is because decreasing θ will increase the opportunity of a successful transmission, thus the active probability of the typical BS is much closer to the active probability in the favorable system. If θ increases, the success probability decreases, the BS holds a long-queue and tends to have a long period of time operating in the active state, which incurs a high interference to the nearby BSs and consequently decreases its successful transmission probability, which leads to the ϵ-stable region being closer to the one of full load system.

In a dynamic communication network, transmission strategies must be adapted according to the state of the network to satisfy an optimality criterion. However, it can be very difficult to analytically derive the optimal transmission strategy when the system becomes complex. In recent years, reinforcement learning has come back to the forefront in its application to network radio resource management, when the agent's interaction with its environment is modeled by a Markov decision process (MDP) [3]. The interest of this approach is that it allows to find an optimal transmission policy in an uncertain environment without an explicit physical model of the communication to perform the resource allocation, but only through trial and error on the part of the agent. In the field of digital communications, AR has been applied to point-to-point systems in order to minimize the energy consumed under delay constraints, e.g., [14, 15], to IoT networks for spectrum management and energy harvesting, e.g.,[115], as well as to the problem of caching in cellular networks using deep RL. However, no work has yet combined reinforcement learning with stochastic geometry in a dynamic network for the evaluation of average RL performance.

In Chapter 6, we deployed RL algorithms to explore transmission strategies for typical BS located at the origin. For simplicity, the powers of all inter-cell interfering BSs are normalized and active with constant probability p in each slot. Meanwhile, in this chapter we refer to [14] to consider a more realistic model, that is, the transmission success probability depends on the state of SINR, and the ARQ protocol is allowed to retransmit packets. We consider two temporal learning algorithms, namely Q-learning and SARSA, to learn the optimal transmission policy to adapt the specific network configurations (BS density, relative location of users to BS, packet arrival intensity, etc.) The experimental results show that while ensuring the same stable probability, the reinforcement learning-based algorithm can greatly reduce the transmission cost compared to greedy algorithm, especially when the packet arrival rate is low. Besides, the stable probability is given in closed-from both in greedy algorithm and RL-based algorithm and validated by simulations.

7.2 Future works

This work opens the direction of reinforcement learning in large-scale cellular networks, in which numerous topics warrant further investigation.

- The chapter 6 focuses on the presence of only one agent in the network, while the others still operate with a greedy policy, which does not make the problem symmetric.

The extension to multi-agent reinforcement learning is a necessary step to analyze the performance of the network with this strategy. However, the extension is non-trivial because instability problems may appear when considering decentralized distributed learning. The theoretical formulation of multi-agent learning and the conditions of convergence remain to be explored.

- Current work focuses on static networks, which means that transmitters and receivers do not change position during transmission. In a next step, mobility could be introduced, which raises the question of modeling charge flows through cells.
- In this work, we consider the FIFO scheduling of packet transmission. In the next work, the effect of different packet transmission scheduling, such as random scheduling and rotating scheduling, can be studied. Meanwhile, most of the current articles based on spatio-temporal models assume that the packet arrival process is based on Bernoulli's distribution. In practice, the traffic varies dramatically from time to time, with traffic arrivals tending to saturation during peak periods and troughs during low periods. The next work can consider more complex spatio-temporal models to approximate the real situation.

Appendix A

Proofs

A.1 Proof of Theorem 4.1

Given the typical UE received data at time slot t, its dynamic coverage probability is written as (to lighten the notation we remove the index t from the channel coefficients)

$$
\begin{aligned}
p_t &= \mathbb{P}^{x_0}(\gamma_t \geq \theta) \\
&= \mathbb{P}^{x_0}\left[\frac{H_{x_0,t}\,\|x_0\|^{-\alpha}}{\sigma^2 + \sum\limits_{x\in\Phi\setminus x_0} H_{x,t}\,\|x\|^{-\alpha}\,\mathbb{1}(x\in\Phi_t)} \geq \theta\right] \\
&\overset{(a)}{=} \int_0^\infty 2\pi\lambda r_0 e^{-\pi\lambda r_0^2}\exp(-\sigma^2\theta r_0^\alpha)\times\mathbb{P}^{x_0}\left[\frac{H_{x_0,t}\,\|x_0\|^{-\alpha}}{\sigma^2 + \sum\limits_{x\in\Phi\setminus x_0} H_{x,t}\,\|x\|^{-\alpha}\,\mathbb{1}(x\in\Phi_t)} \geq \theta \,\middle|\, \|x_0\| = r_0\right] \mathrm{d}r_0 \\
&= \int_0^\infty 2\pi\lambda r_0 e^{-\pi\lambda r_0^2} e^{-\sigma^2\theta r_0^\alpha}\,\mathscr{L}_I(\theta r_0^\alpha)\mathrm{d}r_0
\end{aligned}
\tag{A.1}
$$

where the Laplace transform (LT) of a random variable X in s is denoted as $\mathscr{L}_X(s)$, and (a) follows the distribution of r_0 as $f_{r_0}(r) = e^{-\pi\lambda r^2}2\pi\lambda r$.

The LT $\mathscr{L}_I(s)$ in (A.1), with $s = \theta\|x_0\|^\alpha = \theta r_0^\alpha$, has the form

$$
\begin{aligned}
\mathscr{L}_I(s) &= \mathbb{E}_{\{H_x\},\Phi}\left[\prod_{x\in\Phi\setminus x_0}\exp\left(-sH_x\,\|x\|^{-\alpha}\,\mathbb{1}(x\in\Phi_t)\right)\Big| r_0\right] \\
&\overset{(a)}{=} \mathbb{E}_\Phi\left[\prod_{x\in\Phi\setminus x_0}\mathbb{E}_{H_x}\left[\exp\left(-sH_x\,\|x\|^{-\alpha}\,\mathbb{1}(x\in\Phi_t)\right)\right]\Big| r_0\right] \\
&\overset{(b)}{=} \mathbb{E}_\Phi\left[\prod_{x\in\Phi\setminus x_0}\left(\frac{\mathbb{E}_{\mathbb{1}(x\in\Phi_t)}[\mathbb{1}(x\in\Phi_t)=1]}{1+s\,\|x\|^{-\alpha}\times 1} + \frac{\mathbb{E}_{\mathbb{1}(x\in\Phi_t)}[\mathbb{1}(x\in\Phi_t)=0]}{1+s\,\|x\|^{-\alpha}\times 0}\right)\Big| r_0\right] \\
&\overset{(c)}{=} \mathbb{E}_\Phi\left[\prod_{x\in\Phi\setminus x_0}\left(\frac{q_t}{1+s\,\|x\|^{-\alpha}} + 1 - q_t\right)\Big| r_0\right]
\end{aligned}
\tag{A.2}
$$

where (a) follows from the i.i.d. hypothesis of H_x and further independence from the point process Φ, (b) follows from the law of total expectation and using independence activity of BS [12, Assumption 2], and (c) follows from (4.2).

According to the PGFL of PPP and with $r = \|x\|$, we have

$$\begin{aligned}\mathscr{L}_I(\theta r_0^{\alpha}) &= \exp\left(-2\pi\lambda\int_{r_0}^{\infty}\left(1-\left(\frac{q_t}{1+\theta r_0^{\alpha} r^{-\alpha}}+1-q_t\right)\right) r\mathrm{d}r\right) \\ &\overset{(a)}{=} \exp\left(-\pi\lambda r_0^2\int_1^{\infty}\frac{q_t}{1+u^{\frac{\alpha}{2}}\theta^{-1}}\mathrm{d}u\right)\end{aligned} \tag{A.3}$$

where (a) is obtained by the change of variable $u = (\frac{r}{r_0})^2$.

Combing (A.1), the dynamic coverage probability has the expression

$$p_t(\theta,\xi) = 2\pi\lambda\int_0^{\infty} e^{-\sigma^2\theta r^{\alpha}} e^{-\pi\lambda r^2\left(1+q_t\rho(\alpha,\theta)\right)} r\mathrm{d}r$$

where $\rho(\alpha,\theta) = \int_1^{\infty}[1+u^{\frac{\alpha}{2}}\theta^{-1}]^{-1}du$.

