



HAL
open science

Modèles stochastiques pour les réseaux ad hoc mobiles

Robin Groenevelt

► **To cite this version:**

Robin Groenevelt. Modèles stochastiques pour les réseaux ad hoc mobiles. Réseaux et télécommunications [cs.NI]. Université Nice Sophia Antipolis, 2005. Français. NNT: . tel-00274901

HAL Id: tel-00274901

<https://theses.hal.science/tel-00274901>

Submitted on 21 Apr 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Nice - Sophia Antipolis – UFR Sciences
École Doctorale STIC

THÈSE

Présentée pour obtenir le titre de :
Docteur en Sciences de l'Université de Nice - Sophia Antipolis

Spécialité : INFORMATIQUE

par

Robin GROENEVELT

Équipe d'accueil : Maestro – INRIA Sophia Antipolis

Modèles Stochastiques pour les Réseaux Ad Hoc Mobiles

Soutenue publiquement à l'INRIA le 7 avril 2005 à 14:30 devant le jury composé de :

Président :	Igor	LITOVSKY	Polytech Université de Nice-Sophia Antipolis
Directeur :	Philippe	NAIN	INRIA
Rapporteurs :	Michel	MANDJES	CWI, Université d'Amsterdam
	Ravi	MAZUMDAR	Waterloo University
Examineurs :	Jean-Claude	BERMOND	CNRS
	Ger	KOOLE	Université Libre, Amsterdam

Modèles Stochastiques pour les Réseaux
Ad Hoc Mobiles

Robin GROENEVELT

Titre de la thèse en anglais :

Stochastic Models for Mobile Ad Hoc Networks

*Cette thèse est dédiée
à Magda*

Remerciements

A thesis takes a long time to write. A lot of hard work is required. It takes up such a big part of one's life that afterwards it is difficult to accept that it is actually finished. This work would not have been possible without the help of the many people whom I would like to thank here.

Starting from the beginning, I would especially like to thank Ger Koole for having influenced my life—in many ways, and all of them positive—by presenting me with the opportunity to do a PhD. Because of my decision to move away from the Netherlands it meant that I could unfortunately not do a PhD under his supervision. Nonetheless, I am very happy to have had the occasions to collaborate with him and I hope to be able to continue doing this in the future.

During my initial phase of hesitation before starting a PhD, Bert Kersten—whose career advice I have come to trust blindly over the years—is the person who convinced me to pick up the offer at INRIA. Once again, his advice was correct and I will forever be grateful for the influence he has had in my life over the years.

To both Ger and Bert I would like to express my additional gratitude for the faith that they have had in me.

Although a PhD knows many moments of happiness and sufferings, with no doubt Magdalena Oroń is the person who paid the price for my decision to do a PhD. It is for this reason that I dedicate this thesis not to my mother, but to her. Kashoe, jak smiez?

There are other people as well who have had to pay a price for my decision, and these are of course the friends that I “left behind” in the Netherlands. Luckily though, for some strange reason these friends never seemed to mind too much paying me a visit on the sunny Côte-d'Azur.

During my thesis I have received the help and support of many fellow students of the Mistral and Maestro projects (and sometimes Planet and Mascotte as well). I will not attempt to name all of these people as it would be difficult for me to express the degree of gratitude that I have for each person. Nonetheless, I would like to specifically thank Nidhi, Tuna, Daniele, Urtzi, and Arzad for being such special people and great friends. I feel lucky to have met you and I regret no longer being able to spend (as) much time with you. My gratitude also goes out to Urtzi, Daniele, Uri Yechiali, and Marwan Krunz for their theoretical support.

Thanks go to INRIA and Ephie for having provided me with the necessary resources, as well as the 13 office mates from 10 different countries for having made my stay at INRIA such an interesting experience. Giannis, thank you very much for the help during the last part of my thesis. Tuna, I hope to meet you again in the near future and it is a shame we do not live in the same country.

The following people have been invaluable in helping me with my French: Eric, Florence, and Rabea. I am glad to know that they at least had some fun correcting my French.

Naturally, none of this would have been possible without my supervisor, Philippe Nain. I would like to thank him not only for his time and help in making my thesis a reality, but also for sharing his views on many issues not directly related to the thesis. I have obtained a great respect for his knowledge. Aside from the working environment, I would also like to thank him for all of the hours of running together and for motivating me to run more often.

Eitan Altman, the inexhaustible and indepletable source of ideas and models, thank you VERY much for having shown up at such a crucial time and helping me through it. Your presence was definitely the turning point in my thesis. Also, thank you for showing me how much “fun” research can be.

At the final end of my thesis timeline are the members of the jury to whom I would like to express my gratitude. A big thank you goes to Michel Mandjes and Ravi Mazumdar for having made the journey to INRIA and for taking on the not so small task of being “rapporteurs”. Thanks also go out to Ger Koole, Jean-Claude Bermond, and Igor Litovsky for agreeing to become the other members of the jury.

People that were indirectly involved have been at least equally important as they provide me with the fuel for living life to the maximum. Without their “distractions”, I would never have been able to enjoy life as much or to complete my thesis.

My appreciation will always go out to Sandjai Bhulai for being who he is and for leading a life which runs parallel to my own. The reassuring company (even if far away) means that we know we will never be travelling the road alone. Sandjai Bhulai, thank you for your endless enthusiasm and for proving that oracles do exist at times (just ask him any question).

Milestones in people’s lives are not the same if they are not shared with the people that you have loved all of your life. I wished my mother could be around experience the wonderful phases of her son “growing up”. My father has motivated me unintentionally as he would have preferred seeing me starting an industrial career instead of starting a PhD. However, by following my own desires and accomplishing it on my own it has led to a more gratifying experience. I have learned that following your own heart is what counts and I feel that I have grown as a person because of this. It is therefore now an even greater pleasure to have a father that is proud of me.

Last and definitely not least, I would like to express my thanks and happiness to Valda for being such a fantastic person and beautiful partner in life.

Enjoy the thesis,

A handwritten signature in blue ink that reads "Robin". The signature is written in a cursive style with a large, looping initial 'R' and a long, sweeping underline.

Contents

<i>Présentation des travaux de thèse</i>	i
Un Réseau Ad Hoc Mobile	i
Réseaux Ad Hoc Mobiles	ii
Motivation et la Contribution de la Thèse	iv
Notation	viii
Introduction	1
What is an Ad Hoc Network?	1
Mobile Ad Hoc Networks	2
Motivation and Contribution of the Thesis	3
Organisation of the thesis	5
Notation	7
Part I Message Delay in <i>Mobile</i> Ad Hoc Networks	11
1 Message Delay for One-dimensional Brownian Motions	11
1.1 Introduction	12
1.2 Two mobiles moving along a line segment	14
1.3 A chain of relaying mobiles	18
1.4 Numerical results and discussion	21
1.5 Extensions to the model	23
1.6 Concluding remarks	24
1.A The Brownian motion	25
1.A.1 History of the Brownian motion	25
1.A.2 Brownian motion as a random walker	26
1.A.3 Brownian motion as a diffusion process	27
1.B Proof of Proposition 1.2.2	27
1.C Proof of Proposition 1.3.1	29
2 Message Delay for One-dimensional Random Walkers	31
2.1 Two random walkers jumping on a finite state space	32
2.2 A chain of relaying nodes	38

2.2.1	Numerical results for a chain of nodes	42
2.3	Extensions to the model	43
2.4	Concluding remarks	44
2.A	The stationary distribution of a random walker	45
2.B	Proof of Proposition 2.1.1	45
2.C	Proof of Proposition 2.1.2	49
2.D	Proof of Proposition 2.2.1	60
3	Message Delay in Two-dimensional Networks	63
3.1	Introduction	64
3.2	The stochastic model	66
3.2.1	Definition of the model	66
3.2.2	Performance of relay protocols	67
3.3	Applications	75
3.3.1	Mobility models	75
3.3.2	Simulation Setting	76
3.3.3	Validation of the Poisson meeting times	77
3.3.4	Expected Message Delay	81
3.3.5	Distribution of Number of Copies	81
3.4	Large networks	84
3.5	Concluding Remarks	85
3.A	Proof of Lemma 3.2.1	86
3.B	Derivation of the time until the next event	88
4	λ: The Mysterious Parameter	91
4.1	Introduction	92
4.2	The values for λ	93
4.3	Examples	98
4.4	Implications for the message delay	98
4.5	Concluding remarks	100
4.A	Proof of Theorem 4.2.1.	101
4.B	Proof of Proposition 4.2.1.	103
4.C	Proof of Proposition 4.2.2.	104
4.D	Probability of contact at a random moment	105
Part II Polling Systems with Correlated Switchover Times		109
5	An Alternating-Priority Server with Correlated Switchover Times	109
5.1	Introduction and Motivation	110
5.2	Model Description	111
5.3	Exhaustive/Exhaustive Service System	114
5.3.1	Introduction	114
5.3.2	Performance Measures	116
5.3.3	Identical queues	122

5.3.4	Heavily Unbalanced Traffic	124
5.3.5	Switchover Times Equal to Zero	124
5.4	Exhaustive/Gated Service System	124
5.4.1	Introduction	124
5.4.2	Performance Measures	125
5.4.3	Identical Queues	131
5.4.4	Comparison of Waiting Times in Exhaustive/Gated System	132
5.5	Examples	135
5.5.1	Example 1: Correlated Switchover Times	135
5.5.2	Example 2: Stochastic Recursive Sequence of Switchover Times	137
5.5.3	Example 3: Identical Switchover Times	139
5.5.4	Example 4: Switchover Times Coming from the Same Sequence	142
5.6	Concluding Remarks	143
5.A	Proof of Theorem 5.3.2.	145
5.B	Proof of Theorem 5.3.3.	147
5.C	Alternative proof of Theorem 5.3.4.	150
5.D	Proof of Theorem 5.4.3.	150
5.E	List of Notation	157

Appendix **161**

A The Value Function of a Tandem Queue **161**

A.1	Introduction	162
A.2	A tandem queue with no arrivals	162
A.2.1	The generating function of the value function	163
A.2.2	The value function of a tandem queue	166
A.3	A tandem queue with arrivals	168
A.4	Appendix: Proof of Proposition A.2.1	170
A.5	Appendix: Proof of Proposition A.2.2	172

Bibliography **181**

Résumé – Abstract **184**

Présentation des travaux de thèse

Contents

Un Réseau Ad Hoc Mobile	i
Réseaux Ad Hoc Mobiles	ii
Motivation et la Contribution de la Thèse	iv
Notation	viii

La réhabilitation d'un désastre, l'abaissement des détecteurs d'un avion en vol, le traitement de la progression des bagages aux aéroports, l'arrivée de soldats dans un territoire étranger, le contrôle d'un environnement et des habitats, l'équipement d'une maison en électroménager, la circulation des véhicules sur la route, et l'étiquetage des animaux. Tous ces événements ont un point commun : ils profitent amplement du déploiement des réseaux ad hoc.

Cette thèse a pour objet l'étude de la mobilité sur la performance des réseaux ad hoc mobiles. Nous nous concentrons en particulier sur l'étude du délai d'un message.

D'abord, nous définissons le type des réseaux auxquels nous nous intéressons, puis nous regardons la mobilité dans ces réseaux.

Un Réseau Ad Hoc Mobile

La définition d'*ad hoc* donnée par *Le Petit Larousse* est la suivante :

Ad hoc (adjectif): Qui convient à la situation, au sujet.

Un *réseau ad hoc* est un réseau de télécommunication établi malgré l'absence d'une infrastructure fixe. Des raisons de l'absence d'un réseau fixe peut être l'éloignement d'une région, par exemple dans le cas des animaux étiquetés ou des soldats dans un territoire étranger, ou bien il se peut qu'il soit indésirable, impraticable, ou simplement trop cher d'installer un réseau fixe.

l'idée qui fait la force d'un réseau ad hoc est que l'information peut passer d'un nœud à un autre par différents nœuds. Un *nœud* peut représenter une personne, un portable, un capteur de mesure, ou un dispositif électronique ayant la capacité de communiquer avec d'autres capteurs par des transmissions radio. Le fait que les réseaux ad hoc soient devenus si populaires au cours des dernières années est dû à leur facilité de déploiement, leur coût abordable, leurs avancées technologiques récentes, la réponse à un besoin (notamment après des attaques terroristes), et l'utilisation répandue d'instruments électroniques. On trouve constamment de nouvelles applications pour les réseaux ad hoc et la recherche dans ces réseaux est un domaine très vivant.

Les capteurs peuvent consister à n'être que des détecteurs qui sont créés dans le seul but de rassembler et de transmettre des données (les *réseaux de capteurs*). Les capteurs peuvent aussi faire partie d'un dispositif intégré comme un ordinateur portable ou un téléphone portable. Dans chaque cas, nous avons des messages qui doivent être envoyés d'un capteur à un autre.

Les utilisations des réseaux de capteurs sont nombreuses et semblent ne connaître presque aucune limite. Ils ont été utilisés pour mesurer une activité sismique, contrôler les bagages aux aéroports [20] [29], et les activités environnementales [81] [38], militaires [81] [38], et agricoles [30]. D'autres exemples d'utilisations sont les systèmes d'entretien pour l'identification du réservoir-remplissage ou du composant-remplacement, et la régulation du processus en temps réel pour l'automatisation industrielle. Si nous jetons un coup d'oeil à la surveillance d'habitat, les capteurs peuvent être utilisés pour :

- capter des microclimats
- surveiller les comportements des animaux
- identifier des changements d'habitat
- identifier des événements saisonniers peu communs (comme la migration des oiseaux)
- détecter les feux et empêcher leur expansion
- détecter les des secteurs contaminés et puis sonner des alarmes
- fournir des directions et des informations aux visiteurs.

Réseaux Ad Hoc Mobiles

Les nœuds dans les réseaux de capteurs sont souvent stationnaires et le transfert des données dépend de la construction des routes multi-sauts ou des *infostations* [32] [88]. À l'opposé de ces réseaux il existe les réseaux ad hoc mobiles (MANET), dans lesquels les nœuds sont mobiles. Cette mobilité peut être exploitée pour le transfert des données d'un nœud à un autre, en particulier si les nœuds dans le réseau ne sont pas réellement connectés. L'étude de la mobilité est le sujet de cette thèse.

La connectivité d'un réseau est déterminée principalement par la distance entre les nœuds. Une propriété des réseaux ad hoc mobiles est que les connections entre les nœuds sont continuellement établies et rompues. Une représentation est donnée par un graphique de recouvrement (*overlay graphs*) montrant les nœuds qui sont connectés les uns aux autres. D'autres réseaux similaires aux réseaux ad hoc mobiles et ayant aussi des représentations graphiques de recouvrement sont les graphes *petit-mondes* (small-world) et les réseaux *pair-à-pair*. Dans ces réseaux les connections entre les nœuds sont également établies et rompues et les données sont passées d'un nœud à un autre, en utilisant, quand cela est possible, d'autres nœuds comme relais. Il y a cependant des différences qui séparent les réseaux petit-mondes et pair-à-pair avec les réseaux ad hoc et qui valent donc la peine d'être expliquées:

- Le *phénomène petit-monde* est devenu populaire dans les travaux éminents [83] [84] de Watts & Strogatz. Ce phénomène est décrit par l'envoi d'une lettre accompagnée d'une description du destinataire. À chaque étape, une personne reçoit la lettre et la fait suivre à la personne la plus proche du destinataire qu'elle connaisse. Watts & Strogatz ont découvert qu'en moyenne, il y a environ six sauts de l'expéditeur au destinataire. Ce résultat est surprenant vu qu'il existe plus de six milliards de personnes dans le monde! Un réseau est appelé *un réseau petit-monde* s'il y a un nombre faible de sauts entre chaque paire de nœuds dans le réseau, indépendamment de la taille du réseau. La découverte de ce phénomène a entraîné la construction de modèles importants pour la diffusion de l'information entre les gens [9], sur l'Internet [10], et pour la diffusion des virus physiques et des virus technologiques [2].
- Les *réseaux pair-à-pair* sont devenus populaires par leur capacité à distribuer gratuitement de la musique par l'Internet, en particulier par les logiciels Napster¹ et Gnutella². L'idée sous-jacente au réseau pair-à-pair est que chaque pair, c'est-à-dire un ordinateur attaché au réseau, a des fichiers qui sont disponibles pour les autres pairs. Quand quelqu'un cherche un fichier (souvent un fichier d'une chanson ou d'un film), des nœuds intermédiaires aident dans la recherche et le transfert des données. Comme dans les réseaux ad hoc mobiles et petit-monde, nous pouvons représenter les pairs (nœuds) et leurs connections par un réseau de recouvrement. Comme dans les deux autres types de réseaux, les liens entre les nœuds sont souvent rompus et établis.

Les mesures de performance les plus importantes dans les réseaux petit-mondes et pair-à-pair sont : le nombre de nœuds à traverser avant de trouver un fichier, le nombre de nœuds atteints après un certain nombre de sauts, la méthode de recherche, et le passage à l'échelle d'un réseau. Dans les réseaux ad hoc mobiles, la distance physique entre les nœuds

¹Napster, en étant attaqué en justice pour atteinte aux droits l'industrie musicale, a fait de la publicité aux réseaux pair-à-pair. L'affaire a été perdue par Napster et maintenant elle propose une deuxième version payante aux utilisateurs pour le téléchargement de musique.

²Gnutella a été arrêté immédiatement après que America Online en a pris la direction, mais nãreverse engineeringã lui a donné une nouvelle vie. Tandis que Napster pouvait facilement être poursuivi a justice car il se sert d'une base de données centrale, Gnutella ne pouvait pas être poursuivi car il est décentralisé. La combinaison de leur publicité et leur contribution au trafic d'Internet a généré beaucoup d'activité de recherche [36] [23].

et la mobilité jouent également un rôle. La performance d'un réseau ad hoc mobile est donc caractérisée par la connectivité, le débit et le délai d'un message (en terme de temps au lieu du nombre de sauts).

Motivation et la Contribution de la Thèse

La plupart des publications sur les réseaux ad hoc mobiles concernent la connectivité ou le débit [47] [45] d'un réseau. Le travail de cette thèse est nouveau car nous étudions l'effet de la mobilité sur le délai d'un message. L'introduction d'un composant de mobilité rend les modèles mathématiques très difficiles à résoudre. Pour cette raison, très peu de résultats existent encore en la matière. Cette direction de recherche est toujours en développement et beaucoup de questions concernant l'influence de la mobilité et le délai d'un message sont encore à résoudre.

Le sujet principal dans la première partie de cette thèse est le *délai d'un message*, défini comme étant le temps requis pour transmettre un message d'une source à une cible, en utilisant la mobilité et les ressources d'autres nœuds.

Le débit et le délai d'un message sont influencés de manières différentes dans les réseaux ad hoc mobiles lorsqu'il n'y a pas toujours un lien direct entre deux nœuds. Pour s'assurer une transmission satisfaisante il y a plusieurs possibilités :

- L'existence des routes multi-sauts entre la source et la cible, en utilisant des nœuds intermédiaires, permet de transférer les données, donc elle a l'avantage de réduire le délai d'un message.
- Contrairement aux réseaux ad hoc statiques, les réseaux ad hoc mobiles présentent l'avantage que les nœuds intermédiaires peuvent transporter des données quand ils se déplacent. Dans cette situation, le délai d'un message n'est pas seulement constitué d'un délai de propagation et d'un délai d'une file d'attente, mais aussi par suite du temps (significatif) requis pour des nœuds de se placer près l'un de l'autre. Pour cette raison, les protocoles de relais déterminent aussi la performance d'un réseau ad hoc mobiles.
- Un nœud peut transmettre à une puissance plus haute. Cette augmentation de la gamme de transmission peut créer des routes entre deux nœuds, mais au prix de l'augmentation de l'usure des piles et des interférences. En outre, si les distances entre les nœuds sont grandes, l'augmentation de la gamme de transmission n'aura peut-être pas beaucoup d'effet sur la création des routes entre les nœuds. Cependant, comme nous allons voir dans la première partie de cette thèse, l'augmentation de la gamme de transmission a une (grande) influence sur le délai d'un message si la mobilité est utilisée pour le transfert des données.

À la base d'une étude sur la mobilité nous avons le *modèle de mobilité*, qui définit comment les nœuds se déplacent. Les nœuds sont très dynamiques dans certains modèles,

et dans d'autres ils se déplacent à proximité d'un point, ou ils restent même immobiles. Dans beaucoup de situations de la vie réelle, le modèle de mobilité est souvent inconnu ou est récemment mesuré en raison des avancements technologiques faits ces dernières années. À cause de la variabilité des mouvements des nœuds (souvent inconnus), il est important pour mieux mesurer la performance d'un réseau d'utiliser une variété de modèles de mobilité. Par exemple, la performance d'un nouveau protocole dépend-elle du modèle de mobilité ? Ou est-elle indépendante des mouvements des nœuds ? Ou bien encore, l'existence d'une région centrale change-t-elle la performance du protocole ?

Il y a plusieurs caractéristiques d'un réseau qui sont liées à un modèle de mobilité ; chacune est importante pour des raisons différentes. Ces caractéristiques sont :

- *La distribution spatiale des nœuds*³: la fonction de probabilité d'un nœud dans une certaine position si le système est dans un état stationnaire. Est-ce que les nœuds sont concentrés (groupés) autour des points chauds, ou sont-ils répandus sur la superficie ? Le retard du message peut être significativement plus petit si les nœuds se rassemblent autour des points chauds (en raison d'une région de retour en commun). Si les nœuds sont répandus également, i.e., la distribution stationnaire des nœuds est uniforme, les mathématiques deviennent des fois solubles, comme il a été démontré par les articles fructueux de Gupta & Kumar [47] et Grossglauser & Tse [45]. Cependant, la performance d'un réseau en terme de délai d'un message est un problème largement non résolu et il est une des contributions principales de cette thèse.
- *Le premier temps de rencontre* : à partir d'un temps aléatoire, c'est la différence de temps entre un temps aléatoire et l'instant où deux nœuds se retrouvent pour la première fois dans la gamme de communication de l'autre. Ce délai est important, non seulement pour le temps nécessaire au nœud source pour entrer en contact avec le nœud destination, mais aussi pour le temps nécessaire à des nœuds de relais pour entrer en contact avec la destination. En particulier, nous allons montrer (sous certaines hypothèses), que le temps de transfert d'un message est déterminé par le premier temps de rencontre entre deux nœuds.

La dérivation de ces mesures est faite dans la première partie de cette thèse. Ces calculs sont faits pour un nombre de modèles de mobilité à une dimension et à deux dimensions. Nous utilisons ensuite ces expressions afin d'obtenir le sujet central de la première partie : le temps moyen de transfert d'un message.

Un résultat intéressant et surprenant de cette partie est le rôle de l'intégrale sur le carré de la distribution spatiale des nœuds : celle-ci nous permet d'obtenir le premier temps de rencontre (le Théorème 4.2.1), qui joue à son tour un rôle central pour le délai de message sous un nombre de protocoles de relais (le Théorème 3.2.1).

D'autres points importants pour le délai de message : la vitesse des nœuds, la taille de surface sur laquelle les nœuds se déplacent, la gamme de transmission, et le protocole de relais. On penserait que le modèle de mobilité est aussi important. Cependant—sous

³Distribution stationnaire de la position d'un nœud.

l'hypothèse qu'il n'y pas d'interférences et que la gamme de transmission est minime comparée à la taille de surface—il résulte de cette thèse que seulement $\int \int \pi^2(x, y) dx dy$ et la vitesse moyenne des nœuds sont importantes, et *non* le modèle de mobilité lui-même! Évidemment, le modèle de mobilité détermine ces deux quantités, mais cela signifie que deux modèles de mobilité différents peuvent donner le même délai de message, pourvu que le produit de $\int \int \pi^2(x, y) dx dy$ et la vitesse moyenne soient les mêmes! En outre, nous allons montrer que le délai de message est linéaire dans la superficie et inversement linéaire dans la gamme de transmission et dans la vitesse.

Dans cette thèse il y a des contributions dans les domaines des réseaux ad hoc mobiles, des systèmes à "polling", des équations stochastiques et récursives, et des files d'attentes en tandem. Ces résultats sont distribués en deux parties et une appendice :

- I. Message Delay in Mobile Ad Hoc Networks
- II. Polling Systems with Correlated Switchover Times
- A. The Value Function of a Tandem Queue

La partie I est composée des chapitres 1-4. Il y a plusieurs contributions dans la première partie, mais la plus significative est la combinaison du Corollaire 4.2.2 (page 93) avec le Théorème 3.2.1 (page 67). Le corollaire dit que, sous certaines conditions et pour une certaine classe de mouvements, la distribution du premier temps de rencontre entre deux nœuds dans deux dimensions est exponentiellement distribuée avec un paramètre λ . Ce résultat est utilisé dans le Théorème 3.2.1 qui donne une expression pour le délai de message pour deux types de protocoles et plusieurs modèles de mobilités. En particulier, le délai est inversement lié au paramètre λ .

Bien que chaque chapitre soit construit sur l'antécédent, chaque chapitre se lit à part. Les contributions de chaque chapitre sont comme suit :

Dans le premier chapitre nous étudions un réseau ad hoc mobile à une dimension. Les nœuds se déplacent selon un mouvement Brownien dans des segments finis, ou chaque segment a des frontières réfléchives. Les nœuds peuvent communiquer seulement quand ils sont dans leur propre voisinage. Nous obtenons une expression pour le temps de transfert entre deux nœuds dans un segment, et nous obtenons également une expression pour la distribution de probabilité d'emplacement du transfert. Nous constatons que le temps de transfert s'exprime simplement par la gamme de transmission et la longueur d'intervalle. Ces résultats sont alors utilisés pour obtenir le délai d'un message en traversant une suite de nœuds dans des segments contigus.

Une analyse similaire est faite dans le Chapitre 2, mais cette fois pour des nœuds qui se déplacent selon une marche aléatoire sur un nombre d'états finis. Les résultats sont similaires, mais la nature discrète des mouvements nécessite des approches différentes. Parfois les calculs deviennent plus compliqués, et d'autres fois la notation avec des matrices et des implémentations numériques facilitent l'analyse. En particulier, l'expression pour le délai du message est semblable à l'expression quand les nœuds se déplacent selon un mouvement Brownien, sauf un terme qui apparaît à cause de la nature discrète du mouvement. Ce

terme disparaît pour un nombre d'états croissant. Un autre résultat donné dans ce chapitre est une expression pour le temps, à partir d'un temps aléatoire, jusqu'au moment où deux marches aléatoires sont à un certain nombre d'états l'une de l'autre. Les résultats dans ce chapitre peuvent être utilisés aussi pour des messages qui se promènent dans un réseau de capteurs.

Nous proposons également un modèle stochastique pour le calcul du temps de transfert d'un message dans un réseau ad hoc mobile, dans lequel les nœuds mobiles (ou simplement mobiles) servent de nœuds relais. Le modèle a deux paramètres d'entrée: le nombre de mobiles et l'intensité d'un nombre fini de processus de Poisson homogènes et indépendants. Ces processus de Poisson représentent les instants auxquels deux mobiles arbitraires peuvent communiquer. Nous calculons la transformée de Laplace-Stieltjes du temps de transfert (ou délai) d'un message, défini comme étant le temps nécessaire pour transmettre un message d'une source à une destination. De ce résultat nous déduisons le temps moyen de transfert d'un message, ainsi que son comportement asymptotique pour des réseaux avec beaucoup de mobiles. Nous déterminons également la distribution du nombre de copies d'un message dans le réseau à l'instant où le message atteint sa destination. Ces calculs sont effectués pour deux types de protocoles de routage qui utilisent les mobiles comme relais: le *two-hop relay protocol*, où seule la source peut utiliser les mobiles comme relais, et le *unrestricted relay protocol*, où tous les mobiles peuvent utiliser les autres mobiles comme relais. En dépit de sa simplicité, le modèle est capable de prédire les performances (délai, nombre de copies) de ces deux protocoles pour trois modèles différents de mobilité (Random Waypoint, Random Direction, Random Walker), comme le montrent des résultats de simulations.

La valeur du paramètre de la distribution exponentielle utilisée dans le Chapitre 3 est dérivée et discutée dans le Chapitre 4.

La partie II de cette thèse est constituée du Chapitre 5 dans lequel un système à *polling*, qui est constitué de deux files d'attente servies par un serveur est résolu. Après avoir servi une file d'attente, le serveur a besoin d'un temps de passage après lequel il commence à servir l'autre file d'attente. Les suites des temps de passages peuvent être corrélées. Nous obtenons l'expression de plusieurs quantités, dont notamment le temps d'attente moyen et la taille moyenne de la file d'attente. Grâce à ces expressions, nous comparons les différentes disciplines de service. Finalement, par des exemples nous montrons que la corrélation des temps de passage peut augmenter considérablement le temps d'attente moyen et la taille des files d'attente, indiquant que cette corrélation ne peut pas être ignorée. Cela a des implications importantes pour des systèmes de communication dans lesquels un canal de communication commun est partagé par plusieurs utilisateurs et où le temps entre des transferts de données consécutifs est corrélé, par exemple dans les réseaux ad hoc.

Dans l'appendice, nous étudions deux files d'attente en série avec des "holding costs" pour chaque client dans le système. La fonction de valeur est calculée pour le coût moyen lorsqu'il n'y a pas d'entrée de clients (le système avec l'entrée des clients reste un problème ouvert). Cette fonction est une combinaison des nombres de Catalan et de ballot et peut être utilisée pour l'optimisation des systèmes en série ou pour le calcul complet de la fonction de valeur.

Notation

Étant données des variables aléatoires X et Y , la notation

$$f_X(x) := \frac{\partial P(X \leq x)}{\partial x} \quad \text{et} \quad f_{X,Y}(x,y) := \frac{\partial^2 P(X \leq x; Y \leq y)}{\partial x \partial y},$$

est la distribution de la probabilité de X , respectivement (X, Y) .

Introduction

Disaster recovery, sensors dropped from aircraft, baggage handling at airports, soldiers in a foreign territory, environmental and habitat monitoring, household electronics, cars traveling on a road, and tagged animals all have one element in common: they can benefit greatly by the deployment of ad hoc networks.

The main focus in this thesis is the effect of the mobility on the performance of mobile ad hoc networks, in particular the effect it has on the message delay. Before we look at *mobile* ad hoc networks, let us first look at what an ad hoc network is.

What is an Ad Hoc Network?

The definition of *ad hoc* as given by the Webster dictionary [1] is

Ad hoc (adjective)

- 1.a : concerned with a particular end or purpose <*an ad hoc investigation committee*>
- 1.b : formed or used for specific or immediate problems or needs <*ad hoc solutions*>
- 2 : fashioned from whatever is immediately available : improvised <*large ad hoc parades and demonstrations*>

An *ad hoc network* refers to a communication network which has been established despite the lack of a fixed infrastructure. This lack of a fixed network can be due to the remoteness of a region, for example, for tagged animals or soldiers travelling through a foreign territory. Other reasons for the lack of a fixed infrastructure may be because it is undesirable, unfeasible, or simply too expensive.

The idea, and power, behind an ad hoc network is that information is passed from one node to another. Here a *node* can refer to a person, a measuring sensor, or any general electronic device. The strong increase in the popularity of ad hoc networks in the last couple of years is due to their envisioned ease of deployment, financial benefits, recent technological advancements, terrorist attacks, and widespread use of electronic devices. Because of their wide applicability and cheap installation, other uses of ad hoc network are continuously being found and therefore this field of technology and research will continue to grow for quite some time.

The electronic devices may consist only of sensors which are created with the sole purpose of collecting and passing on data (the so-called *sensor networks*). On the other hand, sensors can also be part of an integral device such as a laptop or a mobile phone. Whatever the form, the idea is that a message needs to be sent from one device to another.

For this reason the applications of sensor networks are enormous and seem to know no bounds. Sensor networks have been used to measure seismic activity, handle baggage at airports [20] [29], and to monitor environmental, military [81] [38], agricultural, and vineyard [30] activities. Other examples are maintenance systems for the early recognition of tank-refill or component-replacement as well as real-time process-control for industrial automation. If we take a look at the example of habitat monitoring in a national park, then the sensors can be used to capture micro-climates, monitor animal behaviour, identify changes in habitat, report unusual seasonal events like bird migration, detect fires and prevent its expansion, set alarms for contaminated land or water areas, and sense approaching visitors and provide online directions.