A.2 Proof of Eq. 4.15

The solution of (4.13) is the solution of

$$\begin{cases} \bar{\xi}x_0 + \bar{\xi}px_1 = x_0 \\ \xi x_0 + (\bar{\xi}\bar{p}+\xi p)x_1 + \bar{\xi}px_2 = x_1 \\ \xi\bar{p}x_1 + (\bar{\xi}\bar{p}+\xi p)x_2 + \bar{\xi}px_3 = x_2 \\ \xi\bar{p}x_2 + (\bar{\xi}\bar{p}+\xi p)x_3 + \bar{\xi}px_4 = x_3 \\ \qquad\vdots \end{cases} \tag{A.4}$$

Following the matrix-analytical method introduced in Section 2.2.2, the solution of (A.4) is

$$x_i = R^i\frac{x_0}{\bar{p}}, \text{ where } R = \frac{\xi\bar{p}}{\bar{\xi}p},\ \forall\ i\in[1,+\infty) \tag{A.5}$$

By the law of total probability we should have $\sum_{i=0}^{\infty}x_i = 1$, it comes

$$x_0 + \frac{x_0}{\bar{p}}\sum_{i=1}^{\infty}R^i \overset{(a)}{=} x_0\left(1+\frac{1}{\bar{p}}\times\frac{R}{1+R}\right) = 1 \tag{A.6}$$

where (a) comes from geometric series on the condition $R<1$, i.e. $p>\xi$. After straightforward algebraic manipulation, the final expression of x_0 is

$$x_0 = \frac{p-\xi}{p}, \forall\ p>\xi \tag{A.7}$$

When $R>1$, the geometric series $\sum_{i=1}^{\infty}R^i$ diverges and the solution of (A.6) is $x_0 = 0$.

A.3 Proof of Lemma 5.2

The bth moment of the transmit success probability in full load is

$$\left[1+\int_1^{\infty}\left[1-\left(\frac{1}{1+\theta v^{-\frac{\alpha}{2}}}\right)^b\right]dv\right]^{-1} = 1+2\int_0^1\left[1-\left(\frac{1}{1+\theta r^{\alpha}}\right)^b\right]r^{-3}dr \tag{A.8}$$

$$\overset{(a)}{=} 1+\frac{2}{\alpha}\int_0^1\left[1-\frac{1}{(1+\theta u)^b}\right]u^{-\frac{2}{\alpha}-1}du \tag{A.9}$$

$$\overset{(b)}{=} 1-\underbrace{\left[1-\frac{1}{(1+\theta u)^b}\right](u^{-\frac{2}{\alpha}})\Big|_0^1}_{A}+\int_0^1 u^{-\frac{2}{\alpha}}d\left[1-\frac{1}{(1+\theta u)^b}\right] \tag{A.10}$$

$$= 1-A+\theta b\int_0^1 u^{-\frac{2}{\alpha}}(1+\theta u)^{-b-1}du \tag{A.11}$$

where (a) comes from changing variable $r^{\alpha}=u$, and (b) comes from the partial integration.

When $u=1$, we have $A=1-(1+\theta)^{-b}$, when $u=0$, A will lead to a indeterminate form. However, by applying Taylor approximation, we have

$$(1+\theta u)^{-b} = 1-b\theta u+\mathcal{O}(\theta^2u^2) \tag{A.12}$$

$$\Rightarrow\left[1-\frac{1}{(1+\theta u)^b}\right](u^{-\frac{2}{\alpha}}) \approx b\theta u^{1-\frac{2}{\alpha}} \tag{A.13}$$

Since α is the pass loss exponent satisfies $\alpha\geq 2$, lead to $1-\frac{2}{\alpha}\geq 0$, lead to $u^{1-\frac{2}{\alpha}}|_{u=0}\to 0$, lead to $A|_{u=0}=0$. Recalling (A.11), we have

$$1-A+\theta b\int_0^1 u^{-\frac{2}{\alpha}}(1+\theta u)^{-b-1}du = (1+\theta)^{-b}+\theta b\int_0^1 u^{-\frac{2}{\alpha}}(1+\theta u)^{-b-1}du \tag{A.14}$$

$$= {}_2F_1(b,-\frac{2}{\alpha};1-\frac{2}{\alpha};-\theta) \tag{A.15}$$

A.4 Proof of Theorem 5.1

Let $Y_l \triangleq \ln(\mu_l)$, the characteristic function of Y_l is

$$\varphi_{Y_l}(w) \triangleq \mathbb{E}[e^{iwY_l}] = \mathbb{E}[\mu_l^{iw}] = M_{iw}, w\in\mathscr{R}, i=\sqrt{-1} \tag{A.16}$$

Using the Gil-Pelaez theorem, and assigning b in (5.2) as $b=iw$, we obtain an exact integral expression for the CCDF of Y_l

$$\begin{aligned}\mathbb{P}(Y_l<\ln u) &= \frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{\mathrm{Im}[e^{-iw\ln u}\varphi_{Y_l}(iw)]}{w}dw\\ &= \frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{\mathrm{Im}[u^{-iw}\varphi_{Y_l}(iw)]}{w}dw\end{aligned} \tag{A.17}$$

Then the probability of μ_l is lower than the average arrival rate ξ follows

$$\mathbb{P}_\Phi\{\mu_l < \xi\} = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{\frac{\xi^{-iw}}{1+\int_1^\infty\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-iw}\right]dv}\right\}dw \tag{A.18}$$

The corresponding lower bound of ϵ-stability region is

$$\mathscr{S}_\epsilon^l = \left\{\xi \in \mathbb{R}^+ : \mathbb{P}\left[\mu_l \le \xi\right] \le \epsilon\right\} \tag{A.19}$$

$$= \left\{\xi \in \mathbb{R}^+ : \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{w} \times \mathrm{Im}\left\{\frac{\xi^{-iw}}{1+\int_1^\infty\left[1-\left(1+\theta v^{-\alpha/2}\right)^{-iw}\right]dv}\right\}dw \le \epsilon\right\} \tag{A.20}$$

A.5 Proof of Lemma 5.3

By applying the Markov inequality [116], the probability of μ_l lower than ξ satisfied

$$\mathbb{P}\{\mu_l < \xi\} = 1 - \mathbb{P}\{\mu_l \ge \xi\}, \tag{A.21}$$

$$\text{and } \mathbb{P}\{\mu_l \ge \xi\} < \frac{\mathbb{E}_\Phi[\mu_l^n]}{\xi^n} = \xi^{-n}\mathbb{E}_\Phi[\mu_l^n] \tag{A.22}$$

Combining the expression of μ_l in (5.5), $\mathbb{E}_\Phi[\mu_l^n]$ follows

$$\begin{aligned}\mathbb{E}_\Phi[\mu_l^n] =& \mathbb{E}_\Phi\left[\prod_{x\in\Phi\setminus x_0}\left(\frac{1}{1+\theta\left\|x_0\right\|^\alpha\left\|x\right\|^{-\alpha}}\right)^n\right] \\ \overset{(a)}{=}& \left[1+\int_1^\infty\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-n}\right]dv\right]^{-1}\end{aligned} \tag{A.23}$$

where (a) can be directly obtained from (5.5) and (5.6) by assigning b as $b = n$.

Combining (A.21) and (A.23), we have

$$\mathbb{P}\{\mu_l < \xi\} \ge 1 - \xi^{-n}\left[1+\int_1^\infty\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-n}\right]dv\right]^{-1} \tag{A.24}$$

Let $S_l = \{\xi \in [0,1] : \mathbb{P}[\mu_l < \xi] \le \epsilon\}$, indicating the ϵ-stable region for the full load system. We have

$$S_l \subset \bigcup_{n\in\mathbb{N}^+}\left\{\xi \in [0,1] : 1 - \xi^{-n}\left[1+\int_1^\infty\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-n}\right]dv\right]^{-1} < \epsilon\right\} \tag{A.25}$$

Taken the supremum of both sides of (A.25) results in

$$\sup S \ge \max_{n\in\mathbb{N}^+}\left[(1-\epsilon)\left[1+\int_1^\infty\left[1-\left(1+\theta v^{-\frac{\alpha}{2}}\right)^{-n}\right]dv\right]\right]^{-\frac{1}{n}} \tag{A.26}$$

A.6 Proof of Theorem 5.2

A favorable system is considered for the upper bound. If the transmission of a packet fails, this packet is dropped instead of being re-transmitted. The interfering transmitter is then active with probability ξ, which is $\mathbb{P}(\mathbb{1}(x\in\Phi_t))=\xi$. Let μ_u be the transmit success probability at typical BS conditioned on Φ in the favorable system, it follows

$$\begin{aligned}\mu_u &= \mathbb{E}\left[\exp\left(-\theta\left\|x_0\right\|^{\alpha}\sum_{x\in\Phi_b\setminus x_0}H_x\mathbb{1}(x\in\Phi_t)\left\|x\right\|^{-\alpha}\right)\middle|\Phi\right]\\ &= \prod_{x\in\Phi\setminus x_0}\left(\frac{\xi}{1+\theta\left\|x_0\right\|^{\alpha}\left\|x\right\|^{-\alpha}}+1-\xi\right)\end{aligned} \tag{A.27}$$

We define Y_u as $Y_u \triangleq \ln(\mu_u)$, and follow the similar steps as Lemma 5.2, the bth moment generation function of Y_u is

$$M_b^u = \left[1+\int_1^{\infty}\left[1-\left(1-\frac{\xi\theta}{\theta+v^{\frac{\alpha}{2}}}\right)^b\right]dv\right]^{-1} \tag{A.28}$$

According to the Gil-Pelaez Theorem, the probability of μ_u is lower than the average arrival rate ξ is

$$\mathbb{P}\{\mu_u<\xi\} = \frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{1}{w}\times\left\{\frac{\xi^{-iw}}{1+\int_1^{\infty}\left[1-\left(1-\frac{\xi\theta}{\theta+v^{\alpha/2}}\right)^{iw}\right]dv}\right\}dw \tag{A.29}$$

The corresponding upper bound of ϵ-stable region is

$$\begin{aligned}\mathscr{S}_{\epsilon}^{u} &= \left\{\xi\in\mathbb{R}^{+}:\mathbb{P}\left[\mu_n^{\Phi}\le\xi\right]\le\epsilon\right\}\\ &= \left\{\xi\in\mathbb{R}^{+}:\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{1}{w}\times\mathrm{Im}\left\{\frac{\xi^{-iw}}{1+\int_1^{\infty}\left[1-\left(1-\frac{\xi\theta}{\theta+v^{\alpha/2}}\right)^{iw}\right]dv}\right\}dw\le\epsilon\right\}\end{aligned} \tag{A.30}$$

Therefore, we get the result in Theorem 5.2.

A.7 Proof of Corollary 5.1

The cumulative distribution function of $\mathbb{P}(\mu_u)$ is

$$\mathbb{P}\{\mu_u<\xi\} = \mathbb{P}\{\mu_u^n<\xi^n\} = \mathbb{P}\{e^{-n\ln(\mu_u)}>e^{-n\ln\xi}\} \tag{A.31}$$

By applying the Markov inequality, we obtain

$$\begin{aligned}\mathbb{P}\{\mu_u<\xi\} &< \frac{1}{e^{-n\ln\xi}}\mathbb{E}\left[e^{-n\ln(\mu_u)}\right]\\ &= \xi^n\left[1+\int_1^{\infty}\left[1-\left(1-\frac{\xi\theta}{\theta+v^{\frac{\alpha}{2}}}\right)^{-n}\right]dv\right]^{-1}\end{aligned} \tag{A.32}$$

Since the above inequality holds for all $n \in \mathbb{N}^+$, we have

$$S_\epsilon \supset \bigcup_{n\in\mathbb{N}^+} \left\{ \xi \in \mathbb{R}^+ : \xi^n \left[1 + \int_1^\infty \left[1 - \left(1 - \frac{\xi\theta}{\theta + v^{\frac{\alpha}{2}}} \right)^{-n} \right] dv \right]^{-1} \leq \epsilon \right\} \tag{A.33}$$

Taken the supremum of both sides of (A.33) results in $\xi_c < \xi_c^u < \tilde{\xi}_c^u$, where $\tilde{\xi}_c^u$ is the solution of the fixed-point equation

$$(\tilde{\xi}_c^u)^n = \epsilon^{\frac{1}{n}} \left[1 + \int_1^\infty \left[1 - \left(1 - \frac{\tilde{\xi}_c^u \theta}{\theta + v^{\frac{\alpha}{2}}} \right)^{-n} \right] dv \right]^{\frac{1}{n}}, \ \forall n \in \mathbb{N}^+ \tag{A.34}$$

A.8 Proof of Remark 5.3

We separately prove the upper and lower bound of the critical arrival rate approach to 0 when $\theta \to 0$. Remark 5.3 is then obtained according to the squeeze theorem [117].

For the lower bound of critical arrival rate

$$\begin{aligned} \xi_c^l &= \sup \left\{ \xi \in \mathbb{R}^+ : \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{w} \times \mathrm{Im} \left\{ \frac{\xi^{-iw}}{1 + \int_1^\infty \left[1 - \left(1 + \theta v^{-\alpha/2} \right)^{-iw} \right] dv} \right\} dw \leq \epsilon \right\} \\ &\overset{(a)}{=} \sup \left\{ \xi \in \mathbb{R}^+ : \frac{1}{2} - \frac{1}{\pi} \times \int_0^\infty \frac{1}{w} \mathrm{Im} \left\{ \frac{\cos(\frac{1}{2} w \log(\xi^2)) - \mathrm{i} * \sin(\frac{1}{2} w \log(\xi^2))}{1 + \int_{\arctan(\frac{1}{\sqrt{\theta}})}^{\frac{\pi}{2}} \left[1 - \left(\frac{\theta \tan^2 x}{\theta \tan^2 x + \theta} \right)^{\mathrm{iw}} \right] \mathrm{d}\sqrt{\theta} \tan x} \right\} \mathrm{dw} \leq \epsilon \right\} \end{aligned} \tag{A.35}$$

where (a) follows from variable changing $v = \sqrt{\theta} \tan x$. We see

$$\lim_{\theta\to 0} \int_{\arctan(\frac{1}{\sqrt{\theta}})}^{\frac{\pi}{2}} \left[1 - (\tan x \sin 2x)^s \right] d\sqrt{\theta} \tan x \tag{A.36}$$

$$= \lim_{\theta\to 0} \sqrt{\theta} \int_{\arctan(\frac{1}{\sqrt{\theta}})}^{\frac{\pi}{2}} d\tan x - 2 \int_{\arctan(\frac{1}{\sqrt{\theta}})}^{\frac{\pi}{2}} \sin^{2s} x \cos^{s-2} x dx \tag{A.37}$$

$$= \lim_{\theta\to 0} \sqrt{\theta} \int_{\frac{\pi}{2}}^{\frac{\pi}{2}} d\tan x - 2 \int_{\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2s} x \cos^{s-2} x dx \tag{A.38}$$

$$= 0 \tag{A.39}$$

further lead to

$$\begin{aligned}
\lim_{\theta\to 0}\xi_c^l &= \sup\left\{\xi\in\mathbb{R}^+ : \frac{1}{2}-\frac{1}{\pi}\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{\frac{\xi^{-iw}}{1+\int_1^\infty\left[1-1^{iw}\right]dv}\right\}dw\le\epsilon\right\} \\
&= \sup\left\{\xi\in\mathbb{R}^+ : \frac{1}{2}-\frac{1}{\pi}\times\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{\xi^{-iw}\right\}dw\le\epsilon\right\} \\
&= \sup\left\{\xi\in\mathbb{R}^+ : \frac{1}{2}+\frac{1}{\pi}\times\frac{\pi}{2}\mathrm{sgn}(\ln(\xi))\le\epsilon\right\} \\
&= \sup\left\{\xi\in\mathbb{R}^+ : \frac{1}{2}+\frac{1}{\pi}\times\frac{\pi}{2}(-1)\le\epsilon\right\} \qquad \text{(A.40)} \\
&\overset{(a)}{=} \sup\left\{\xi\in\mathbb{R}^+ : 0\le\epsilon\right\} \\
&= 1 \qquad \text{(A.41)}
\end{aligned}$$

where (a) comes from $\mathrm{sgn}(\ln(\xi)) = -1, \forall\xi\in[0,1]$. This lead to the critical arrival rates are intuitively reasonable for any ϵ.

Similarly, the upper bound of critical arrival rate has the expression

$$\lim_{\theta\to 0}\xi_c^u = \sup\{\xi\in[0,1] : 0\le\epsilon\} = 1 \qquad \text{(A.42)}$$

According to the squeeze theorem, we obtain the results in Remark 5.3.

A.9 Proof of Theorem 5.3

Based on Lemma 5.1, the transmit success probability has the expression

$$\mu = \prod_{x\in\Phi\setminus x_0}\left(\frac{q}{1+\theta\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right) \qquad \text{(A.43)}$$

Defining $Y \triangleq \ln\mu$, the moment generating function of Y is

$$\begin{aligned}
&\mathbb{E}\left[\exp(sY)\right] \\
&= \mathbb{E}_\Phi\left[\prod_{x\in\Phi\setminus x_0}\left(\frac{q}{1+\theta\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right)^s\right] \\
&\overset{(a)}{=} \mathbb{E}_\Phi\left[\exp\left(-\lambda\int_{\|x_0\|}^\infty\left[1-\left(\frac{q}{1+\theta\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right)^s\right]d\|x\|\right)\right] \\
&\overset{(b)}{=} \int_0^\infty 2\pi\lambda r\exp(-\lambda\pi r^2)\times\exp\left(-2\pi\lambda\times\int_r^\infty\left[1-\left(\frac{q}{1+\theta\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right)^s\right]xdx\right)dr \\
&\overset{(c)}{=} \int_0^\infty 2\pi\lambda r e^{-\lambda\pi r^2}\exp\left(-\lambda\pi r^2\int_1^\infty\left[1-\left(\frac{q}{1+\theta v^{-\frac{\alpha}{2}}}\right)^s\right]dv\right)dr \\
&= \left[1+\int_1^\infty\left[1-\left(1-\frac{q\theta}{\theta+v^{\frac{\alpha}{2}}}\right)^s\right]dv\right]^{-1} \qquad \text{(A.44)}
\end{aligned}$$

where (a) follows from the probability generation functional of the PPP; (b) is obtained by using the PDF of $\|x_0\|$, which is $f_{\|x_0\|}(r) = 2\pi\lambda re^{\pi\lambda r^2}dr$, and the approximation that the correlation

between active BSs is ignored when DTMC has converged [12]; (c) is obtained using the change of variable $v^{\frac{1}{2}} = \frac{\|x\|}{\|x_0\|}$.