Mobile Ad Hoc Networks

As opposed to sensor networks which are often stationary and rely on a multi-hop path or infostations [32] [88] to transfer data, the nodes in mobile ad hoc networks (MANET) are mobile and form the main focus of this thesis. This mobility can easily be exploited to ensure the transfer of data from one node to another, especially if the nodes in the network are not well connected to each other.

In an ad hoc network, the connectivity of two nodes is generally determined by the physical distance between them. A particular property of mobile ad hoc networks is that connections between nodes are continuously set up and broken down. At any time, the network can be represented by an overlay graph showing the nodes which are connected to each other at that moment in time. Closely related to this are *small-world graphs* and *peer-to-peer* networks. Also in these networks, connections between nodes come and go and there is data that is passed from one node to another node, possibly by making use of intermediary nodes. There are however a number of distinctions which make these networks different from ad hoc networks and are therefore worth pointing out.

The *small-world phenomenon* was made popular by Watts & Strogatz in their seminal works [83] [84]. The basic idea starts with a letter that must be sent to someone in the world. The letter is accompanied by a description of the person to whom the letter needs to be sent, and whenever a person receives the letter (s)he passes it on to an acquaintance whom (s)he thinks is closer to the destination. It was found that everyone in the world is separated by an average of approximately six such hops from each other. This is a rather surprising result given that there are over 6 billion people in the world! Nowadays, if in a network—irrespective of its size—two nodes are only a small number of hops away from each other, then the network is said to have the small-world phenomenon. This study has led to important models for the spread of not only information between people, but also for the spread of both (physical) viruses and (technological) viruses which are sent over the internet through emails.

Peer-to-peer networks were made popular by Napster⁴ and Gnutella⁵ due to their publicity and their ability to distribute music on the internet for free. The basic idea is that each peer (i.e., computer attached to the peer-to-peer network) has files on its computer that are available to all of the other peers. When someone searches for a specific file (usually a music or a movie file), intermediary nodes are often used for both searching and for transferring the data. Here the network can also be represented by an overlay network showing the connections between the nodes at a certain moment in time. Just like in mobile ad hoc networks, the links between nodes are often broken down and rebuilt again.

In small-world and peer-to-peer networks there is large emphasis on the number of hops, the number of nodes reached per hop, search methods, and the scalability of the network. In mobile ad hoc networks however, the *physical* distances between the nodes play a role and therefore the performance of the network is instead characterized through the connectivity, throughput, and message delay (in terms of time instead of the number of hops).

Motivation and Contribution of the Thesis

Most of the existing literature on mobile ad hoc networks is concerned with either the connectivity or the throughput [47] [45] of a network. The work in this thesis is novel in the sense that the effect of the *mobility* on the message delay is studied. The introduction of a mobile component in models often renders them mathematically intractable and, for this reason, very few results exist on this matter. This direction of research is still in its infancy and much work needs to be done to better understand the influence of mobility *and* the mobility pattern on the delay of a message.

The main issue addressed in the first part of this thesis is the *message delay*, defined as the time required to transfer a message from a source to a destination node, while possibly making use of intermediary mobile nodes.

One characteristic of mobile ad hoc networks is that a direct link between two nodes does not always exist, and this affects the throughput and the message delay in different ways. To ensure a satisfactory transmission of the data a number of different options exist.

- If a network is not sparsely distributed, then multi-hop paths between a source and a destination node may exist. The intermediary nodes can then relay the data thus greatly reducing the message delay.
- As opposed to static ad hoc networks, mobile ad hoc networks have the additional

⁴Napster brought peer-to-peer networks into the news by being engaged in a legal lawsuit concerning music rights. The court case was lost by Napster and now a second version of Napster charges users to download music.

⁵Gnutella was discontinued immediately after America Online took it over but with reverse engineering it was brought back to life again. Whereas Napster made use of a central database, Gnutella is decentralized, which meant that it could not be pursued by lawsuits. The idea of all nodes contributing information to the network and the nodes helping each other find the information has led to a surge of research activity [36] [23].

advantage that an intermediary node can relay data while it is moving around. The latter situation means that the message delay is not only made up of a propagation and queueing delay, but also of a (possibly significant) amount of time required for nodes to move closer to each other. In this case, relay protocols help determine the performance of the network.

- Another method of influencing the performance of the network is to have a node transmit at a higher power, thus increasing its transmission range. This may cause a path between two nodes to come into existence, but it this does come at the price of valuable battery power. If the distances between the nodes are large, then increasing the transmission range may not have much effect on the creation of a direct path between a source and a destination node. However, as will be seen in the first part of this thesis, this will definitely have a positive influence on the message delay if the mobility of nodes is used to transmit a message.

At the basis of the study of the effects of mobility is the *mobility model*. The mobility model defines how nodes move around. In some cases, the nodes are very dynamic; in other cases, nodes remain around the same location or even pause. In many real-life situations, the mobility pattern is often unknown or is only just being measured due to the technological advancements made in recent years. Because of the variability of, and often unknown, movement patterns, it is important to study the performance of a network under a variety of different mobility models. For example, do conclusions of the performance of a new protocol depend on the mobility model or do they hold in general?

Fixing a mobility model results in a number of characteristics, each of them important in a different way. These characteristics are:

- *Spatial node distribution*⁶: the probability of a node being in a certain location once the system has reached steady-state. Are the nodes clustered around hot-spots or are they spread out evenly over the surface area? If the nodes congregate around hot-spots, then the message delay can be significantly smaller due to nodes coming back, or hanging around, these hot-spots. If the nodes are spread out evenly, i.e., the spatial node distribution is uniform, then systems often become mathematically tractable as exemplified by the seminal papers of Gupta & Kumar [47] and Grossglauser & Tse [45]. However, the performance of a network in terms of message delay remains largely an open question and this forms the main contribution of the first part of this thesis.
- *First-meeting times*: starting from a random time, this is the time until two nodes are within communication range of one another for the first time. This quantity is important not only for the time needed for the source node to come into contact with the destination node, but also for the time needed for relay nodes to come into contact with each other and with the destination node. In particular, it will be shown that under a number of realistic assumptions, the message delay is determined by the first-meeting times of nodes.

⁶Also referred to as the steady-state location or the stationary distribution of the location of the nodes.

The derivation of these measures for a number of the mobility models is performed in the first part in this thesis, in both one and two dimensions. These quantities will then be used to derive the central topic of the first part, namely the mean message delay. One very interesting and surprising result is that the integral over the square of the spatial node distribution plays a central role for the first-meeting times (Theorem 4.2.1), which in turn plays a central role for the message delay under a number of different relay protocols (Theorem 3.2.1).

What else is of importance for the message delay? Naturally the speed of the nodes, the surface area, the transmission range and the relay protocol play a role. One would think that the mobility model itself plays an important role as well. However, assuming there is no interference and $r \ll L$, then one important result derived in this thesis is that only $\int \int \pi^2(x, y) dx dy$ and the average relative speed of the nodes is of importance and NOT the mobility pattern itself! Naturally the mobility pattern determines these two quantities, but this means that completely different mobility models can give the same message delay (as long as the product of $\int \int \pi^2(x, y) dx dy$ and the average relative speed is the same)! Furthermore, it will be shown that the message delay is linear in the area size and inversely linear in the transmission range and the speed.

Organisation of the Thesis

This thesis makes a number of contributions to the fields of mobile ad hoc networks, polling systems, stochastic recursive equations, and tandem queues. The results are distributed over two parts and an appendix:

- I. Message Delay in Mobile Ad Hoc Networks;
- II. Polling Systems with Correlated Switchover Times;
 - A. The Value Function of a Tandem Queue.

Part I is composed of Chapters 1-4. Although many contributions are given in the first part, perhaps most valuable are the combination of Corollary 4.2.2 on page 93 and Theorem 3.2.1 on page 67. The former states that for certain movement patterns, under certain conditions, the first-meeting time between two nodes in two dimensions is exponentially distributed with parameter λ . The latter theorem makes use of this result to give closed form expressions for the message delay using two different relaying protocols for a number of different mobility patterns in two dimensions. In particular, the message delay is inversely related to λ .

Although there is a build-up and link between the chapters, each chapter is self-contained. The contributions of each chapter are outlined in the following paragraphs.

In Chapter 1, a simple one-dimensional ad hoc network topology is analysed. Nodes move as Brownian motions in finite segments, where each segment has reflecting boundaries. Communications (or relays) between nodes can occur only when they are within transmission range of each other. An expression for the transfer time between two nodes in one segment

is obtained as well as a closed form expression for the probability density function of the location where a relay takes place. It is worth noting that the average message delay between two nodes reveals a simple expression in terms of the transmission range and the length of the interval. These results are then used to obtain the message delay across of sequence of nodes moving in adjacent segments.

A similar analysis is repeated in Chapter 2 but then for nodes moving as Random Walkers on a finite number of states. Although similar expressions are obtained, the discrete nature requires a slightly different analysis. At times the mathematics is more involved and at other times matrix notation and the possibility of numerical implementation simplify matters. In particular, the same expression for the message delay as under the Brownian Motion mobility model is obtained up to a factor difference which accounts for the discrete nature of the process. This factor vanishes for an increasing number of states. A nice result given in this chapter is an expression for the time it takes two random walkers on the same state space to come with a certain number of states from each other. These results can also be used for two messages travelling across a sensor network.

In Chapter 3, a simple stochastic model is introduced that accurately models the message delay in two-dimensional mobile ad hoc networks where nodes relay messages for each other. The model has only two input parameters: the number of nodes and the parameter of an exponential distribution which models the first-meeting time between two nodes. Closed-form expressions are obtained for the Laplace-Stieltjes transform of the message delay. From this result, the expected message delay in closed-form is obtained as well as its asymptotic expansion for networks with many nodes. The probability distribution of the number of copies of the message at the time the message is delivered is also computed. These calculations are carried out for two relay protocols: the two-hop relay and the unrestricted relay protocols. Despite its simplicity, the model is able to accurately predict the performance of both relay protocols for a number of mobility models (random waypoint, random direction and the random walker mobility models), as shown by simulations.

The exponential parameter used in Chapter 3 is derived in Chapter 4.

Part II of this thesis is made up of a single chapter, namely Chapter 5. In this chapter, a queueing system is analysed in which service is alternated between two queues and the server requires a (finite) switchover time to switch from one queue to the other. The distinction from classical results is that the sequence of switchover times from each of the queues need not be i.i.d. nor independent from each other; each sequence is merely required to form a stationary ergodic sequence. With the help of stochastic recursive equations, explicit expressions are derived for a number of performance measures, most notably for the average delay of a customer and the average queue lengths under different service disciplines. With these expressions, a comparison is made between the service disciplines, and the influence of correlation is studied. Finally, through a number of examples, the correlation is shown to significantly increase the mean delay and the average queue lengths indicating that the correlation between switchover times should not be ignored. This has important implications for communication systems in which a common communication channel is shared amongst various users and where the time between consecutive data transfers is correlated (one can also think of a narrow bridge with cars travelling over the bridge from one direction at a

time). The position of the server then corresponds to the direction data is travelling in. A similar situation arises in ad hoc networks; there is a common channel which needs to be shared amongst various users. The higher the number of users, the longer one has to wait before being able to capture the channel necessary to (re)transmit data. In particular, if one has to wait a long time before being able to transmit data, then it is very likely that there are many users around and that the next time one has to wait once again for a long period of time. For this reason the correlation of the number of users over time in an ad hoc network inherently introduces correlation between the waiting (switchover) times, and this in turn leads to an increase in the mean delay and queue lengths.

The appendix of the thesis lies separate from the rest of the thesis as it deals with the derivation of the value function (also known as the bias vector) of the average costs of the tandem queue. It remains an open problem in queueing theory. Once the value function is obtained, it can be used for important optimisation purposes. Although the final expression remains out of reach, a closed-form expression is however obtained for the value function if there is no inflow of customers into the first queue. This expression is in terms of Catalan and ballot numbers which give an intuitive explanation to the problem at hand.

Notation

For any random variables X and Y , let

$$f_X(x) := \frac{\partial P(X \leq x)}{\partial x} \quad \text{and} \quad f_{X,Y}(x, y) := \frac{\partial^2 P(X \leq x; Y \leq y)}{\partial x \partial y},$$

be the probability density function (pdf) of respectively X and (X, Y) .

Message Delay in *Mobile* Ad Hoc Networks

Message Delay for One-dimensional Brownian Motions

Contents

1.1	Introduction	12
1.2	Two mobiles moving along a line segment	14
1.3	A chain of relaying mobiles	18
1.4	Numerical results and discussion	21
1.5	Extensions to the model	23
1.6	Concluding remarks	24
1.A	The Brownian motion	25
1.B	Proof of Proposition 1.2.2	27
1.C	Proof of Proposition 1.3.1	29

Mobile ad hoc networks are characterized by a lack of a fixed infrastructure and by node mobility. In these networks data transfer can be improved by using mobile nodes as relay nodes. As a result, transmission power and the movement pattern of the nodes have a key impact on the performance. In this chapter the impact of node mobility is studied through the analysis of a simple one-dimensional ad hoc network topology. Nodes move in adjacent segments with reflecting boundaries according to Brownian motions. Communications (or relays) between nodes can occur only when they are within transmission range of each other. The expected time to relay a message is determined along with the probability density function of the relay locations. Finally, an approximation formula is provided for the expected relay time between any pair of mobiles.

Note: The material presented in this chapter has been published in the proceedings of WiOpt¹ [41] and has been submitted to the ACM Kluwer journal of Wireless Networks.

¹Unfortunately in [41] the formulas published are a factor $\sqrt{2}$ off from the correct result.

1.1 Introduction

Ad hoc networks can be deployed when a fixed network structure is not available. The lack of a fixed infrastructure may arise in emergency situations, remote regions, hostile areas or, as is often the case, due to the financial costs involved in the deployment of a fixed infrastructure.

As a consequence of the absence of a fixed infrastructure, components (or nodes) of an ad hoc network need to behave as routers by relaying messages in order to improve communications. Instances of nodes in ad hoc networks are laptops, planes [76], cars, electronic tags on animals [73], mobile phones, et cetera. If nodes are mobile then operating these networks become even more complex, as mobility will impact routing protocols, control of transmission power, quality of service (e.g., interference, path loss, shadowing effects), battery usage, to name but a few.

As long as data does not have to be transferred directly between two mobiles and that nodes are willing to relay messages, their mobility may have a positive impact on the performance, as shown in [45]. This has led to the design of protocols that take advantage of node mobility to enhance the performance of some applications (e.g., messaging applications in [63]). Data relaying cuts down transmission power, interferences and increases battery usage. On the other hand, it may increase latency—since the existence at any time of a “path” between two mobiles is not guaranteed—even if (intermediary) nodes can be used as routers to convey a message from its source to its destination.

This chapter studies the impact of mobility on the latency in the case of nodes acting as relay nodes. This is done for one-dimensional ad hoc network topologies and under the assumption that nodes move according to (independent) Brownian motions.

A natural approach (but not the only one, see [71] for another approach) to modeling a mobile ad hoc network with relaying nodes consists of looking down at the earth and representing it as a finite two-dimensional plane. If two mobiles are within a fixed transmission range of each other then a message can be relayed/transmitted (see Figure 1.1). Furthermore, mobiles move according to a certain movement pattern. Unfortunately, this simple model of an ad hoc network (no physical restrictions in the area covered by the nodes, nodes are homogeneous, etc.) is extremely difficult to analyse, even with simple movement patterns such as, for example, the Random waypoint mobility model (which is described in Section 3.3.1.1 or in [86]). For instance, obtaining an exact expression for the stationary distribution of the location of the mobiles under the Random waypoint mobility model in two dimensions remains an open problem.

Obtaining any results characterizing the first instance of time when two mobiles come within transmission range of each other is a problem of even greater complexity. For this reason, this chapter focuses on a one-dimensional topology—a model that already reveals interesting properties. Although the extension of the theory presented in this chapter to a two dimensional scenario does not seem to be possible, the insights gained have led to the developments of the results in Chapter 3, of which most notably the Laplace-Stieltjes transform of the message delay.