Aforesaid (5.17), $\mathbb{E}_\Phi[q] = \xi/\mathbb{E}_\Phi[\mu], \forall\ \mathbb{E}_\Phi[\mu] > \xi$. And it can be noticed that $\mathbb{E}_\Phi[\mu]$ is the particular case when $s = 1$ in (A.44). After straightforward algebraic manipulations, we have

$$\mathbb{E}[q] = \begin{cases} \frac{\xi}{1-\theta\xi\rho(\theta,\alpha)}, & \text{if } \frac{1}{1+\theta\rho(\theta,\alpha)} > \xi \\ 1, & \text{if } \frac{1}{1+\theta\rho(\theta,\alpha)} \leq \xi \end{cases} \tag{A.45}$$

where $\rho(\alpha,\theta) = \int_1^\infty [\theta + u^{\frac{\alpha}{2}}]^{-1} du$.

The CDF of Y, denoted by $F_Y(u) = \mathbb{P}[Y \leq u]$, follows from the Gil-Pelaez's Theorem as

$$\begin{aligned} F_Y(u) &= \mathbb{P}(Y < \ln(u)) \\ &= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{w}\mathrm{Im}\left\{ \frac{u^{-iw}}{1+\int_1^\infty \left[1-\left(1-\frac{k\theta}{\theta+v^{\alpha/2}}\right)^{iw}\right] dv} \right\} dw \end{aligned}$$

and the proof is complete.

A.10 Proof of Lemma 6.1

Given λ, the state-value function for a policy π given s and Φ can be expressed as

$$L_\pi(s,\lambda) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \eta^t L(t,\lambda) | S(0) = s, \Phi\right] \tag{A.46}$$

$$= \mathbb{E}_\pi\left[L(0,\lambda) + \eta\sum_{t=1}^{\infty} \eta^{t-1} L(t,\lambda)\Big| s, \Phi\right] \tag{A.47}$$

$$= \mathbb{E}_\pi\left[L(0,\lambda) + \eta L_\pi(s',\lambda)\ \Big|\ S(0) = s, \Phi\right] \tag{A.48}$$

$$= \sum_a \pi(a|s) \sum_{s'\in\mathcal{S}} \sum_{l\in\mathcal{L}} p(s',l|s,a)\left[L(0,\lambda) + L_\pi(s',\lambda)\right] \tag{A.49}$$

where $\pi(a|s) = \mathbb{P}(A = a|S = s)$. Similarly, the action-value function $q_\pi(s,a) : \mathcal{S}\times\mathcal{A} \to \mathbb{R}$, which satisfies

$$\begin{aligned} q_\lambda(s,a) &= \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \eta^t L(t,\lambda) | S(0) = s, A(0) = a, \Phi\right] \\ &= \sum_{s'\in\mathcal{S}} \sum_{l\in\mathcal{L}} p(s,l|s',a)\left[L(0,\lambda) + L_\pi(s',\lambda)\right] \end{aligned} \tag{A.50}$$

The relationship between state values and Q-values is

$$L_\pi(s,\lambda) = \sum_a \pi(a|s) q_\lambda(s,a) \tag{A.51}$$

The goal of solving unconstrained MDP is to find an optimal policy to obtain a minimum cost. An optimal policy, can be defined from the perspective of state-value function, as

$$L_\pi^*(s,\lambda) = \min_\pi L_\pi(s,\lambda), s \in \mathscr{S} \tag{A.52}$$

And for the optimal Q-values, we have

$$q_\lambda^*(s,a) = \min_a q(s,a), s \in \mathscr{S}, a \in \mathscr{A} \tag{A.53}$$

Substituting (A.51) to (A.53), the optimal state value equation in (A.49) can be reformulated as

$$L_\pi^*(s,\lambda) = \min_a q_\lambda^*(s,a) \tag{A.54}$$

where the fact that $\sum_a \pi(a|s) q_\lambda^*(s,a) \geq \min_a q_\lambda^*(s,a)$ was applied to obtain (A.54). Note that the optimal state value equation is a minimization over the action space instead of the strategy space. By combing (A.49) and (A.53) and (A.54), we have the following dynamic programming:

$$L_\pi^*(s,\lambda) = \min_a \sum_{s' \in \mathscr{S}} \sum_{l \in \mathscr{L}} p(s',l|s,a) \left[l + L_\pi^*(s',\lambda) \right] \tag{A.55}$$

$$q_\lambda^*(s,a) = \sum_{s' \in \mathscr{S}} \sum_{l \in \mathscr{L}} p(s',l|s,a) \left[l + \min_{a'} q_\lambda^*(s',a') \right] \tag{A.56}$$

A.11 Proof of Theorem 6.2

For greedy policy, the average transmit success probability is

$$r_\Phi(\tilde{\pi}) = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=0}^{T} f_s(S_c(t)) | S(0) = s, \Phi \right] \tag{A.57}$$

$$= \mathbb{E}\left[f_s(S_c) | S(0) = s, \Phi \right] \tag{A.58}$$

$$= \sum_{m=1}^{M-1} f_m \mathbb{P}(\theta_m < \gamma < \theta_{m+1} | \Phi) + f_M \mathbb{P}(\gamma > \theta_M | \Phi) \tag{A.59}$$

$$\overset{(a)}{=} \sum_{m=1}^{M} (f_m - f_{m-1}) \mathbb{P}(\gamma > \theta_m | \Phi) \tag{A.60}$$

$$\overset{(b)}{=} \sum_{m=1}^{M} (f_m - f_{m-1}) \prod_{x \in \Phi \setminus x_0} \left(\frac{q}{1 + \theta_m \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q \right) \tag{A.61}$$

where in (a) we set $f_0 = 0, f_{M+1} = 1$ for convenience, (b) follows the similar to the steps in (A.2) in Appendix A.1.

Then the moment generation function of M_b has the expression

$$M_b = = \mathbb{E}_\Phi \left[\left(\sum_{m=1}^{M} (f_m - f_{m-1}) \mathbb{P}(\gamma \geq \theta_m | \Phi) \right)^b \right] \tag{A.62}$$

$$= \mathbb{E}_\Phi \left[\left(\sum_{m=1}^{M} (f_m - f_{m-1}) \prod_{x \in \Phi \setminus x_0} \left(\frac{q}{1 + \theta_m \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q \right) \right)^b \right] \tag{A.63}$$

There are two ways to derive (A.63), which lead to the same expression. We detailed separately.

Method 1 Since (A.63) depends on the BS locations only through the relative distance, we can apply the PGFL of the RDP defined in (2.12) and directly obtain

$$M_b = \Delta f^b \sum_{k=0}^{b} \binom{b}{k} \left[1 + 2 \int_0^1 \left[1 - \left(1 - \frac{q\theta_1}{\theta_1 + u^{-\alpha}} \right)^k \left(1 - \frac{q\theta_2}{\theta_2 + u^{-\alpha}} \right)^{b-k} \right] u^{-3} du \right]^{-1}, \forall b \in \mathbb{N} \tag{A.64}$$

Method 2 We assume $f_m - f_{m-1} = \Delta f,\ M = 2$ thus

$$M_b = \mathbb{E}_\Phi \left[\left(\sum_{m=1}^{2} \Delta f \mathbb{P}(\gamma \geq \theta_m | \Phi) \right)^b \right] \tag{A.65}$$

$$= \Delta f^b \mathbb{E}_\Phi \left[\left(\prod_{x \in \Phi \setminus x_0} \underbrace{\left(\frac{q}{1 + \theta_1 \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q \right)}_{f(\theta_1, x)} + \prod_{x \in \Phi \setminus x_0} \underbrace{\left(\frac{q}{1 + \theta_2 \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q \right)}_{f(\theta_2, x)} \right)^b \right] \tag{A.66}$$

$$\overset{(a)}{=} \Delta f^b \mathbb{E}_\Phi \left[\sum_{k=0}^{b} \binom{b}{k} \prod_{x \in \Phi \setminus x_0} f(\theta_2, x)^k \prod_{x \in \Phi \setminus x_0} f(\theta_1, x)^{b-k} \right] \tag{A.67}$$

$$= \Delta f^b \sum_{k=0}^{b} \binom{b}{k} \mathbb{E}_\Phi \left[\prod_{x \in \Phi \setminus x_0} f(\theta_2, x)^k f(\theta_1, x)^{b-k} \right] \tag{A.68}$$

$$= \Delta f^b \sum_{k=0}^{b} \binom{b}{k} \int_0^\infty 2\pi\lambda_b r \exp(-\lambda_b \pi r^2) \times \underbrace{\exp\left(-2\pi\lambda_b \int_r^\infty \left[1 - f(\theta_2, x)^k f(\theta_1, x)^{b-k} \right] x dx \right)}_{g(\theta_1, \theta_2)} dr \tag{A.69}$$

where (a) comes from the proposition A.1 and the fact that

$$\frac{\prod_{x \in \Phi \setminus x_0} f(\theta_2, x)}{\prod_{x \in \Phi \setminus x_0} f(\theta_1, x)} = \prod_{x \in \Phi \setminus x_0} \frac{f(\theta_2, x)}{f(\theta_1, x)} = \prod_{x \in \Phi \setminus x_0} \frac{\frac{q}{1 + \theta_2 \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q}{\frac{q}{1 + \theta_1 \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q} < 1 \tag{A.70}$$

Note that (A.67) is to ensure b can be replaced by a complex number, which is a necessary step if using Gil-Pelaez Theorem. By changing variable $\frac{\|x\|}{\|x_0\|} = v^{\frac{1}{2}}$, we can substitute $g(\theta_1, \theta_2)$ as

$$g(\theta_1, \theta_2) = \exp\left(-\pi\lambda_b r^2 \int_1^\infty \left[1 - \left(1 - \frac{q\theta_1}{\theta_1 + v^{\frac{\alpha}{2}}} \right)^k \left(1 - \frac{q\theta_2}{\theta_2 + v^{\frac{\alpha}{2}}} \right)^{b-k} \right] dv \right) \tag{A.71}$$

Thus (A.69) equals to

$$M_b = \Delta f^b \sum_{k=0}^{b} \binom{b}{k} \left[1 + \int_1^\infty \left[1 - \left(1 - \frac{q\theta_1}{\theta_1 + v^{\frac{\alpha}{2}}} \right)^k \left(1 - \frac{q\theta_2}{\theta_2 + v^{\frac{\alpha}{2}}} \right)^{b-k} \right] dv \right]^{-1} \tag{A.72}$$

Note that (A.64) can be translated to (A.72) by changing variable $u = v^{-\frac{1}{2}}$, and the proof is complete.

Proposition A.1. *If $s \in \mathbb{C}$ is an arbitrary complex number and is an arbitrary complex number and $|x| < 1$, then*

$$(1+x)^b = \sum_{k=0}^{\infty} \binom{b}{k} x^k, \forall b \in \mathbb{C}, |x| < 1 \tag{A.73}$$

Whether (A.73) converges depends on the value of the complex numbers b and x. More precisely:

- *If $|x| < 1$, the series converges absolutely for any complex number b;*
- *If $|x| = 1$, the series converges absolutely if and only if either $Re(b) > 0$ or $b = 0$;*
- *If $|x| > 1$, the series diverges, unless b is a non-negative integer (in which case the series is a finite sum).*

For general expression $(x+y)^b, \forall s \in \mathscr{C}$, we have $(x+y)^b = \sum_{k=0}^{\infty} \binom{b}{k} x^k y^{b-k}$, If $|x| < |y|$.

A.12 Proof of Theorem 6.4

The multinomial theorem illustrated that for any positive integer k and any non-negative integer b, the following equation is satisfied

$$(a_1 + a_2 + \cdots + a_k)^b = \sum_{\substack{n_1, n_2, \cdots, n_k \geq 0 \\ n_1 + n_2 + \cdots + n_k = b}} \frac{b!}{n_1! n_2! \cdots n_k!} a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k} \tag{A.74}$$

Given $b \in \mathbb{N}$, the moment generating function of $r_\phi(\tilde{\pi})$ is

$$M_b = \mathbb{E}_\Phi \left[\left(\sum_{m=1}^{M} \Delta f_m \mathbb{P}(\gamma \geq \theta_m | \Phi) \right)^b \right] \tag{A.75}$$

$$= \mathbb{E}_\Phi \left[\left(\sum_{m=1}^{M} \Delta f_m \prod_{x \in \Phi \setminus x_0} \underbrace{\left(\frac{q}{1 + \theta_m \|x_0\|^\alpha \|x\|^{-\alpha}} + 1 - q \right)}_{f(\theta_m, x)} \right)^b \right] \tag{A.76}$$

$$= \mathbb{E}_\Phi \left[\sum_{\substack{n_1, n_2, \cdots, n_M \geq 0 \\ n_1 + n_2 + \cdots + n_M = b}} \frac{b!}{n_1! n_2! \cdots n_M!} \left(\Delta f_1 \prod_{x \in \Phi \setminus x_0} f(\theta_1, x) \right)^{n_1} \cdots \left(\Delta f_M \prod_{x \in \Phi \setminus x_0} f(\theta_M, x) \right)^{n_M} \right] \tag{A.77}$$

$$= \mathbb{E}_\Phi \left[\sum_{\substack{n_1, n_2, \cdots, n_M \geq 0 \\ n_1 + n_2 + \cdots + n_M = b}} \frac{b!}{n_1! n_2! \cdots n_M!} \prod_{m=1}^{M} \Delta f_m^{n_m} \times \prod_{x \in \Phi \setminus x_0} (f(\theta_m, x))^{n_m} \right] \tag{A.78}$$

$$\overset{(a)}{=} \sum_{\substack{n_1, n_2, \cdots, n_M \geq 0 \\ n_1 + n_2 + \cdots + n_M = b}} \frac{b!}{n_1! n_2! \cdots n_M!} \prod_{m=1}^{M} \Delta f_m^{n_m} \left[1 + \int_1^\infty \left[1 - \prod_{m=1}^{M} \left(1 - \frac{q \theta_m}{\theta_m + v^{\frac{\alpha}{2}}} \right)^{n_m} \right] dv \right]^{-1} \tag{A.79}$$

where (a) comes from the PGFL of the PPP, similar and executive steps are given in (A.44).

A.13 Proof of Lemma 6.2

Given the stationary policy π obtained by RL and the PPP realization Φ, the average transmit success probability is

$$r_\Phi(\pi) = \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T} f_s(A(t),S_c(t))\Big|S(0)=s,\Phi\right] \tag{A.80}$$

$$= \sum_{a\in\mathscr{A}}\sum_{S_c\in\mathscr{C}} f_s(A,S_c)\mathbb{P}\left(S_c=s_c\Big|S(0)=s,\Phi\right)\mathbb{P}\left(A=a\Big|S_c=s_c,S(0)=s,\Phi\right) \tag{A.81}$$

$$\overset{(a)}{=} \sum_{a\in\mathscr{A}}\sum_{S_c\in\mathscr{C}} f_s(A,S_c)\mathbb{P}\left(S_c=s_c\Big|\Phi\right)\mathbb{P}\left(A=a\Big|S_c=s_c,\Phi\right) \tag{A.82}$$

$$= \sum_{S_c\in\mathscr{C}}\left[f_s(A=1,S_c)\mathbb{P}\left(s_c\Big|\Phi\right)\mathbb{P}\left(A=1\Big|S_c=s_c,\Phi\right)+f_s(A=0,S_c)\mathbb{P}\left(s_c\Big|\Phi\right)\mathbb{P}\left(A=0\Big|S_c=s_c,\Phi\right)\right] \tag{A.83}$$

$$\overset{(b)}{=} \sum_{S_c\in\mathscr{C}} f_s(A=1,S_c)\mathbb{P}\left(s_c\Big|\Phi\right)\mathbb{P}\left(A=1\Big|S_c=s_c,\Phi\right) \tag{A.84}$$

$$\overset{(c)}{=} \sum_{m=1}^{M-1} f_s(1,s_c^m)\pi\left(1\Big|s_c^m,\Phi\right)\left[\mathbb{P}(\gamma>\theta_m)-\mathbb{P}\left(\gamma>\theta_{m-1}\Big|\Phi\right)\right]+f_s(1,s_c^M)\pi(1\Big|s_c^M,\Phi)\mathbb{P}\left(\gamma>\theta_M\Big|\Phi\right) \tag{A.85}$$

$$\overset{(d)}{=} \sum_{m=1}^{M}\left[f_m\pi(1|s_c^m,\Phi)-f_{m-1}\pi(1|s_c^{m-1},\Phi))\right]\mathbb{P}\left(\gamma>\theta_m\Big|\Phi\right) \tag{A.86}$$

$$\overset{e}{=} \sum_{m=1}^{M}\left[f_m\pi(1|s_c^m,\Phi)-f_{m-1}\pi(1|s_c^{m-1},\Phi))\right]\prod_{x\in\Phi\setminus x_0}\left(\frac{q}{1+\theta_m\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right) \tag{A.87}$$

where (a) comes from the $r_\Phi(\pi)$ only depends on current channel state s_c without depending on initialization state $S(0)$ under stationary policy π; (b) comes from the fact $f_s(A=0,S_c)=0,\forall S_c\in\mathscr{C}$; (c) from the definition $\pi(a|s):\mathbb{P}(A=a|S=s)$; (d) we set $f_0=0,f_{M+1}=1$ for convenience; (e) follows the similar to the steps in (A.2) in Appendix A.1.