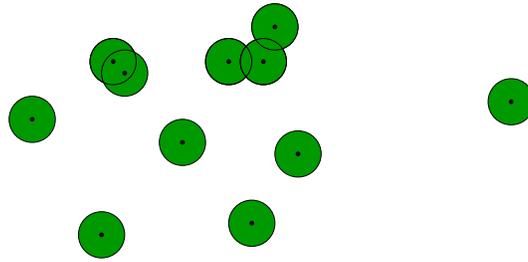


Figure 1.1: Visual representation of an ad hoc network. Node can transfer a message only if they are with each others transmission range. In this case only two of the nodes can communicate with each other.

When analyzing a mobile ad hoc network, an important consideration is the movement pattern. Are mobiles restricted in their movement by roads, physical objects, waterways, or mountains? Do they roam around a central point? It has been shown that the latter is the case for the Random waypoint mobility model, where there is a higher concentration of mobiles around a central region (see Section 4.B for an example) .

The following scenarios are addressed in this chapter. In Section 1.2 the situation is considered where two mobiles move along a segment with reflecting boundaries (see Figure 1.2). Both mobiles move along the segment according to independent Brownian motions. We are interested in computing the expected time until both mobiles come within communication range of each other. This quantity is computed for any given initial locations (Proposition 1.2.1) as well as for the case where each Brownian motion is initially in steady-state (Proposition 1.2.3). It is known (see Section 1.3) that the latter assumption implies that both mobiles are uniformly distributed over the segment. The uniform spatial distribution over the coverage area has attracted attention lately and several fundamental results [8] [47] [45] have been obtained in this setting. However, the model presented here is different from the models considered in those papers.

In Section 1.3, I mobiles on I segments are considered, one mobile per segment, as depicted in Figure 1.5. The mobiles move along their respective segment (with reflecting boundaries) according to independent Brownian motions. The goal is to determine the expected message delay between the first and last mobile in the sequence (Proposition 1.3.2), that is, the expected time to transfer a message from the first to the last mobile. As an additional result, the probability density function (pdf) of the position of a mobile at a relay epoch (Proposition 1.3.1) is identified. Numerical results are reported in Section 1.4. These results suggest an accurate and scalable approximation for the expected message delay in one dimension (see equation (1.15)). The possible extensions of the model are discussed in Section 1.5.

1.2 Two mobiles moving along a line segment

Consider two mobiles (say mobiles \mathbf{X} and \mathbf{Y}) moving along segment $[0, L]$. See Figure 1.2. Communications between these two mobiles occur only when the distance between them is less than or equal to $r \leq L$. The objective of this section is to determine the expected *first-meeting time*, defined as the first time when both mobiles come with a distance r of each other.

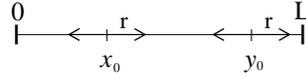


Figure 1.2: Two mobiles moving along $[0, L]$ with transmission range r .

Let $x(t)$ and $y(t)$ be the position of mobiles \mathbf{X} and \mathbf{Y} , respectively, at time t . We assume that $\mathbf{X} = \{x(t), t \geq 0\}$ and $\mathbf{Y} = \{y(t), t \geq 0\}$ are identical and independent Brownian motions with drift 0 and diffusion coefficient² D , both moving along the segment $[0, L]$ with *reflecting* boundaries at the edges. Let $T_{L,r}$ be the first-meeting time, namely,

$$T_{L,r} = \inf\{t \geq 0 : |y(t) - x(t)| \leq r\}. \quad (1.1)$$

Set $x(0) = x_0$ and $y(0) = y_0$. By convention we assume that $T_{L,r} = 0$ if $|y_0 - x_0| \leq r$. From now on we assume that $|y_0 - x_0| > r$.

We are interested in

$$T_{L,r}(x_0, y_0) := \mathbb{E}[T_{L,r} \mid x(0) = x_0, y(0) = y_0], \quad 0 < x_0, y_0 < L,$$

the expected first-meeting time given that mobiles \mathbf{X} and \mathbf{Y} are located at position x_0 and y_0 , respectively, at time $t = 0$. The following result holds:

Proposition 1.2.1 (Expected first-meeting time with given initial positions)

For $0 \leq x_0 < y_0 \leq L$ with $x_0 + r < y_0$ and $0 \leq r \leq L$

$$T_{L,r}(x_0, y_0) = \frac{32(L-r)^2}{D\pi^4} \sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \frac{\sin\left(\frac{m\pi(y_0+x_0-r)}{2(L-r)}\right) \sin\left(\frac{n\pi(y_0-x_0-r)}{2(L-r)}\right)}{mn(m^2+n^2)}. \quad (1.2)$$

◇

The proof of Proposition 1.2.1 is based on the following intermediary result that gives the expected time for a two-dimensional Brownian motion \mathbf{Z} evolving in an R by R square to hit any boundary of the square.

²i.e., $x(t+h) - x(t)$ (respectively $y(t+h) - y(t)$) is normally distributed with mean 0 and variance $2Dh$ for all $h > 0$, and non-overlapping time intervals are independent of each other. See Appendix 1.A for a more detailed description of the Brownian motion.

Proposition 1.2.2 (Two Brownian motions in a square)

Consider two independent and identical one-dimensional Brownian motions $\{u(t), t \geq 0\}$ and $\{v(t), t \geq 0\}$, with zero drift and diffusion coefficient D . Define the two-dimensional Brownian motion $\mathbf{Z} = \{z(t) = (u(t), v(t)), t \geq 0\}$. Set $u_0 = u(0)$ and $v_0 := v(0)$ and assume that $0 \leq u_0 \leq R$ and $0 \leq v_0 \leq R$.

Let

$$\tau_R := \inf\{t \geq 0 : u(t) \in \{0, R\} \text{ or } v(t) \in \{0, R\}\}$$

be the first time when the process \mathbf{Z} hits the boundary of a square of size R by R .

Define $\tau_R(u_0, v_0) = \mathbb{E}[\tau_R | z(0) = (u_0, v_0)]$. Then,

$$\tau_R(u_0, v_0) = \frac{16R^2}{D\pi^4} \sum_{\substack{m \geq 1 \\ m \text{ odd}}}^{\infty} \sum_{\substack{n \geq 1 \\ n \text{ odd}}}^{\infty} \frac{\sin\left(\frac{m\pi u_0}{R}\right) \sin\left(\frac{n\pi v_0}{R}\right)}{mn(m^2 + n^2)}. \quad (1.3)$$

◇

The proof of Proposition 1.2.2 is given in Appendix 1.B. We are now in a position to prove Proposition 1.2.1.

Proof of Proposition 1.2.1: Let $x_0 + r < y_0 \leq L$. An equivalent way to view the Brownian motions \mathbf{X} and \mathbf{Y} at time $t = 0$ is to consider that the point (x_0, y_0) is located in the upper triangle in Figure 1.3 delimited by the lines $x = 0$, $y = L$ and $y = x + r$. If we assume that the boundaries $x = 0$ and $y = L$ are reflecting boundaries in Figure 1.3, then we see that $T_{L,r}(x_0, y_0)$ is nothing but the expected time needed for the two-dimensional Brownian motion $\{(x(t), y(t)), t \geq 0\}$ to hit the diagonal of the triangle (i.e., to hit the line $y = x + r$) given that $(x(0), y(0)) = (x_0, y_0)$. (The process $\{(x(t), y(t)), t \geq 0\}$ is a two-dimensional Brownian motion since $\{x(t), t \geq 0\}$ and $\{y(t), t \geq 0\}$ are both independent Brownian motions.)

By using the classical method of images (see e.g., [68, p. 81]), it can be seen that this time is itself identical to the expected time needed to hit the boundary of the square of size $\sqrt{2}(L - r)$ by $\sqrt{2}(L - r)$ shown in Figure 1.4 given that $(x(0), y(0)) = (x_0, y_0)$. This is due to the reflecting boundaries at $x = 0$ and $y = L$ acting as mirrors.

In order to apply the result in Proposition 1.2.2, we need to compute the coordinates (x'_0, y'_0) of (x_0, y_0) in a new system of coordinates (x', y') depicted in Figure 1.4 and which is rotated 45° from the original coordinate system. We find $(x'_0, y'_0) = ((y_0 + x_0 - r)/\sqrt{2}, (y_0 - x_0 - r)/\sqrt{2})$ and we may conclude, from Proposition 1.2.2, that

$$T_{L,r}(x_0, y_0) = \tau_{\sqrt{2}(L-r)}\left(\frac{(y_0 + x_0 - r)}{\sqrt{2}}, \frac{(y_0 - x_0 - r)}{\sqrt{2}}\right). \quad (1.4)$$

By using (1.3) in the right hand side of (1.4) we see that (1.2) holds. ■

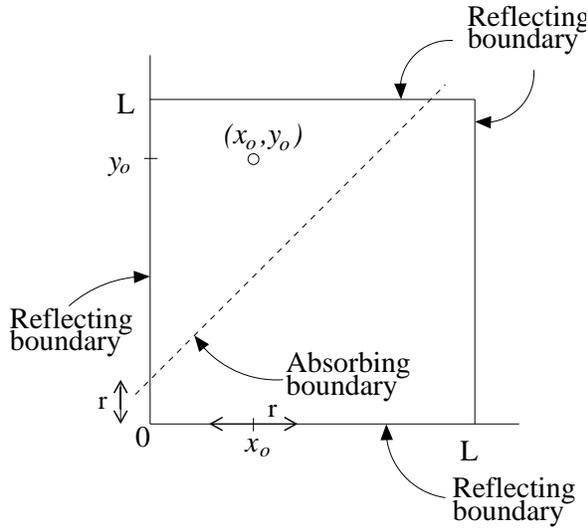


Figure 1.3: When mobiles \mathbf{X} and \mathbf{Y} are at a distance r of each other they are located on the line $y = x + r$ ($y_0 > x_0 + r$).

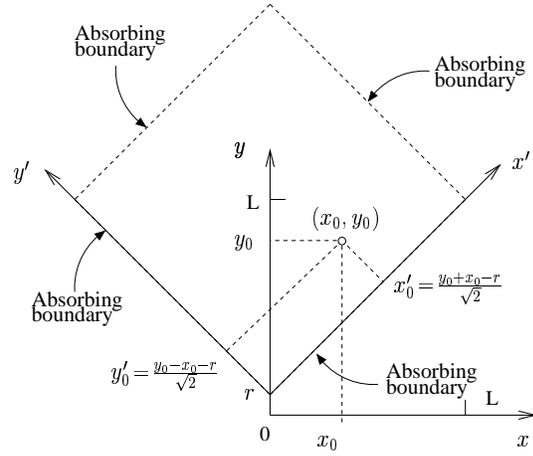


Figure 1.4: Since reflecting barriers at $x = 0$ and $y = L$ act as mirrors, the method of images turns the problem into a 2D Brownian motion inside four absorbing barriers.

An example of the expected first-meeting time $T_{L,r}(x_0, y_0)$ is displayed in Figure 1.6 (see Section 1.4 for comments).

We conclude this section by giving the expected first-meeting time when both mobiles are uniformly distributed over the segment $[0, L]$ at time $t = 0$. We will see in the next section that this case corresponds to the situation where both Brownian motions \mathbf{X} and \mathbf{Y} are in steady-state at time $t = 0$.

Proposition 1.2.3 (Expected first-meeting time for uniform initial positions)

Assume that both mobiles \mathbf{X} and \mathbf{Y} are uniformly distributed over $[0, L]$ at time $t = 0$ and $0 \leq r \leq L$. The expected first-meeting time $\mathbb{E}[T_{L,r}]$ is

$$\mathbb{E}[T_{L,r}] = \frac{128(L-r)^4}{D\pi^6 L^2} C_0, \quad (1.5)$$

where C_0 is a constant given by $C_0 = \sum_{m=1}^{\infty} \sum_{n \text{ odd}} \frac{1}{m^2 n^2 (m^2 + n^2)} \approx 0.527927$. \diamond

Proof: Since \mathbf{X} and \mathbf{Y} are uniformly distributed at $t = 0$, we have

$$\begin{aligned}
 \mathbb{E}[T_{L,r}] &= \frac{1}{L^2} \int_0^L \int_0^L \mathbb{E}[T_{L,r} | x(0) = x_0, y(0) = y_0] dx_0 dy_0 \\
 &= \frac{1}{L^2} \int_{x_0+r < y_0 \leq L} T_{L,r}(x_0, y_0) dx_0 dy_0 + \frac{1}{L^2} \int_{y_0+r < x_0 \leq L} T_{L,r}(y_0, x_0) dx_0 dy_0 \\
 &= \frac{2}{L^2} \int_{x_0+r < y_0 \leq L} T_{L,r}(y_0, x_0) dx_0 dy_0 \\
 &= \frac{64(L-r)^2}{D\pi^4 L^2} \int_{x_0+r < y_0 \leq L} h(y_0+x_0-r, y_0-x_0-r) dx_0 dy_0.
 \end{aligned}$$

where

$$h(u, v) := \sum_{\substack{m \geq 1 \\ m \text{ odd}}}^{\infty} \sum_{\substack{n \geq 1 \\ n \text{ odd}}}^{\infty} \frac{\sin(mu\beta) \sin(nv\beta)}{mn(m^2 + n^2)}, \quad \beta := \frac{\pi}{\sqrt{2}(L-r)}.$$

Define the new variables $u = (y_0 + x_0 - r)/\sqrt{2}$ and $v = (y_0 - x_0 - r)/\sqrt{2}$. We find

$$\begin{aligned}
 \mathbb{E}[T_{L,r}] &= \frac{64(L-r)^2}{D\pi^4 L^2} \left[\int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=0}^u h(u, v) |J(u, v)| dv du \right. \\
 &\quad \left. + \int_{u=\frac{L-r}{\sqrt{2}}}^{\sqrt{2}(L-r)} \int_{v=0}^{\sqrt{2}(L-r)-u} h(u, v) |J(u, v)| dv du \right] \tag{1.6}
 \end{aligned}$$

where $|J(u, v)|$ ($=1$) is the determinant of the Jacobian matrix

$$J(u, v) = \begin{pmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

It remains to evaluate the two double integrals in (1.6). By making use of the identity $h(u, v) = h(\sqrt{2}(L-r) - u, v)$ we see that both integrals in the right hand side of (1.6) are equal, since

$$\int_{u=\frac{L-r}{\sqrt{2}}}^{\sqrt{2}(L-r)} \int_{v=0}^{\sqrt{2}(L-r)-u} h(u, v) dv du = \int_{u=\frac{L-r}{\sqrt{2}}}^{\sqrt{2}(L-r)} \int_{v=0}^{\sqrt{2}(L-r)-u} h(\sqrt{2}(L-r)-u, v) dv du = \int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=0}^u h(u, v) dv du.$$

The first integral can be evaluated by using the symmetry $h(u, v) = h(v, u)$. This gives

$$\begin{aligned}
 \int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=0}^u h(u, v) dv du &= \int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=0}^u h(v, u) dv du = \int_{v=0}^{\frac{L-r}{\sqrt{2}}} \int_{u=v}^{\frac{L-r}{\sqrt{2}}} h(v, u) du dv \\
 &= \int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=u}^{\frac{L-r}{\sqrt{2}}} h(u, v) dv du.
 \end{aligned}$$



Figure 1.5: A chain of relaying mobiles.