A.14 Proof of Theorem 6.5

The bth moments of $M_b\triangleq\mathbb{E}[(r_\Phi(\pi))^b]$ has the expression

$$M_b = \mathbb{E}_\Phi\left[\left(\sum_{m=1}^{M}\left[f_m\pi(1|s_c^m,\Phi)-f_{m-1}\pi(1|s_c^{m-1},\Phi)\right]\mathbb{P}(\gamma\geq\theta_m|\Phi)\right)^b\right],\ b\in\mathbb{N} \tag{A.88}$$

$$= \mathbb{E}_\Phi\left[\left(\sum_{m=1}^{M} b_m\prod_{x\in\Phi\setminus x_0}\underbrace{\left(\frac{q}{1+\theta_m\|x_0\|^\alpha\|x\|^{-\alpha}}+1-q\right)}_{f(\theta_m,x)}\right)^b\right] \tag{A.89}$$

where $b_m:=f_m\pi(1|s_c^m,\Phi)-f_{m-1}\pi(1|s_c^{m-1},\Phi)$.

Based on multinomial series definition, (A.89) can be expressed as

$$M_b = \mathbb{E}_\Phi \left[\sum_{\substack{n_1,n_2,\cdots,n_M \geq 0 \\ n_1+n_2+\cdots+n_M=b}} \frac{b!}{n_1!n_2!\cdots n_M!} \left(b_1 \prod_{x\in\Phi\setminus x_0} f(\theta_1,x) \right)^{n_1} \cdots \left(b_M \prod_{x\in\Phi\setminus x_0} f(\theta_m,x) \right)^{n_M} \right] \tag{A.90}$$

$$= \sum_{\substack{n_1,n_2,\cdots,n_M \geq 0 \\ n_1+n_2+\cdots+n_M=b}} \frac{b!}{n_1!n_2!\cdots n_M!} \mathbb{E}_\Phi \left[\left(b_1 \prod_{x\in\Phi\setminus x_0} f(\theta_1,x) \right)^{n_1} \cdots \left(b_M^\Phi \prod_{x\in\Phi\setminus x_0} f(\theta_M,x) \right)^{n_M} \right] \tag{A.91}$$

$$= \sum_{\substack{n_1,n_2,\cdots,n_M \geq 0 \\ n_1+n_2+\cdots+n_M=b}} \frac{b!}{n_1!n_2!\cdots n_M!} \mathbb{E}_\Phi \left[\prod_{x\in\Phi\setminus x_0} \prod_{m=1}^{M} \left(b_m f(\theta_m,x) \right)^{n_m} \right] \tag{A.92}$$

where

$$\mathbb{E}_\Phi \left[\prod_{x\in\Phi\setminus x_0} \prod_{m=1}^{M} \left(b_m f(\theta_m,x) \right)^{n_m} \right] \tag{A.93}$$

$$\overset{(a)}{=} \int_0^\infty 2\pi\lambda_b \exp(-\lambda_b \pi r^2) \exp\left(-2\pi\lambda_b \int_r^\infty \left[1 - \prod_{m=1}^{M} b_m^{n_m} \left(\frac{q}{1+\theta_m x_0^\alpha x^{-\alpha}} + 1 - q \right)^{n_m} \right] x dx \right) dr \tag{A.94}$$

$$\overset{(b)}{=} \left[1 + \int_1^\infty \left[1 - \prod_{m=1}^{M} \mathbb{E}_\Phi[b_m^{n_m}] \left(1 - \frac{q\theta_m}{\theta_m + v^{\frac{\alpha}{2}}} \right)^{n_m} \right] dv \right]^{-1} \tag{A.95}$$

where (a) follows the distribution of $\|x_0\|$ as $f_{\|x_0\|}(r) = e^{-\pi\lambda r^2} 2\pi\lambda r$ and PGFL of PPP.

Further, $\mathbb{E}_\Phi[b_m^{n_m}]$ represents the n_mth moments of b_m, where $n_m \in \mathbb{N}^+$ and satisfying $n_m \leq M$, $b_m = f_m \pi(1|s_c^m,\Phi) - f_{m-1}\pi(1|s_c^{m-1},\Phi)$.

$$\mathbb{E}_\Phi[b_m^{n_m}] = \mathbb{E}_\Phi \left[\left[f_m \pi(1|s_c^m,\Phi) - f_{m-1}\pi(1|s_c^{m-1},\Phi) \right]^{n_m} \right] \tag{A.96}$$

$$= \mathbb{E}_\Phi \left[\sum_{i=1}^{n_m} \left(f_m \pi(1|s_c^m,\Phi) \right)^i \left(-f_{m-1}\pi(1|s_c^{m-1},\Phi) \right)^{n_m-i} \right] \tag{A.97}$$

$$= \sum_{i=0}^{n_m} f_m^i (-1)^{n_m-i} f_{m-1}^{n_m-i} \mathbb{E}_\Phi \left[\pi(1|s_c^m,\Phi)^i \right] \mathbb{E}_\Phi \left[\pi(1|s_c^{m-1},\Phi)^{n_m-i} \right] \tag{A.98}$$

where $\mathbb{E}_\Phi \left[\pi(1|s_c^m,\Phi)^i \right]$ is the ith moment of policy averaged over Φ, and $\mathbb{E}_\Phi \left[\pi(1|s_c^{m-1},\Phi)^{n_m-i} \right]$ can be further interpreted as

$$\mathbb{E}_\Phi \left[\pi(1|s_c^{m-1},\Phi)^{n_m-i} \right] = \int_{\mathcal{N}} \pi(1|s_c^m,\Phi)^{n_m-i} \mathbb{P}(d\varphi) \tag{A.99}$$

where $\varphi = \{x_1, x_2, \cdots\}$ is viewed as a locally finite countable subset of $\mathbb{R}^2$, $\mathcal{N}$ is the set of all φ, and a point process Φ is a random choice of one of the φ in $\mathcal{N}$. $\mathbb{P}(d\varphi)$ satisfied $\int_{\mathcal{N}} \mathbb{P}(d\varphi) = 1$, which can be seen as the probability of number of points in little displace of φ.

Substituting (A.95), (A.98) to (A.92), we completed the proof.

Particularly, when $n_m = 1$, we have

$$\mathbb{E}_\Phi[b_m] = \mathbb{E}_\Phi\left[f_m \pi(1|s_c^m, \Phi) - f_{m-1}\pi(1|s_c^{m-1}, \Phi)\right] \tag{A.100}$$

$$= f_m \mathbb{E}_\Phi\left[\pi(1|s_c^m, \Phi)\right] - f_{m-1}\mathbb{E}_\Phi\left[\pi(1|s_c^{m-1}, \Phi)\right] \tag{A.101}$$

where $\mathbb{E}_\Phi\left[\pi(1|s_c^m, \Phi)\right]$ can be interpreted as the proportion of policy choose active at a predefined state average over Φ.

Bibliography

[1] R. S. Sutton, A. G. Barto *et al.*, *Introduction to reinforcement learning*. MIT press Cambridge, 1998, vol. 135.

[2] V. P. Rekkas, S. Sotiroudis, P. Sarigiannidis, S. Wan, G. K. Karagiannidis, and S. K. Goudos, "Machine learning in beyond 5G/6G networks—state-of-the-art and future trends," *Electronics*, vol. 10, no. 22, pp. 2786–2804, 2021.

[3] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[4] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, 2011.

[5] M. Haenggi, "The Meta Distribution of the SIR in Poisson Bipolar and Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2577–2589, 2016.

[6] K. Stamatiou and M. Haenggi, "Random-access poisson networks: Stability and delay," *IEEE Communications Letters*, vol. 14, no. 11, pp. 1035–1037, 2010.

[7] B. Blaszczyszyn, M. Jovanovic, and M. K. Karray, "Performance laws of large heterogeneous cellular networks," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2015, pp. 597–604.

[8] A. AlAmmouri, J. G. Andrews, and F. Baccelli, "Stability of wireless random access systems," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 1190–1197.

[9] Y. Zhong, T. Q. S. Quek, and X. Ge, "Heterogeneous cellular networks with spatio-temporal traffic: Delay analysis and scheduling," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1373–1386, June 2017.

[10] Y. Zhong, W. Zhang, and M. Haenggi, "Delay analysis in static poisson network," in *2015 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2015, pp. 1–5.

[11] H. S. Dhillon, R. K. Ganti, J. G. Andrews, and F. Baccelli, "Load-aware modeling and analysis of heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 550–560, 2013.

[12] M. Gharbieh, H. Elsawy, A. Bader, and M. S. Alouini, "Spatiotemporal Stochastic Modeling of IoT Enabled Cellular Networks: Scalability and Stability Analysis," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3585–3600, 2017.

[13] N. Jiang, Y. Deng, X. Kang, and A. Nallanathan, "Random access analysis for massive IoT networks under a new spatio-temporal model: A stochastic geometry approach," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5788–5803, Nov 2018.