Hence,

$$\int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=0}^u h(u, v) dv du = \frac{1}{2} \int_{u=0}^{\frac{L-r}{\sqrt{2}}} \int_{v=0}^{\frac{L-r}{\sqrt{2}}} h(u, v) dv du$$

so that

$$\mathbb{E}[T_{L,r}] = \frac{64(L-r)^2}{D\pi^4 L^2} \int_0^{\frac{L-r}{\sqrt{2}}} \int_0^{\frac{L-r}{\sqrt{2}}} h(u, v) dv du. \quad (1.7)$$

Since the double series in $h(u, v)$ are uniformly bounded in the variables $u, v \in [0, \sqrt{2}(L-r)]$ (its absolute value is bounded from above by $(\sum_{k \geq 1} 1/k^2)^2 = \pi^4/36$), we may invoke the bounded convergence theorem to interchange the integral and summation signs in (1.7). This gives

$$\begin{aligned} \mathbb{E}[T_{L,r}] &= \frac{64(L-r)^2}{D\pi^4 L^2} \sum_{\substack{m \geq 1 \\ m \text{ odd}}}^{\infty} \sum_{\substack{n \geq 1 \\ n \text{ odd}}}^{\infty} \frac{1}{mn(m^2 + n^2)} \int_{u=0}^{\frac{L-r}{\sqrt{2}}} \sin(mu\beta) du \int_{v=0}^{\frac{L-r}{\sqrt{2}}} \sin(nv\beta) dv \\ &= \frac{128(L-r)^4}{D\pi^6 L^2} \sum_{\substack{m \geq 1 \\ m \text{ odd}}}^{\infty} \sum_{\substack{n \geq 1 \\ n \text{ odd}}}^{\infty} \frac{1}{m^2 n^2 (m^2 + n^2)}. \end{aligned}$$

The last line follows because $\cos(j\pi/2) = 0$ for j odd. ■

1.3 A chain of relaying mobiles

Consider the situation depicted in Figure 1.5. There are I adjacent segments, each of length L , and there is a single mobile per segment. We denote by \mathbf{X}_i the mobile in segment i . Let $0 \leq x_i(t) \leq L$ ($i = 1, \dots, I$) be the *relative* position of the i -th mobile in its segment. We assume that the process $\mathbf{X}_i = \{x_i(t), t \geq 0\}$ is a Brownian motion with zero drift and diffusion coefficient D and that $\mathbf{X}_1, \dots, \mathbf{X}_I$ are mutually independent processes. Last, we assume that each segment has reflecting boundaries at the ends.

Let $T_1 = \inf\{t \geq 0 : x_1(t) + r \geq L + x_2(t)\}$ be the transfer time between mobiles \mathbf{X}_1 and \mathbf{X}_2 , that is T_1 is the first time when \mathbf{X}_1 and \mathbf{X}_2 are located at a distance less than or equal to r from each other. The relay times $T_2 \leq \dots \leq T_{I-1}$ between mobiles \mathbf{X}_2 and \mathbf{X}_3 , \dots , \mathbf{X}_{I-1} and \mathbf{X}_I , respectively, are recursively defined by

$$T_i = \inf\{t \geq T_{i-1} : x_i(t) + r \geq L + x_{i+1}(t)\}, \quad i = 2, \dots, I-1.$$

The objective of this section is to compute $\mathbb{E}[T_i]$ for $i = 1, \dots, I - 1$.

Throughout this section it will be assumed that $L \leq r \leq 2L$. This assumption is made for the sake of mathematical tractability. Indeed, a few seconds of reflection will convince the reader that when³ $L \leq r \leq 2L$ and $(x_1(0), x_2(0)) = (x_0, y_0)$ the time needed to transfer a message between two adjacent segments is the same as $T_{2L,r}(x_0, y_0 + L)$, the expected first-meeting time obtained in Section 1.2 for a segment of length $2L$ (with the given initial conditions). This observation allows us to find at once the expected transfer time between mobiles \mathbf{X}_1 and \mathbf{X}_2 for any initial conditions $x_1(0)$ and $x_2(0)$. We find

$$\mathbb{E}[T_1 \mid x_1(0) = x, x_2(0) = y] = \mathbb{E}[T_{2L,r}(x, y + L)]. \quad (1.8)$$

The difficulty arises when trying to find the expected relay times between mobiles \mathbf{X}_i and \mathbf{X}_{i+1} for $i = 2, \dots, I - 1$, since the position of \mathbf{X}_i when the transfer between \mathbf{X}_{i-1} and \mathbf{X}_i takes place is not uniform in $[iL, (i + 1)L]$.

To overcome this difficulty, we assume that the Brownian motions $\mathbf{X}_1, \dots, \mathbf{X}_I$ are all in steady-state at time $t = 0$. This assumption implies, ² in particular, that the position of each mobile at time $t = 0$ is uniformly distributed over its segment (i.e., the pdf of $x_i(0)$ is uniform over $[0, L]$). The same holds of course at any arbitrary time (i.e., the pdf of $x_i(t)$ is uniform over $[0, L]$ if t is arbitrary).

Another consequence of this assumption is that the position of mobile \mathbf{X}_{i+1} at time T_{i-1} (i.e., when \mathbf{X}_i receives a message from \mathbf{X}_{i-1}) is still uniformly distributed over $[0, L]$. This property will be used later on.

Proposition 1.3.1 below addresses the location of a mobile at the time when a relay occurs. For later reference, we state the result in a general form. Consider two adjacent segment, each of length L , with a single mobile in each segment (mobile \mathbf{X} in the first segment and \mathbf{Y} in the second segment). Both mobiles move in their segment (with reflecting boundaries) according to independent and identical Brownian motions with zero drift and coefficient diffusion D . We assume that the Brownian motion representing the movement of \mathbf{Y} is in steady state at time $t = 0$. As usual, a relay will occur the first time when both mobiles come within a distance r of each other, with $L \leq r \leq 2L$.

Proposition 1.3.1 (Pdf of location at relay epoch)

Fix $L \leq r \leq 2L$. Let $q(y; x)$, $y \in [0, L]$, be the pdf of the (relative) position of mobile \mathbf{Y} at the relay epoch, given that at time $t = 0$ the mobile \mathbf{X} is at position x and the position of mobile \mathbf{Y} is uniform.

³When $L \leq r \leq 2L$ the reflecting boundaries conditions are at $x_1(t) = 0$ and at $x_2(t) = 2L$. For $0 \leq r \leq L$ there are two additional reflecting boundary conditions at $x_1(t) = L$ and at $x_2(t) = L$ which lead to a much more difficult problem which we were not able to solve.

²Hint: let $\pi(x)$ be the stationary density probability that the mobile is in position $x \in [0, L]$. Solving the diffusion equation $D\partial^2\pi(x)/dx^2 = 0$ with the reflecting conditions $d\pi(x)/dx = 0$ for $x \in \{0, L\}$ and the normalizing condition $\int_0^L \pi(x)dx = 1$ yields $\pi(x) = 1/L$ for $x \in [0, L]$ – see e.g., [26, p. 223].

We have

$$q(y; x) = \frac{\mathbf{1}_{\{y \leq x+r-L\}} + f(x, y) \mathbf{1}_{\{y \geq r-L, x < 2L-r\}}}{L}, \quad (1.9)$$

where

$$\begin{aligned} f(x, y) &= \frac{4}{\pi^2} \sum_{m \geq 1} \sum_{\substack{n \geq 1 \\ n \neq m}}^{\infty} \frac{n(a_{m,n} + b_{m,n} + c_{m,n})}{m^2 + n^2} \sin\left(\frac{m\pi(y-r+L)}{2L-r}\right) \\ &\quad + \frac{2}{\pi(2L-r)} \sum_{m \geq 1} \frac{d_m + e_m}{m} \sin\left(\frac{m\pi(y-r+L)}{2L-r}\right), \end{aligned}$$

and

$$\begin{aligned} a_{m,n} &= \frac{2m \sin(n\theta) - 2n \sin(m\theta)}{m^2 - n^2}, \\ b_{m,n} &= \frac{\sin((m-n)\pi + n\theta) + \sin((m-n)\pi - m\theta)}{m-n}, \\ c_{m,n} &= -\frac{\sin((m+n)\pi - n\theta) + \sin((m+n)\pi - m\theta)}{m+n}, \\ d_m &= 2(2L-r-x) \cos(m\theta) \\ e_m &= \frac{2L-r}{m\pi} \left(\sin(m\theta) - \sin(2m\pi - m\theta) \right), \quad \theta = \frac{\pi x}{2L-r}. \quad \diamond \end{aligned}$$

The proof of Proposition 1.3.1 is given in Appendix 1.C. We are now in a position to compute the expected transfer times $\mathbb{E}[T_i]$ for $i = 1, \dots, I-1$.

Define $f_i(x)$ ($0 \leq x \leq L$) as the pdf of $x_i(T_{i-1})$ for $i = 1, \dots, I-1$ (that is, $P(x_i(T_{i-1}) < y) = \int_0^y f_i(x) dx$). Note that $f_1(x) = 1/L$ for $x \in [0, L]$ thanks to the assumption that mobile \mathbf{X}_1 is in steady-state at time $t = 0$ (recall that $T_0 = 0$ by convention). Let us first compute $\mathbb{E}[T_1]$. We find

$$\begin{aligned} \mathbb{E}[T_1] &= \frac{1}{L^2} \int_0^L \int_0^L \mathbb{E}[T_1 | x_1(0) = x, x_2(0) = y] dx dy \\ &= \frac{1}{L^2} \int_{\{x+r < y+L\}} T_{2L,r}(x, y+L) dx dy, \end{aligned} \quad (1.10)$$

by using (1.8) and $T_{2L,r}(x, y+L) = 0$ if $x+r \leq y+L$. Similar to the derivation of (1.5) we get

$$\mathbb{E}[T_1] = \frac{64(2L-r)^4}{D\pi^6 L^2} C_0. \quad (1.11)$$

Next the computation of $\mathbb{E}[T_i]$ for $i = 2, \dots, I - 1$. We have

$$\mathbb{E}[T_i] = \mathbb{E}[T_{i-1}] + \frac{1}{L} \int_0^L \int_0^L \mathbb{E}[T_i - T_{i-1} | x_i(T_{i-1}) = x, x_{i+1}(T_{i-1}) = y] f_i(x) dx dy \quad (1.12)$$

$$= \mathbb{E}[T_{i-1}] + \frac{1}{L} \int_{\{x+r < y+L\}} T_{2L,r}(x, y+L) f_i(x) dx dy, \quad (1.13)$$

where we have used (1.8) to derive (1.13). To derive (1.12) we have used the fact that the position of mobile \mathbf{X}_{i+1} is uniformly distributed over its segment at time T_{i-1} (i.e., when the relay between mobiles \mathbf{X}_{i-1} and \mathbf{X}_i occurs), and that it is independent of the position of mobile \mathbf{X}_{i-1} at time T_{i-1} . It remains to evaluate the functions $f_i(x)$ for $i = 2, \dots, I - 1$. Differentiating in y on both sides of the identity

$$P(x_i(T_{i-1}) < y) = \int_0^L P(x_i(T_{i-1}) < y | x_{i-1}(T_{i-2}) = x) f_{i-1}(x) dx,$$

and then using Proposition 1.9, gives

$$f_i(y) = \int_0^L q(y; x) f_{i-1}(x) dx, \quad 0 \leq y \leq L, \quad (1.14)$$

for $i = 2, \dots, I - 1$. These results are summarized in the next proposition.

Proposition 1.3.2 (Expected relay times of one-dimensional Brownian motions)

The expected relay times $\mathbb{E}[T_i]$ for $i = 1, \dots, I - 1$, are given by (1.11) and (1.13), where the functions $f_i(x)$, $i = 2, \dots, I - 1$, satisfy the recursion (1.14) with $f_1(x) = 1/L$. In particular,

$$\mathbb{E}[T_1] = \frac{64(2L - r)^4}{D\pi^6 L^2} C_0. \quad \diamond$$

1.4 Numerical results and discussion

The expected first-meeting time $T_{L,r}(x_0, y_0)$ between two nodes is displayed in Figure 1.6 as a function of the initial position x_0 and y_0 of the mobiles, for $L = 30$, $r = 5$ and $D = 1/4$ (recall that D is the diffusion coefficient of the Brownian motions \mathbf{X} and \mathbf{Y}). The figure shows that the expected message delay grows (roughly) linearly as the initial distance between both mobiles increases and neither of the mobiles is near the boundaries of the interval $[0, L]$. We used (1.14) to determine the mapping $x \rightarrow f_2(x)$ for $0 \leq x \leq L$, the pdf of the location of mobile \mathbf{X}_2 when the relay with \mathbf{X}_1 occurs. This mapping is plotted in Figure 1.7 for different values of the starting position of mobile \mathbf{X}_1 ($x_1(0) = 5, 10, 15, 20$) and for $L = 30$, $r = 35$, $D = 1/4$. It is interesting to observe that $f_2(x)$ is uniform in

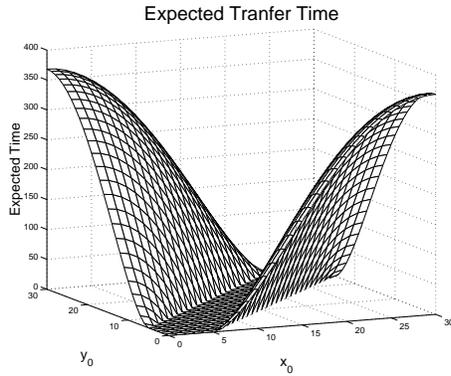


Figure 1.6: The mapping $(x_0, y_0) \rightarrow T_{L,r}(x_0, y_0)$ (expected first-meeting time between mobiles \mathbf{X} and \mathbf{Y} starting from x_0 and y_0 , respectively, at $t = 0$). See (1.2) for $L = 30$, $r = 5$, $D = 1/4$.

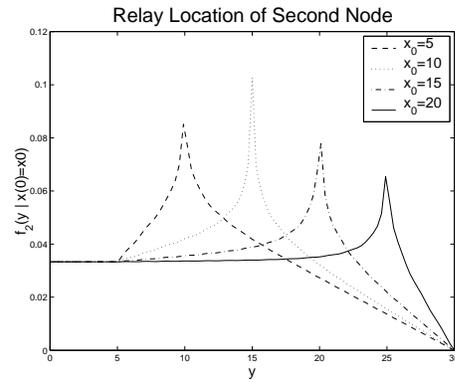


Figure 1.7: Mapping $x \rightarrow f_2(x)$ (pdf of location of mobile \mathbf{X}_2 at the relay epoch) for when mobile \mathbf{X}_1 is at position $x_0 \in \{5, 10, 15, 20\}$ at time $t = 0$, for $D = 1/4$, $L = 30$, $r = 35$.

$[0, x_1(0)]$. This is easily explained by the fact that if \mathbf{X}_2 is located in $[0, r - L]$ at time T_1 then it was necessarily located in this interval prior to time T_1 , since otherwise the relay would have occurred before T_1 . Each peak corresponds to the most likely value y in $[0, L]$ where mobile \mathbf{X}_2 will be located at time T_1 . This value is around $y = x_1(0) + r$.

Figure 1.8 displays mappings $x \rightarrow f_i(x)$ for $i \in \{2, 3, 100\}$ (evaluated from (1.14) with uniformly distributed initial locations). It is worth observing that these functions converge very rapidly (already $f_3(x)$ and $f_{100}(x)$ are extremely close to each other).

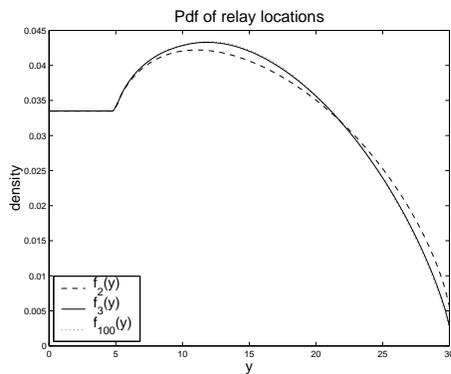


Figure 1.8: Mappings $x \rightarrow f_i(x)$ for $i \in \{2, 3, 100\}$ (pdf of starting location of mobiles) for $L = 30$, $r = 35$, and $D = 1/4$.

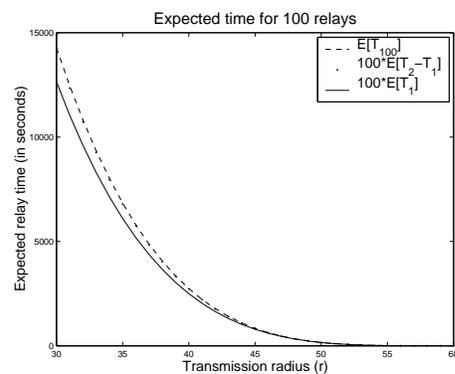


Figure 1.9: Comparison of mappings $r \rightarrow \mathbb{E}[T_{100}]$, $r \rightarrow 100 \times \mathbb{E}[T_2 - T_1]$, and $r \rightarrow 100 \times \mathbb{E}[T_1]$ for $L = 30$ and $D = 1/4$.

Figure 1.9 displays mappings $r \rightarrow \mathbb{E}[T_{100}]$, $r \rightarrow 100 \times \mathbb{E}[T_2 - T_1]$, and $r \rightarrow 100 \times \mathbb{E}[T_1]$. This figure carries two important messages. First, it shows for different values of the

transmission range r , that the approximation $\mathbb{E}[T_{100}] \sim 100 \times \mathbb{E}[T_2 - T_1]$ is very close to the exact result $\mathbb{E}[T_{100}]$ (derived from Proposition 1.3.2)), thereby suggesting the approximation

$$\mathbb{E}[T_i] \sim i \times \mathbb{E}[T_2 - T_1] \quad (1.15)$$

for the expected time to relay a message from mobile \mathbf{X}_1 to mobile \mathbf{X}_{i+1} . This approximation is based on the fact that the relay location convergences extremely rapidly and, with the exception of the first relay, the relay locations, and therefore also the relay times, of the consecutive relays are very similar. It has been verified that (1.15) is accurate for small values of i as well as for large values (i.e., larger than 100). Second, it shows that the approximation $\mathbb{E}[T_{100}] \sim 100 \times \mathbb{E}[T_1]$ may not be accurate for small transmission ranges, thereby ruling out the approximation $\mathbb{E}[T_i] \sim i \times \mathbb{E}[T_1]$. This is so because the latter approximation does not account for the fact that mobile \mathbf{X}_i does not start from a “uniform location” at time T_{i-1} (as opposed to mobile \mathbf{X}_1 whose position is uniformly distributed over $[0, L]$ at time $t = 0$).

1.5 Extensions to the model

In this paper the message delay over a one-dimensional network was analysed for mobiles which move as Brownian motions in adjacent segments. The extension of this theory to more than two mobiles per segment, or with leakage from one domain to the next, does not seem to be mathematically tractable. The reason for this is the following. If there are N mobiles in a segment then their positions needs to be mapped to a single N -dimensional Brownian motion. So far no problem. The model starts becoming more complex though when one has to take into account the positions of the N -dimensional Brownian motion which correspond to two nodes being in each others transmission range. For two one-dimensional Brownian motions this resulted in a diagonal line (see Figure 1.3). For three nodes this leads to three intersecting planes in a three-dimensional space. On top of this one has to take into account the (reflecting) border conditions for each of the mobiles. Although writing down these conditions is still feasible, the “real” problem lies in keeping track of which nodes have or have not received a copy of the message, the correlation between these nodes their positions, and finding an expression for the function which corresponds to the expected transfer time and which satisfies all of the necessary conditions.

A similar problem arises for N two-dimensional Brownian motions. In this case the positions of the mobiles can be mapped to a single $2N$ -dimensional Brownian motion. Since the area in which two nodes can communicate is given by a circle, it means that with the method of images the corresponding boundary conditions are no longer in the simple form of a square. Once again, finding a function which corresponds to the expected message delay, while keeping track of which mobile have a copy of the message, does not seem to be feasible.

As a side remark it is worth mentioning that the process which models the distance between two two-dimensional Brownian motions in free space (i.e., with no boundaries)

is known as a *Bessel process*. For this process various results exist, and, just as for two Brownian motions on an infinite line, the expected time until two Brownian motions in free space come within each others range is infinite. In this paper a bounded region was assumed to ensure a finite transfer time.

1.6 Concluding remarks

Besides message delay, the question of power control is also central in ad hoc networking. Ongoing research is concerned with determining the minimum transmission range that will ensure communication between mobiles (within a certain probability) before the battery power runs out, and with introducing utility functions into the model.

As soon as two nodes come within each others communication range there is the important issue of how long their *contact time* is. If these times are too short then a successful transfer of a message can not be guaranteed. Hence either the contact times or the probability of a successful transfer should be taken into account to more accurately reflect reality.

With a certain amount of overlap to this work, the next chapter considers the situation where instead of mobiles moving as Brownian motions they move as Random walkers over a discrete state space. This model also corresponds to messages being passed around a sensor network. The results presented there are verified through simulations and are in correspondence to the results presented in this chapter (in the limit for an infinite number of states).

As mentioned earlier, the problem with keeping track of which mobiles do or do not have a copy of the message in combination with their positions greatly increases the complexity and limits the extendibility of the theory presented in this paper. However, by assuming that $r \ll L$ and by working in (at least) two dimensions it has been found that the number of copies in the network can be decoupled from the positions of the copies. This greatly simplifies the analysis and has led to generic results and formulas given in Chapter 3 which hold under a variety of mobility models. Oddly enough, the theory developed there can not be used for the analysis of one-dimensional mobility models and hence there remains a need to study the one- and the two-dimensional settings separately.

Acknowledgments

This work presented in this chapter was partially supported by the EURO NGI network of excellence. I would also like to thank Marwan Krunz for stimulating discussions at the beginning of this work.

1.A The Brownian motion

A Brownian motion with zero drift is a stochastic process $\{x(t); t \geq 0\}$ with the properties

- Each increment $x(t+h) - x(t)$ is normally distributed with mean 0 and variance $2Dh$ for all $h > 0$. The constant D is known as the diffusion coefficient.³
- Non-overlapping time intervals are independent of each other.

Although the proof lies outside the scope of this thesis, we point out that if there is an absorbing barrier at x' then we must have $p(x', t) = 0$. If there is a reflecting barrier at x' then we must have $\left. \frac{\partial p_{x_0, x}(t)}{\partial x} \right|_{x'} = 0$. See [68] for the underlying theory,

If one is interested in the first-passage time to a boundary, then the behaviour of the Brownian motion *after* it has reached the boundary for the first time is *not* of influence to the distribution of the first-passage time to that boundary. This means that the boundary can be taken to be absorbing or reflecting as it will not change the first-passage time distribution. In the presence of an absorbing boundary, the first-passage time to the boundary can be calculated by taking the flux at the boundary, that is, the first-passage time is given by the space derivative of the concentration orthogonal to the boundary. See [68] for the motivation and proofs of this result.

The extension of the one-dimensional Brownian motion to a two-dimensional movement pattern is done by taking two independent one-dimensional Brownian motions and mapping each of these movements to each of the dimensions. Similarly, from a multi-dimensional Brownian motion we can extract a one-dimensional Brownian motion by looking at the movement in only one of the dimensions.

1.A.1 History of the Brownian motion

The Brownian motion was first “discovered” by the English botanist Robert Brown in 1827 and later developed, in a mathematical sense, by Einstein in 1905. Einstein’s treatment considers a small particle which is exposed to an enormous amount of molecular collisions. Although each collision has a negligible effect, the superposition of all collisions gives the particle a motion. This motion, a type of diffusion process, is the Brownian motion. The first rigorous mathematical formulation was given by Wiener in his 1918 dissertation and following papers and for this reason the Brownian motion is also often referred to as a *Wiener-process*. Brownian motion are also used in finance—in particular stock markets—to model the fluctuations of (stock) prices.

Other, more complicated, diffusion processes can also be used to model a movement (such as the Ornstein-Uhlenbeck process). As opposed to the random walk, a Brownian motion is a continuous time, continuous state space Markov process. This movement is in

³In the literature there is often a confusion in the exact definition of the diffusion coefficient D and the drift μ . The constant 2 can be part of, or it can be left out of the definition of the constants.

general constructed in either of two ways: either as a limiting form of a random walker or as a diffusion process.

1.A.2 Brownian motion as a random walker

First the approach starting from a random walker on an infinite state space. We will provide only the outline of the proof as it provides insight into the way a Brownian motion moves. A rigorous treatment is rather involved and can be found in the references mentioned at the end of this section.

The evolution of the occupation probability a random walker on an infinite state space is described by the master-equation

$$p_{x_0,x}(n+1) = p \cdot p_{x_0,x-1}(n) + q \cdot p_{x_0,x+1}(n). \quad (1.16)$$

In the following we will replace this discrete space and time process by a continuous space and time process. First we make the time continuous by letting each step take Δ_t units of time. During a time t the random walk jumps about t/Δ_t times. The master-equation then becomes

$$p_{x_0,x}(t + \delta_t) = p \cdot p_{x_0,x-1}(t) + q \cdot p_{x_0,x+1}(t).$$

By taking the Taylor expansion of this expression we obtain

$$p_{x_0,x}(t) + \delta_t \frac{\partial p_{x_0,x}(t)}{\partial t} + \sum_{k \geq 2} \frac{\delta_t^k}{k!} \frac{\partial^k p_{x_0,x}(t)}{\partial t^k} = p \cdot p_{x_0,x-1}(t) + q \cdot p_{x_0,x+1}(t).$$

Since the term δ_t will be sent to zero the summation will become negligible. We now have

$$\frac{\partial p_{x_0,x}(t)}{\partial t} = \frac{p}{\delta_t} \cdot p_{x_0,x-1}(t) + \frac{q}{\delta_t} \cdot p_{x_0,x+1}(t) - \frac{1}{\delta_t} \cdot p_{x_0,x}(t)$$

Next the discrete space will be “continuized⁴” by letting each step have length δ_x . By once again taking the Taylor expansion (but now on the spatial coordinate x), collecting terms, and “throwing away” all terms with a δ_x^k and k larger than three, we obtain the *convection-diffusion* equation

$$\frac{\partial p_{x_0,x}(t)}{\partial t} + \nu \frac{\partial p_{x_0,x}(t)}{\partial x} = D \frac{\partial^2 p_{x_0,x}(t)}{\partial x^2}, \quad (1.17)$$

where $\nu := (p-q)\delta_x/\delta_t$ is the *bias velocity* or *drift* and $D := (\delta_x)^2/2\delta_t$ is the *diffusion coefficient*. For the symmetric random walk, i.e., $p = q = 1/2$, the drift is equal to zero and the simpler *diffusion equation*

$$\frac{\partial p_{x_0,x}(t)}{\partial t} = D \frac{\partial^2 p_{x_0,x}(t)}{\partial x^2} \quad (1.18)$$

⁴Although this word officially does not exist, here it is used to mean “made continuous”.

is obtained. This Brownian motion is referred to as the (standard) Brownian motion with zero drift. Whereas for random walkers one talks about the occupation probability of a state, for Brownian motions one refers *concentration* in x .

The approach presented here is mainly heuristic in nature, for a truly rigorous treatment δ_t and δ_x must be sent to zero in the correct way. See for example [26, page 203] [53]. For more information about the relationship between the discrete hopping process and the continuum the reader is referred to [35], [85], and [82].

1.A.3 Brownian motion as a diffusion process

Small particles execute Brownian motions owing to collisions with the molecules in the gas or liquid in which they are suspended. Let $x(t)$ represent, at time t , the location of a particle along one axis (for example, along a horizontal axis). The concentration $p_{x_0,x}(t)$ then represent the probability density in x at time t for a particle starting in $x(0) = x_0$. Since $p_{x_0,x}(t)$ is a density we have the properties

$$p_{x_0,x}(t) \geq 0, \quad \int_{-\infty}^{\infty} p_{x_0,x}(t) dx = 1, \quad (1.19a)$$

for all $t \geq 0$. Furthermore, since at time $t = 0$ the particle starts in x_0 , we require

$$\lim_{t \rightarrow 0} p_{x_0,x}(t) = 0, \quad \text{for } x \neq x_0. \quad (1.19b)$$

From physical properties Einstein showed that $p_{x_0,x}(t)$ must satisfy (1.18), where D is given by $D = 2RT/Nf$. Here R is a gas constant, T is the temperature, N is Avogardo's number, and f is a coefficient of friction.

1.B Proof of Proposition 1.2.2

The density probability $q(x, t; u_0)$ that the Brownian motion $\{u(t), t \geq 0\}$ is in position $x \in (0, R)$ at time t , given that $u(0) = u_0$ and that the Brownian motion has not been absorbed up to time t , is [68, p. 255, formula (8.2.1)] [85, page 177]

$$w(x, t; u_0) = \frac{2}{R} \sum_{n \geq 1} e^{-(n\pi/R)^2 Dt} \sin\left(\frac{n\pi x}{R}\right) \sin\left(\frac{n\pi u_0}{R}\right).$$

Since $\{u(t), t \geq 0\}$ and $\{v(t), t \geq 0\}$ are independent and identical Brownian motions, we deduce from the above that the density probability $p(x, y, t; u_0, v_0)$ that the two-dimensional Brownian motion \mathbf{Z} is in position (x, y) at time t , without having hit one of the sides of the squares up to time t , is given by

$$p(x, y, t; u_0, v_0) = w(x, t; u_0) w(y, t; v_0). \quad 0 < x, y < R. \quad (1.20)$$

Conditioned on $z(0) = (u_0, v_0)$, the probability $S(t; u_0, v_0) = P(\tau_R > t)$ that the process has not hit the boundaries at time t (often called the survival probability [68]) is given by

$$S(t; u_0, v_0) = \int_0^R \int_0^R p(x, y, t; u_0, v_0) dx dy.$$

Therefore,

$$\begin{aligned} S(t; u_0, v_0) &= \int_0^R w(x, t; u_0) dx \int_0^R w(y, t; v_0) dy \\ &= \frac{4}{R^2} \sum_{m \geq 1} e^{-(m\pi/R)^2 Dt} \sin\left(\frac{m\pi u_0}{R}\right) \int_0^R \sin\left(\frac{m\pi x}{R}\right) dx \times \\ &\quad \sum_{n \geq 1} e^{-(n\pi/R)^2 Dt} \sin\left(\frac{n\pi v_0}{R}\right) \int_0^R \sin\left(\frac{n\pi y}{R}\right) dy \\ &= \frac{16}{\pi^2} \sum_{\substack{m \geq 1 \\ m \text{ odd}}} \sum_{\substack{n \geq 1 \\ n \text{ odd}}} \frac{\sin\left(\frac{m\pi u_0}{R}\right) \sin\left(\frac{n\pi v_0}{R}\right)}{mn} e^{-\frac{\pi^2}{R^2}(m^2+n^2)Dt}, \end{aligned} \quad (1.21)$$

where the uniform convergence of the series $w(x, t; \cdot)$ in $x \in [0, \infty)$ (because $|w(x, t; \cdot)| \leq 1/(1 - \exp(-(\pi/R)^2 Dt))$) allows one to interchange integral and summation signs in (1.21). Note that, as expected, $S(0; u_0, v_0) = 1$ since $\sum_{i \geq 1} \sin((2i-1)x)/(2i-1) = \pi/4$ for all x [37, Formula 1.442.1].