[14] M. H. Ngo and V. Krishnamurthy, "Monotonicity of constrained optimal transmission policies in correlated fading channels with ARQ," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 438–451, 2009.

[15] N. Mastronarde and M. van der Schaar, "Joint physical-layer and system-level power management for delay-sensitive wireless communications," *IEEE Transactions on Mobile Computing*, vol. 12, no. 4, pp. 694–709, 2012.

[16] N. Zhao, Y.-C. Liang, D. Niyato, Y. Pei, M. Wu, and Y. Jiang, "Deep reinforcement learning for user association and resource allocation in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5141–5152, 2019.

[17] E. Ghadimi, F. D. Calabrese, G. Peters, and P. Soldati, "A reinforcement learning approach to power control and rate adaptation in cellular networks," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–7.

[18] L. Qi, M. Peng, Y. Liu, and S. Yan, "Advanced user association in non-orthogonal multiple access-based fog radio access networks," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8408–8421, 2019.

[19] K. Zia, N. Javed, M. N. Sial, S. Ahmed, A. A. Pirzada, and F. Pervez, "A distributed multi-agent rl-based autonomous spectrum allocation scheme in d2d enabled multi-tier hetnets," *IEEE Access*, vol. 7, pp. 6733–6745, 2019.

[20] T. Sanguanpuak, S. Guruacharya, N. Rajatheva, M. Bennis, and M. Latva-Aho, "Multi-operator spectrum sharing for small cell networks: A matching game perspective," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3761–3774, 2017.

[21] Y. J. Chun, S. L. Cotton, H. S. Dhillon, A. Ghrayeb, and M. O. Hasna, "A stochastic geometric analysis of device-to-device communications operating over generalized fading channels," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4151–4165, 2017.

[22] M. Di Renzo, A. Guidotti, and G. E. Corazza, "Average rate of downlink heterogeneous cellular networks over generalized fading channels: A stochastic geometry approach," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 3050–3071, 2013.

[23] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 248–257, 2012.

[24] Y. Deng, L. Wang, M. Elkashlan, M. Di Renzo, and J. Yuan, "Modeling and analysis of wireless power transfer in heterogeneous cellular networks," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5290–5303, Dec 2016.

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

[26] H. ElSawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4454–4469, Aug 2014.

[27] H. ElSawy, A. Sultan-Salem, M. S. Alouini, and M. Z. Win, "Modeling and Analysis of Cellular Networks Using Stochastic Geometry: A Tutorial," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 1, pp. 167–203, 2017.

[28] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Load-aware modeling and analysis of heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1666–1677, 2013.

[29] Y. Zhong, M. Haenggi, T. Q. Quek, and W. Zhang, "On the stability of static poisson networks under random access," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 2985–2998, 2016.

[30] R. R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 918–930, 1988.

[31] Q. Liu, J. Baudais, and P. Mary, "A tractable coverage analysis in dynamic downlink cellular networks," in *IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020.

[32] ——, "Queue analysis with finite buffer by stochastic geometry in downlink cellular networks," in *IEEE 93rd Vehicular Technology Conference(VTC)*, 2021.

[33] N. Sapountzis, T. Spyropoulos, N. Nikaein, and U. Salim, "An analytical framework for optimal downlink-uplink user association in hetnets with traffic differentiation," in *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–7.

[34] H. H. Yang and T. Q. S. Quek, "Spatio-temporal analysis for SINR coverage in small cell networks," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5520–5531, 2019.

[35] A. Attahiru S, *Applied Discrete-Time Queues*. New York,NY,USA:Springer, 2016.

[36] W. Ni, J. A. Zhang, Z. Fang, M. Abolhasan, R. P. Liu, and Y. J. Guo, "Analysis of finite buffer in two-way relay: A queueing theoretic point of view," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3690–3694, 2018.

[37] E. F. Morales and J. H. Zaragoza, "An introduction to reinforcement learning," *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*, pp. 63–80, 2011.

[38] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. C. Liang, and D. I. Kim, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[39] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.

[40] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[41] B. Błaszczyszyn, M. Haenggi, P. Keeler, and S. Mukherjee, *Stochastic geometry analysis of cellular networks*. Cambridge University Press, 2018.

[42] M. Haenggi, "A geometric interpretation of fading in wireless networks: Theory and applications," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5500–5510, 2008.

[43] R. K. Ganti and M. Haenggi, "Asymptotics and approximation of the SIR distribution in general cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2130–2143, 2015.

[44] D. G. Kendall, "Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain," *The Annals of Mathematical Statistics*, pp. 338–354, 1953.

[45] P. Purdue, "The M/M/1 queue in a Markovian environment," *Operations Research*, vol. 22, no. 3, pp. 562–569, 1974.

[46] S. R. Chakravarthy *et al.*, "The batch Markovian arrival process: a review and future work," *Advances in probability theory and stochastic processes*, vol. 1, pp. 21–49, 2001.

[47] R. A. Vitale, "On stochastic dependence and a class of degenerate distributions," *Lecture Notes-Monograph Series*, pp. 459–469, 1990.

[48] T. Yang and H. Li, "On the steady-state queue size distribution of the discrete-time geo/g/1 queue with repeated customers," *Queueing systems*, vol. 21, no. 1, pp. 199–215, 1995.

[49] A. Gravey, J.-R. Louvion, and P. Boyer, "On the geo/d/1 and geo/d/1/n queues," *Performance Evaluation*, vol. 11, no. 2, pp. 117–125, 1990.

[50] T. Altiok, "On the phase-type approximations of general distributions," *IIE Transactions*, vol. 17, no. 2, pp. 110–116, 1985.

[51] B. Balcioglų, D. L. Jagerman, and T. Altiok, "Approximate mean waiting time in a GI/D/1 queue with autocorrelated times to failures," *IIE Transactions*, vol. 39, no. 10, pp. 985–996, 2007.

[52] O. Brun and J.-M. Garcia, "Analytical solution of finite capacity m/d/1 queues," *Journal of Applied Probability*, vol. 37, no. 4, pp. 1092–1098, 2000.

[53] D. L. Iglehart, "Extreme values in the gi/g/1 queue," *The Annals of Mathematical Statistics*, pp. 627–635, 1972.

[54] A. Chydzinski, R. Wojcicki, and G. Hryn, "On the number of losses in an MMPP queue," in *International Conference on Next Generation Wired/Wireless Networking*. Springer, 2007, pp. 38–48.

[55] Q. M. He, "The classification of matrix GI/M/1 type Markov chains with a tree structure and its applications to queueing," *Journal of Applied Probability*, vol. 40, no. 4, pp. 1087–1102, 2003.

[56] P. Mary, V. Koivunen, and C. Moy, "Reinforcement learning for PHY layer communications," arXiv preprint, Jul. 2021. [Online]. Available: https://arxiv.org/abs/2106.11595

[57] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[58] H. Van Seijen, H. Van Hasselt, S. Whiteson, and M. Wiering, "A theoretical and empirical analysis of expected Sarsa," in *2009 IEEE symposium on adaptive dynamic programming and reinforcement learning*. IEEE, 2009, pp. 177–184.

[59] F. G. Nocetti, I. Stojmenovic, and J. Zhang, "Addressing and routing in hexagonal networks with applications for tracking mobile users and connection rerouting in cellular networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp. 963–971, 2002.

[60] K. B. Baltzis, "Analytical and closed-form expressions for the distribution of path loss in hexagonal cellular networks," *Wireless Personal Communications*, vol. 60, no. 4, pp. 599–610, 2011.

[61] S. Lee and K. Huang, "Coverage and economy of cellular networks with many base stations," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1038–1040, 2012.

[62] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Stochastic geometry and architecture of communication networks," *Telecommunication Systems*, vol. 7, no. 1, pp. 209–227, 1997.

[63] B. Błaszczyszyn, M. K. Karray, and H. P. Keeler, "Using poisson processes to model lattice cellular networks," in *2013 Proceedings IEEE INFOCOM*. IEEE, 2013, pp. 773–781.

[64] A. Guo and M. Haenggi, "Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5800–5812, 2013.

[65] N. Deng, W. Zhou, and M. Haenggi, "Heterogeneous cellular network models with dependence," *IEEE Journal on selected Areas in Communications*, vol. 33, no. 10, pp. 2167–2181, 2015.

[66] M. Haenggi, "Meta distributions-part 1: Definition and examples," *IEEE Communications Letters*, vol. 25, no. 7, pp. 2089–2093, 2021.

[67] M. Afshang and H. S. Dhillon, "Fundamentals of modeling finite wireless networks using binomial point process," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3355–3370, 2017.

[68] M. Haenggi, "Mean interference in hard-core wireless networks," *IEEE Communications Letters*, vol. 15, no. 8, pp. 792–794, 2011.