Finally,

$$\begin{aligned} \tau_R(u_0, v_0) &= \int_0^\infty S(t; u_0, v_0) dt \\ &= \frac{16}{\pi^2} \int_0^\infty \sum_{\substack{m \geq 1 \\ m \text{ odd}}} \sum_{\substack{n \geq 1 \\ n \text{ odd}}} \frac{\sin\left(\frac{m\pi u_0}{R}\right) \sin\left(\frac{n\pi v_0}{R}\right)}{mn} e^{-\frac{\pi^2}{R^2}(m^2+n^2)Dt} dt \end{aligned} \quad (1.22)$$

$$\begin{aligned} &= \frac{16}{\pi^2} \sum_{\substack{m \geq 1 \\ m \text{ odd}}} \sum_{\substack{n \geq 1 \\ n \text{ odd}}} \frac{\sin\left(\frac{m\pi u_0}{R}\right) \sin\left(\frac{n\pi v_0}{R}\right)}{mn} \int_0^\infty e^{-\frac{\pi^2}{R^2}(m^2+n^2)Dt} dt \\ &= \frac{16R^2}{D\pi^4} \sum_{\substack{m \geq 1 \\ m \text{ odd}}} \sum_{\substack{n \geq 1 \\ n \text{ odd}}} \frac{\sin\left(\frac{m\pi u_0}{R}\right) \sin\left(\frac{n\pi v_0}{R}\right)}{mn(m^2+n^2)}, \end{aligned} \quad (1.23)$$

where we have used the property that the series $S(t; \cdot, \cdot)$ is uniformly convergent in $[0, \infty)$ (since $S(t; \cdot, \cdot) \leq 1$ for all $t \geq 0$ by definition of $S(t; \cdot, \cdot)$) to interchange the summation and the integral signs in (1.22) to give (1.23). This concludes the proof. \blacksquare

1.C Proof of Proposition 1.3.1

Let $x(t)$ and $y(t)$ be the relative positions at time t of mobiles \mathbf{X} and \mathbf{Y} in $[0, L]$ and $[L, 2L]$, respectively. Let T the first time when $x(t) + r \geq y(t) + L$. Observe that $T = 0$ if $x(0) + r \geq y(0) + L$. We have

$$\begin{aligned}
 P(y(T) < y \mid x(0) = x_0) &= \frac{1}{L} \int_0^L P(y(T) < y \mid x(0) = x_0, y(0) = y_0) dy_0 \\
 &= \frac{1}{L} \int_0^L \mathbf{1}_{\{x_0+r \geq L+y_0\}} \mathbf{1}_{\{y > y_0\}} dy_0 \\
 &\quad + \frac{1}{L} \int_0^L \mathbf{1}_{\{x_0+r < L+y_0, y \geq r-L\}} P(y(T) < y \mid x(0) = x_0, y(0) = y_0) dy_0 \\
 &= \frac{1}{L} \min(x_0 + r - L, y) \\
 &\quad + \frac{1}{L} \mathbf{1}_{\{y \geq r-L, x_0 < 2L-r\}} \int_{x_0+r-L}^L P(y(T) < y \mid x(0) = x_0, y(0) = y_0) dy_0,
 \end{aligned}$$

where the indicator function $\mathbf{1}_{\{y \geq r-L\}}$ in the second integral in the second equality accounts for the fact that if the transfer does not take place at $t = 0$ (under the condition $x_0 + r < y + L$ then necessarily $T > 0$) then mobile \mathbf{Y} can not be located in $[L, L-r)$ at time T as otherwise the relay would have occurred before time T . Differentiating both sides of the above relation with regards to y gives

$$q(y; x_0) = \frac{1}{L} \mathbf{1}_{\{y \leq x_0+r-L\}} + \frac{1}{L} \mathbf{1}_{\{y \geq r-L, x_0 < 2L-r\}} \int_{x_0+r-L}^L g(y; x_0, y_0) dy_0, \quad (1.24)$$

with $g(y; x_0, y_0) := (\partial/\partial y)P(y(T) < y \mid x(0) = x_0, y(0) = y_0)$. It remains to evaluate $g(y; x_0, y_0)$. To this end, we will use again the method of images (see proof of Proposition 1.2.1).

Consider a square of size R by R , with $R = \sqrt{2}(2L - r)$, delimited by the (absorbing) boundaries $x' = 0$, $x' = R$, $y' = 0$ and $y' = R$. Starting from position (x'_0, y'_0) at time $t = 0$, the pdf $p(x', y', t; x'_0, y'_0)$ of the location of a two-dimensional Brownian motion at time t , given that the mobile has not been absorbed up to time t , is given by (see equation (1.20))

$$\begin{aligned}
 p(x', y', t; x'_0, y'_0) &= \frac{4}{R^2} \sum_{n \geq 1} \sum_{m \geq 1} e^{-(m^2+n^2)(\pi/R)^2 Dt} \times \\
 &\quad \sin\left(\frac{m\pi x'}{R}\right) \sin\left(\frac{n\pi y'}{R}\right) \sin\left(\frac{m\pi x'_0}{R}\right) \sin\left(\frac{n\pi y'_0}{R}\right).
 \end{aligned} \quad (1.25)$$

This expression will be used later on to derive the pdf of the location where the Brownian motion hits the side of the square for the first time.

Let $\xi(x', y'; x'_0, y'_0)$ ($0 \leq x', y', x'_0, y'_0 \leq R$), be the pdf of the absorption occurring at point (x', y') . Since we have applied the method of images we find that $g(y; x_0, y_0)$ is the sum of four of these components. Namely, with $x' = \sqrt{2}(y+L-r)$, it is the sum of the densities of hitting the side of the square $R \times R$ at the points $(x', 0)$, $(0, x')$, $(R-x', R)$, and $(R, R-x')$. With $x'_0 = (y_0+x_0+L-r)/\sqrt{2}$ and $y'_0 = (y_0-x_0+L-r)/\sqrt{2}$ this gives

$$g(y; x_0, y_0) = \xi(x', 0; x'_0, y'_0) + \xi(0, x'; x'_0, y'_0) + \xi(R-x', R; x'_0, y'_0) + \xi(R, R-x'; x'_0, y'_0).$$

Onward calculations can be simplified slightly by making use of symmetry arguments. Continuous rotation of the square by 90° means that each of the terms can be replaced by the density of the probability of hitting the side of the square at $(x', 0)$ while starting from, respectively, (x'_0, y'_0) , (y'_0, x'_0) , $(R-x'_0, R-y'_0)$, or $(R-y'_0, R-x'_0)$. This gives

$$\begin{aligned} g(y; x_0, y_0) = & \xi(x', 0; x'_0, y'_0) + \xi(x', 0; y'_0, x'_0) \\ & + \xi(x', 0; R-x'_0, R-y'_0) + \xi(x', 0; R-y'_0, R-x'_0). \end{aligned} \quad (1.26)$$

Note that although $\xi(x', 0; \cdot, \cdot)$ no longer contains y' , it still depends on y , x_0 , and y_0 through $x' = \sqrt{2}(y+L-r)$, $x'_0 = (y_0+x_0+L-r)/\sqrt{2}$, and $y'_0 = (y_0-x_0+L-r)/\sqrt{2}$. It remains to solve $\xi(x', 0; x'_0, y'_0)$ for any set of initial conditions (x'_0, y'_0) . We shall do this through the help of the first-passage probability of the point $(x', 0)$.

If $j(x', t)$ is the pdf of the first-passage probability of hitting the absorbing boundary of the square for the first time in the point $(x', 0)$ at time t , then naturally

$$\xi(x', 0; x'_0, y'_0) = \int_0^\infty j(x', t) dt, \quad (1.27)$$

since it is the probability density of hitting the boundary for the first time in $(x', 0)$ over all time.

It is known [68, p. 25, p. 45] that $j(x', t)$ is equal to the flux going out from the point $(x', 0)$, i.e.,

$$j(x', t) = D \frac{\partial p(x', y', t; x'_0, y'_0)}{\partial y'} \Big|_{y'=0},$$

with $p(x', y', t; x'_0, y'_0)$ the pdf of the location of the Brownian motion at time t given by equation (1.25). Combining this with (1.27) gives

$$\begin{aligned} \xi(x', 0; x'_0, y'_0) = & D \int_0^\infty \frac{\partial p(x', y', t; x'_0, y'_0)}{\partial y'} \Big|_{y'=0} dt \\ = & \frac{4}{R\pi} \sum_{n \geq 1} \sum_{m \geq 1} \frac{n}{m^2 + n^2} \sin\left(\frac{m\pi x'}{R}\right) \sin\left(\frac{m\pi x'_0}{R}\right) \sin\left(\frac{n\pi y'_0}{R}\right). \end{aligned} \quad (1.28)$$

Finally, plugging (1.26) and (1.28) into (1.24) yields (1.9) after some tedious algebra. \blacksquare

Message Delay for One-dimensional Random Walkers

Contents

2.1	Two random walkers jumping on a finite state space	32
2.2	A chain of relaying nodes	38
2.3	Extensions to the model	43
2.4	Concluding remarks	44
2.A	The stationary distribution of a random walker	45
2.B	Proof of Proposition 2.1.1	45
2.C	Proof of Proposition 2.1.2	49
2.D	Proof of Proposition 2.2.1	60

Mobile ad hoc networks are characterized by a lack of a fixed infrastructure and by node mobility. In these networks key factors for the performance of the network are the node mobility and the transmission power. In addition to these two factors, relaying also plays an important role in determining the time required to transfer a message between mobiles. In this chapter we study these three elements through the analysis of a simple one-dimensional ad hoc network topology. Nodes move as random walkers and communication can take if they are within a certain number of states from each other. Each mobile remains an arbitrary amount of time in a state after which it possibly jumps to one of the neighbouring states. Closed-form expressions are derived for the expected time until two nodes communicate directly with each other, departing from either given initial positions or from a stationary regime. Although the resulting expressions are rather complex, a simple and accurate approximation is derived which holds for an arbitrary number of states, communication ranges, and sojourn time distributions at each state. Finally, these results are

extended to obtain the message delay when relaying a message across a sequence of mobiles. The results presented here can also be used for two messages travelling through a sensor network. In addition, in the limit (for the number of states), the Random Walker turn into Brownian motions and all of the results converge to those presented in Chapter 1 or in [41].

The rest of this chapter is organised as follows. In Section 2.1 the situation is considered where two nodes move as two independent random walkers over the same set of states (see Figure 2.1). We are interested in computing the expected time until both nodes come within a certain number of states from each other. This quantity is derived for any given initial locations (Proposition 2.1.1) as well as for the case where each random walk is assumed to start from steady-state (Proposition 2.1.2). It is known (see the assumption after Proposition 2.1.1) that this assumption implies that both nodes are uniformly distributed over the segment. The uniform spatial distribution over the coverage area has attracted attention lately and several fundamental results [45, 34, 8] have been obtained in this setting.

These results are then used and extended in Section 2.2 where a message is relayed across a chain of nodes. Numerical results are given in section 2.2.1. Finally, concluding remarks are given in Section 2.4.

2.1 Two random walkers jumping on a finite state space

We consider two mobiles (henceforth called *nodes* or *random walkers*) \mathbf{X} and \mathbf{Y} jumping over $L > 1$ points, labelled $0, 1, \dots, L - 1$. Both node are assumed to move independently of each other according to the following mobility pattern: each mobile visits state $i \in \{0, 1, \dots, L-1\}$ for a random duration (called the *visit* time) and then jumps instantaneously to one of the two neighbouring states with equal probability if $i \in \{1, 2, \dots, L-2\}$; if $i = 0$ (resp. $i = L - 1$) then the mobile either stays in state 0 (resp. $L - 1$) or it jumps to state 1 (resp. $L - 2$) with equal probability. After this the process starts over again but from the new state. See Figure 2.1.

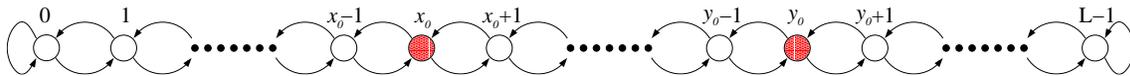


Figure 2.1: Two nodes (at x_0 and y_0) represented as two random walker on L states.

We assume that the visit times at the states are independent and identically distributed random variables with finite mean $\mathbb{E}[S]$. In particular, the cases will be considered where the visit times are either all exponentially distributed or all constant. Let T_n denote the time when the n^{th} jump (of either mobile) occurs which gives $\{T_n\}_n$ as a Poisson process with intensity $2/\mathbb{E}[S]$. In the latter case (constant visit times) we assume that the mobiles jump at exactly the same time¹ so that $T_{2n-1} = T_{2n} = n\mathbb{E}[S]$. In both cases $T_0 = 0$.

¹The analysis of two mobiles jumping at different times is messy due to need of keeping track which mobile jumped last. If nodes are assumed to jump at the same time all the results will be almost the same.

Let $x(n), y(n) \in \{0, 1, \dots, L-1\}$, respectively, be the state nodes \mathbf{X} and \mathbf{Y} are in just after the n^{th} jump (i.e., at time T_n+) and let $x_0 := x(0)$ and $y_0 := y(0)$.

Nodes \mathbf{X} and \mathbf{Y} can communicate as long as they are not separated by more than r states. More precisely, if we denote by $N_{L,r}$ the number of jumps until the first communication takes place, then

$$N_{L,r} = \inf\{n \in \mathbb{Z}^* : |y(n) - x(n)| \leq r\}, \quad \text{with } \mathbb{Z}^* := \{0, 1, 2, \dots\} \quad (2.1)$$

Note that $N_{L,r} = 0$ if $|y(0) - x(0)| \leq r$.

Our first objective in this section is to compute

$$N_{L,r}(x_0, y_0) := \mathbb{E}[N_{L,r} | x(0) = x_0, y(0) = y_0], \quad x_0, y_0 \in \{0, 1, \dots, L-1\}, \quad (2.2)$$

the conditional expected number of jumps required until both nodes can communicate. From this the unconditional expected number of jumps, $\mathbb{E}[N_{L,r}]$, will be derived as well as the expected time, $\mathbb{E}[T_{N_{L,r}}]$, until the two nodes can transfer a message between each other. The following proposition gives the conditional expected number of jumps.

Proposition 2.1.1 (Expected number of jumps with given initial positions)

Let $x_0, y_0, r \in \{0, 1, \dots, L-1\}$ and $x_0 + r < y_0$ (the case $x_0 - r > y_0$ can be solved by interchanging x_0 and y_0). The expected number of jumps for the random walkers \mathbf{X} and \mathbf{Y} starting in x_0 and y_0 to come within r states from each other is given by

$$N_{L,r}(x_0, y_0) = \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{F_j F_k g_{jk}(x_0, y_0)}{1 - \cos(j\theta) \cos(k\theta)} \quad (2.3a)$$

in the case of exponential visit times and by

$$N_{L,r}(x_0, y_0) = \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{2F_j F_k g_{jk}(x_0, y_0)}{1 - (\cos(2j\theta) + \cos(2k\theta))/2} \quad (2.3b)$$

in the case of deterministic visit times. Here

$$R := 2(L - r), \quad F_j := \frac{2 \sin(j\theta)}{R(1 - \cos(j\theta))},$$

$$\theta := \frac{\pi}{R} = \frac{\pi}{2(L - r)}, \quad g_{jk}(x_0, y_0) := \sin\left(\frac{\pi j(y_0 - r + x_0 + 1)}{2(L - r)}\right) \sin\left(\frac{\pi k(y_0 - r - x_0)}{2(L - r)}\right).$$

The proof is given Appendix 2.B. ◇

We have not succeeded in simplifying equation (2.3) any further. As an example, the conditional expected number of jumps $N_{L,r}(x_0, y_0)$ as a function of x_0 and y_0 is displayed in Figure 2.2 for exponential visit times and $L = 30$ states and $r = 5$. Deterministic visit times gives an identical figure. The figure shows that the expected number of jumps grows (roughly) linearly as the initial distance between both nodes increases and neither of the nodes is near the outer states of the segment.

As Proposition 2.1.1 gives the conditional expected number of jumps, the next question that naturally arises is what the unconditional expected number of jumps could be. To answer this question, we assume that the random walkers \mathbf{X} and \mathbf{Y} are in steady-state at jump $n = 0$. This assumption implies² in particular, that the position of each random walker at jump $n = 0$ is uniformly distributed, i.e., the *probability density function* (pdf) of $x(0)$ or $y(0)$ is uniform over $\{0, \dots, L-1\}$. The same holds of course at any arbitrary time (i.e., the pdf of $x(n)$ or $y(n)$ is uniform over $\{0, \dots, L-1\}$ if n is arbitrary). With this assumption we are now in a position to give the unconditional expected number of jumps needed for two nodes to come within communication range of each other.

Proposition 2.1.2 (Expected number of jumps for uniform initial positions)

Let $r \in \{0, 1, \dots, L-1\}$ and assume that two random walkers \mathbf{X} and \mathbf{Y} are in steady-state at jump $n = 0$. The expected number of jumps, $\mathbb{E}[N_{L,r}]$, in the case of exponential visit times is

$$\mathbb{E}[N_{L,r}] = \frac{4}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{G_j G_k \cos(k\theta)}{1 - \cos(j\theta) \cos(k\theta)}, \quad (2.4a)$$

whereas in the case of deterministic visit times

$$\mathbb{E}[N_{L,r}] = \frac{8}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{G_j G_k \cos(k\theta)}{1 - (\cos(2j\theta) + \cos(2k\theta))/2}. \quad (2.4b)$$

Here

$$R := 2(L - r), \quad \theta := \frac{\pi}{R} = \frac{\pi}{2(L - r)}, \quad G_j := \frac{1}{1 - \cos(j\theta)}.$$

Furthermore, there is an insensitivity property³ towards the underlying visit time distribution since for both exponential and deterministic visit times⁴

$$\mathbb{E}[N_{L,r}] = \frac{512(L - r)^4 C}{\pi^6 L^2} + \mathcal{O}\left(\frac{(L - r)^2}{L^2}\right), \quad (2.5)$$

²See Section 2.A for the derivation.