[69] Y. Li, F. Baccelli, H. S. Dhillon, and J. G. Andrews, "Statistical modeling and probabilistic analysis of cellular networks with determinantal point processes," *IEEE Transactions on communications*, vol. 63, no. 9, pp. 3405–3422, 2015.

[70] S. Wang and M. Di Renzo, "On the mean interference-to-signal ratio in spatially correlated cellular networks," *IEEE Wireless Communications Letters*, vol. 9, no. 3, pp. 358–362, 2019.

[71] S. M. Azimi-Abarghouyi, B. Makki, M. Haenggi, M. Nasiri-Kenari, and T. Svensson, "Stochastic geometry modeling and analysis of single-and multi-cluster wireless networks," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4981–4996, 2018.

[72] A. Guo, Y. Zhong, W. Zhang, and M. Haenggi, "The gauss–poisson process for wireless networks and the benefits of cooperation," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1916–1929, 2016.

[73] A. M. Hayajneh, S. A. R. Zaidi, D. C. McLernon, M. Di Renzo, and M. Ghogho, "Performance analysis of uav enabled disaster recovery networks: A stochastic geometric framework based on cluster processes," *IEEE Access*, vol. 6, pp. 26 215–26 230, 2018.

[74] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint transmission in heterogeneous networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4134–4146, 2014.

[75] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Modeling non-uniform UE distributions in downlink cellular networks," *IEEE Wireless Communications Letters*, vol. 2, no. 3, pp. 339–342, 2013.

[76] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, 2014.

[77] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2365–2380, 2015.

[78] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 550–560, 2012.

[79] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," arXiv preprint, Oct. 2016. [Online]. Available: https://arxiv.org/abs/1604.03183

[80] H. Elsawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 4454–4469, 2014.

[81] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2669–2679, 2012.

[82] M. Di Renzo and P. Guan, "Stochastic geometry modeling and system-level analysis of uplink heterogeneous cellular networks with multi-antenna base stations," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2453–2476, 2016.

[83] A. H. Sakr and E. Hossain, "Cognitive and energy harvesting-based d2d communication in cellular networks: Stochastic geometry modeling and analysis," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1867–1880, 2015.

[84] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in hetnets," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5360–5373, 2015.

[85] L. Zhang, W. Nie, G. Feng, F.-C. Zheng, and S. Qin, "Uplink performance improvement by decoupling uplink/downlink access in hetnets," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 6862–6876, 2017.

[86] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, 2015.

[87] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5038–5057, 2015.

[88] A. Abdallah, M. M. Mansour, and A. Chehab, "A distance-based power control scheme for D2D communications using stochastic geometry," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2017, pp. 1–6.

[89] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of los and nlos transmissions in dense cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2365–2380, 2016.

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

[91] Y. Hmamouche, M. Benjillali, S. Saoudi, H. Yanikomeroglu, and M. Di Renzo, "New trends in stochastic geometry for wireless networks: A tutorial and survey," *Proceedings of the IEEE*, 2021.

[92] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[93] E. Leonardi, M. Mellia, F. Neri, and M. Ajmone Marsan, "Bounds on average delays and queue size averages and variances in input-queued cell-based switches," in *2001 Proceedings IEEE INFOCOM*, vol. 2, 2001, pp. 1095–1103 vol.2.

[94] H. Al-Zubaidy, J. Liebeherr, and A. Burchard, "Network-Layer Performance Analysis of Multihop Fading Channels," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 204–217, 2016.

[95] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *IEEE/ACM transactions on networking*, vol. 15, no. 6, pp. 1333–1344, 2007.

[96] M. Haenggi, "The mean interference-to-signal ratio and its key role in cellular and amorphous networks," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 597–600, 2014.

[97] M. Haenggi and R. Smarandache, "Diversity polynomials for the analysis of temporal correlations in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5940–5951, 2013.

[98] G. Nigam, P. Minero, and M. Haenggi, "Spatiotemporal cooperation in heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1253–1265, 2015.

[99] M. A. Kishk and H. S. Dhillon, "Joint uplink and downlink coverage analysis of cellular-based RF-powered IoT network," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 446–459, 2017.

[100] K. Koufos and C. P. Dettmann, "Temporal correlation of interference and outage in mobile networks over one-dimensional finite regions," *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 475–487, 2018.

[101] F. Baccelli and P. Brémaud, *Elements of queueing theory: Palm Martingale calculus and stochastic recurrences*. Springer Science & Business Media, 2013, vol. 26.

[102] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[103] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[104] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 732–742, 2008.

[105] H. A. Al-Rawi, M. A. Ng, and K.-L. A. Yau, "Application of reinforcement learning to routing in distributed wireless networks: a review," *Artificial Intelligence Review*, vol. 43, no. 3, pp. 381–416, 2015.

[106] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 2, pp. 257–265, 2018.

[107] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in hetnets: A utility perspective," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1025–1039, 2015.

[108] S. O. Somuyiwa, A. György, and D. Gündüz, "A reinforcement-learning approach to proactive caching in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1331–1344, 2018.

[109] R. G. Bartle, *The elements of integration and Lebesgue measure*. John Wiley & Sons, 2014.

[110] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[111] E. Altman, *Constrained Markov decision processes: stochastic modeling*. Routledge, 1999.

[112] K. I. Ahmed and E. Hossain, "A deep q-learning method for downlink power allocation in multi-cell networks," arXiv preprint, Apr. 2019. [Online]. Available: https://arxiv.org/abs/1904.13032

[113] S. Boyd and A. Mutapcic, "Stochastic subgradient methods," *Lecture Notes for EE364b, Stanford University*, 2008.

[114] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC press, 2004.

[115] C. Man, L. Hang, L. Xuewen, and C. Shuguang, "Reinforcement learning-based multiaccess control and battery prediction with energy harvesting in IoT systems," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2009–2020, 2019.

[116] A. Shadrin, "Twelve proofs of the markov inequality," *Approximation theory: a volume dedicated to Borislav Bojanov*, pp. 233–298, 2004.

[117] H. H. Sohrab, *Basic Real Analysis*, 2nd ed. Boston, Basel, Berlin: Springer, 2005.

Titre : **Analyse des performances de la liaison descendante des réseaux cellulaires dynamiques**

Mots clés : géométrie stochastique, théorie des files d'attente, apprentissage par renforcement, probabilité de couverture, région de stabilité

Résumé : Cette thèse caractérise de la région de stabilité d'un réseau aléatoire lorsqu'un modèle de trafic est intégré à la description de la géométrie du réseau. Premièrement, nous caractérisons la probabilité de couverture stable du réseau. À partir de la notion de probabilité de couverture dynamique, l'interaction entre les états des files d'attentes dans le réseau est prise en compte à l'aide d'une modélisation par chaîne de Markov discrète des files d'attente. Les cas des files d'attente à taille finie et infinie sont traités. La région de stabilité indique à partir de quelle intensité de trafic au moins une file d'attente dans le réseau diverge. Une description plus fine du phénomène est faite en répondant à la question “quelle est la proportion de files d'attente instables dans le réseau ?". Dans ce cas, la notion de epsilon-stabilité est exploitée, elle décrit l'ensemble des intensités de trafic pour lesquelles une file d'attente prise au hasard a une probabilité de diverger inférieure à epsilon. Enfin, la dépendance entre la géométrie, la dynamique du réseau et la stratégie d'allocation rend la caractérisation des regions de stabilité avec l'allocation de ressources très difficile. Le caractère dynamique du réseau est descrit par un processus décisionnel markovien utilisant l'apprentissage par renforcement. La région de stabilité est donc étudiée lorsque la station de base typique peut choisir d'émettre ou de rester silencieuse selon l'état du réseau observé.

Title : **Performance Analysis of Dynamic Downlink Cellular Networks**

Keywords : stochastic geometry, queuing theory, reinforcement learning, coverage probability, stability region, transmission policies

Abstract : The main question posed in this thesis is the characterization of the stability region of a random network when a traffic model is integrated into the description of the network geometry. First, we characterized the stable coverage probability of a random network. Starting from the notion of dynamic coverage probability, the interaction between the queue states in the network is taken into account using a discrete Markov chain modeling of the queues, where the typical user's service rate depends on the dynamic coverage probability. The cases of buffer with finite and infinite size are both taken into account. The stability region indicates from which traffic intensity at least one queue in the network diverges. A more refined description of the phenomenon is made by answering the question, "what is the proportion of unstable queues in the network?". In this case, the notion of epsilon-stability is exploited, which describes the set of traffic intensities for which a queue taken at random has a probability of diverging less than epsilon. Finally, the characterization of the stable regions by considering the resource allocation is very difficult to obtain, because of the dependence between the geometry and the dynamic of the network and the allocation strategy. However, the dynamic nature of the network considered in this thesis lends itself perfectly to description by a Markovian decision process for which reinforcement learning strategies can be proposed. The region of stability is therefore investigated where the typical base station can choose to transmit or remain silent depending on the observed network state.