³It is interesting to point out that the limiting form of the random walkers, two Brownian motions moving on a line, gives an expression [41, Proposition 3] similar to equation (2.5) up to a constant factor ($D = 1/4$) difference which accounts for the transition from the discrete to the continuous state space.

⁴One writes $f(N) = \mathcal{O}(g(N))$ if $|f(N)/g(N)|$ is bounded from above as $N \rightarrow \infty$.

where

$$C := \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 k^2 (j^2 + k^2)} \approx 0.5279.$$

In both cases a simple lower bound is given by

$$\frac{512(L-r)^4 C}{\pi^6 L^2} - \frac{(2 + \pi^2)(L-r)^2}{2\pi^2 L^2} \leq \mathbb{E}[N_{L,r}] \quad (2.6)$$

◇

The proof is forwarded to Appendix 2.C. Expressions for the upper bounds can also be obtained (they are given in Appendix 2.C) but they are rather lengthy and have therefore been omitted here as they do not provide additional information.

Examples of equations (2.4a) and (2.4b) are given in Figure 2.3 for $L = 25$ states and varying transmission ranges. Also shown in the figure are simulated values of the expected number of jumps for a range of different visit time distributions, along with the 90% confidence intervals of the simulations. Each point in the figures is based on $n = 5.000.000$ runs to estimate the expected number of jumps. The confidence interval around each simulated mean is $[\bar{x} \pm \frac{\sigma}{\sqrt{n}} \Phi(0.95)]$, where $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ is the simulated expected number of jumps (x_i is the number of jumps needed in the i^{th} simulation run) and $\sigma := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ is the simulation variance. Here $\Phi(0.95) \approx 1.645$ is the 95 percentile of the normal distribution function.

Striking is the fact that in all cases the theoretical and the simulated expected number of jumps are in such close vicinity to one another that they can hardly be distinguished from one another. This is due to an insensitivity towards the underlying visit time distribution. The reason for this, we argue, is that although directly after the jump of one node the elapsed visit time of the other node may not be ignored (unless it is exponential), after a number of jumps the expected number of jumps of each of the two nodes become more or less independent of each other. Putting it differently, let W_n be a sequence of numbers such that $W_n = 0$ if the first node jumps at jump n and $W_n = 1$ if the second node jumps at jump n . Then, unless the visit times are exponential, the sequences of zeros and ones are correlated. However, as the number of steps increases, the mean number of zeros becomes more independent of the order of the zeros and ones. As proposition 2.1.2 presents the expected number of jumps, regardless of the order in which the nodes jump, it is to be expected that the expected number of jumps is more or less (but not completely) independent of the visit time distribution.

Another interesting feature of Figure 2.3 is that equations (2.4a), (2.4b), and $\frac{32R^4}{\pi^6 L^2}$ can be seen to lie extremely close to each other, indicating that the $\mathcal{O}(\cdot)$ term in (2.5) is relatively small (even for $L = 25$). Taking different values of L and r have continuously led to the same conclusions.

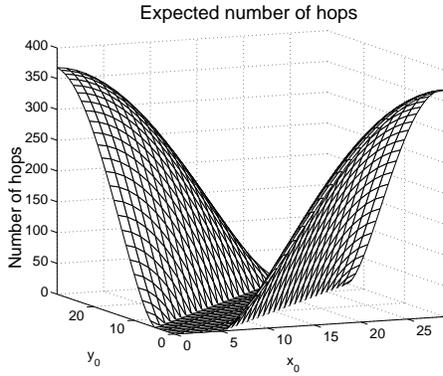


Figure 2.2: Expected number of jumps $N_{30,5}(x_0, y_0)$ (see Proposition 2.1.1) for random walkers \mathbf{X} and \mathbf{Y} to come within r states from each other, starting at x_0 and y_0 at jump $n = 0$.

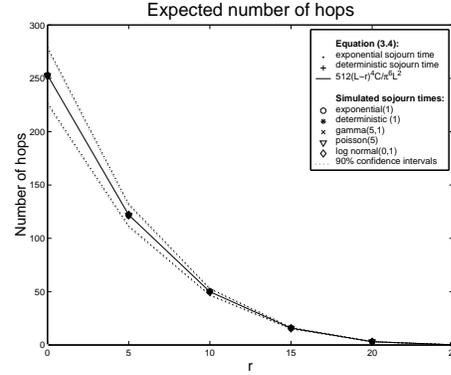


Figure 2.3: Expected number of jumps for different radii, given uniform starting positions and $L = 25$ states.

To get an indication of the estimates given in equations (2.5) and (2.6) take a look at Figure 2.4. Shown are, for exponential visit times (deterministic visit times gives an identical figure), the function $\frac{32R^4}{\pi^6 L^2}$, the lower bound given by (2.6), and the exact values calculated with (2.4). As can be seen, even for small values of L and r , the exact expected number of jumps for deterministic and exponential visit times, and the function $\frac{32R^4}{\pi^6 L^2} C$ lie extremely close to one another. Also, as L increases we see that the lower bound gets (relatively) closer to the exact values.

As a note it must be observed that although they are all extremely close, the expected number of jumps for exponential and deterministic visit time are not identical, nor is either of them exactly equal to $\frac{32R^4}{\pi^6 L^2} C$.

Proposition 2.1.2 gave the expected number of jumps until the two nodes are within communication range of each other. The following proposition extends this result to the expected time for this to happen.

Proposition 2.1.3 (Expected transfer time for uniform initial positions)

Let two random walkers jump on the same L states and let them start from a steady-state distribution. For exponentially and deterministically distributed visit times with mean $\mathbb{E}[S]$, the expected transfer time $\mathbb{E}[T_{N_{L,r}}]$ is

$$\mathbb{E}[T_{N_{L,r}}] = \frac{\mathbb{E}[S]}{2} \mathbb{E}[N_{L,r}], \quad (2.7)$$

where $\mathbb{E}[N_{L,r}]$ is given in Proposition 2.1.2. For both exponentially or deterministically

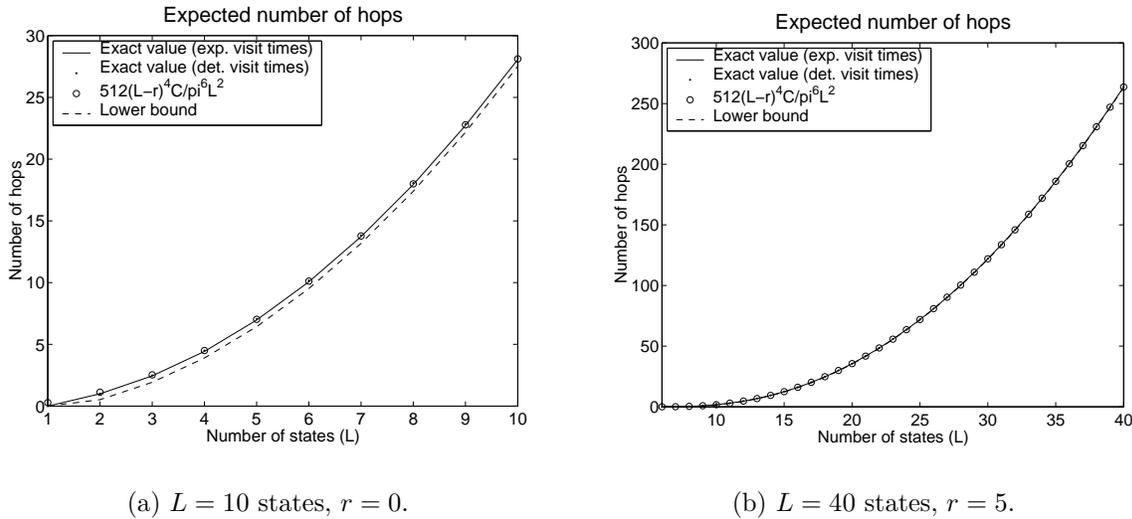


Figure 2.4: Expected number of jumps (exact values, approximation, and lower bound) until two random walkers on L states are within r states from each other .

distributed visit times we have, with the help of (2.5),

$$\mathbb{E}[T_{N_{L,r}}] = \frac{256(L-r)^4 C \mathbb{E}[S]}{\pi^6 L^2} + \mathbb{E}[S] \cdot \mathcal{O}\left(\frac{(L-r)^2}{L^2}\right) \quad (2.8)$$

Just as in Proposition 2.1.2, we see that there is an insensitivity property towards the underlying visit time distribution. \diamond

Proof. We start by considering exponential visit times. Let the visit times, S_1 and S_2 , of the two nodes be independent and exponentially distributed with mean $\mathbb{E}[S]$. Since $T_n - T_{n-1}$ is the visit time between the $(n-1)^{th}$ and the n^{th} jump, it follows that $T_n - T_{n-1}$ is distributed as an exponential variable with mean $S/2$. To see why $T_n - T_{n-1} \sim \exp(2/\mathbb{E}[S])$, first realize that by definition $T_1 - T_0 = \min(S_1, S_2)$. Then

$$P(T_1 - T_0 > t) = P(\min(S_1, S_2) > t) = P(S_1 > t; S_2 > t) = P(S_1 > t)P(S_2 > t) = e^{-2t/\mathbb{E}[S]}.$$

The visit times for both nodes ‘start over’ after each jump—due to the memoryless property of the exponential distribution—and therefore $T_n - T_{n-1} \sim T_1 - T_0$. This gives the expected transfer time as

$$\mathbb{E}[T_{N_{L,r}}] = \mathbb{E}\left[\sum_{i=1}^{N_{L,r}} T_i - T_{i-1}\right] = \mathbb{E}[T_1 - T_0] \mathbb{E}[N_{L,r}] = \frac{\mathbb{E}[S]}{2} \mathbb{E}[N_{L,r}].$$

In the case of deterministic visit times of duration $\mathbb{E}[S]$, the visit time $T_n - T_{n-1}$ between the $(n-1)^{th}$ and the n^{th} jump is either of duration $\mathbb{E}[S]$ if n is odd and zero if n is even

(since the nodes were assumed to jump at the same time). This results in

$$\mathbb{E}[T_{N_{L,r}}] = \mathbb{E}\left[\sum_{i=1}^{N_{L,r}} T_i - T_{i-1}\right] = \mathbb{E}[T_1 - T_0] \frac{\mathbb{E}[N_{L,r}]}{2} = \frac{\mathbb{E}[S]}{2} \mathbb{E}[N_{L,r}]. \quad \blacksquare$$

2.2 A chain of relaying nodes

In the previous section we considered two nodes moving on the same state space. Next consider the situation depicted in Figure 2.5. There are I adjacent segments, each with L states, and there is a single node per segment. We denote by \mathbf{X}_i the node in segment i ($i \in \{1, \dots, I\}$). Let $x_i(n) \in \{0, \dots, L-1\}$ be the *relative* position of the i^{th} node in its segment. We assume that the process $\mathbf{X}_i = \{x_i(n), n \in \mathbb{Z}^*\}$ is a random walker with equal probability of jumping to the right or the left and that the random walker remains a random amount of time (the *visit* time) in each state. Whenever a random walker is in a state at the edge of its segment, then once the visit time has expired the random walker jumps to the same state or to the neighbouring state (within its segment) with equal probability. Additionally, assume that $\mathbf{X}_1, \dots, \mathbf{X}_I$ are mutually independent processes.

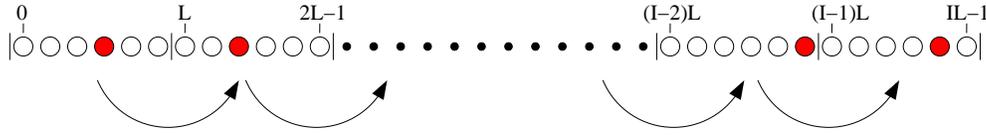


Figure 2.5: There are I segments with a random walker jumping in each segment.

Let $M_1 = \inf\{n \in \mathbb{Z}^* : x_1(n) + r \geq L - 1 + x_2(n)\}$ be the expected number of jumps until \mathbf{X}_1 and \mathbf{X}_2 are located at most r states from each other for the first time. The expected number of jumps $M_2 \leq \dots \leq M_{I-1}$ between nodes \mathbf{X}_2 and $\mathbf{X}_3, \dots, \mathbf{X}_{I-1}$ and \mathbf{X}_I , respectively, are recursively defined by

$$M_i = \inf\{n \geq M_{i-1} : x_i(n) + r \geq L - 1 + x_{i+1}(n)\}, \quad i = 2, \dots, I - 1, \quad n \in \mathbb{Z}^*. \quad (2.9)$$

Our objective in this section is to compute $\mathbb{E}[M_i]$, the expected number of jumps needed to transfer a message from the first node to the $(i+1)^{\text{th}}$ node, for $i = 1, \dots, I-1$.

Throughout this section it will be assumed that $L \leq r \leq 2L$. This assumption is made for the sake of mathematical tractability. Indeed, a few seconds of reflection will convince the reader that when⁵ $L \leq r \leq 2L$ and $(x_1(0), x_2(0)) = (x_0, y_0)$ the expected number of jumps needed to transfer a message between two adjacent segments is the same as $N_{2L,r}(x_0, y_0 + L)$, the expected number of jumps obtained in Section 2.1 for a segment with $2L$ states and with

⁵When $L \leq r \leq 2L$ the reflecting boundaries conditions are at $x_1(t) = 0$ and at $x_2(t) = 2L$. For $0 \leq r \leq L$ there are two additional reflecting boundary conditions at $x_1(t) = L$ and at $x_2(t) = L$ which lead to a much more difficult problem which we were not able to solve.

the given initial conditions. This observation allows us to find at once the expected number of jumps between nodes \mathbf{X}_1 and \mathbf{X}_2 for any initial conditions $x_1(0)$ and $x_2(0)$. We find

$$\mathbb{E}[M_1 | x_1(0) = x, x_2(0) = y] = N_{2L,r}(x, y + L). \quad (2.10)$$

The difficulty arises when trying to find the expected number of jumps between nodes \mathbf{X}_i and \mathbf{X}_{i+1} for $i = 2, \dots, I - 1$, since the position of \mathbf{X}_i right after the transfer between \mathbf{X}_{i-1} and \mathbf{X}_i takes place is not uniform.

To overcome this difficulty, we assume that the random walkers $\mathbf{X}_1, \dots, \mathbf{X}_I$ are all in steady-state at jump $n = 0$. This assumption implies, in particular, that the position of each node at jump $n = 0$ is uniformly distributed over its segment (i.e., the pdf of $x_i(0)$ is uniform over $\{0, \dots, L-1\}$). The same holds at any arbitrary time, i.e., the pdf of $x_i(n)$ is uniform over $\{0, \dots, L-1\}$ if n is arbitrary.

Another consequence of this assumption is that the position of node \mathbf{X}_{i+1} at jump M_{i-1} (i.e., when \mathbf{X}_i receives a message from \mathbf{X}_{i-1}) is uniformly distributed over $\{0, \dots, L-1\}$. This property will be used later on.

Proposition 2.2.1 below addresses the location of a node at the moment when a relay occurs. For later reference, the result is stated in a general form. Consider two adjacent segment, each with L states, with a single node in each segment (node \mathbf{X} in the first segment and \mathbf{Y} in the second segment). Both nodes move in their segment as independent random walkers with equal probability of jumping in either direction (and at the outermost states the random walker can once again jump back to the same state). Let the random walker representing the movement of \mathbf{Y} be in steady state at jump $n = 0$. As usual, a transfer can occur the first time both nodes are within r states from one another.

Proposition 2.2.1 (Pdf of location at relay epoch for exponential visit times)

Fix $L \leq r \leq 2L$. Let $q(x, y)$ be the pdf of the (relative) position $y \in \{0, \dots, L-1\}$ of node \mathbf{Y} at the relay epoch, given that at jump $n = 0$ node \mathbf{X} is at position $x \in \{0, \dots, L-1\}$ and that the position of node \mathbf{Y} is uniform. Let N be the number of jumps until the relay can take place. The function $q(x, y)$ can be calculated through a series of functions.

We have

$$\begin{aligned} q(x_0, y) &= P(y(N) = y | x(0) = x_0) \\ &= \frac{\mathbf{1}_{\{y \leq x_0 + r - L + 1\}}}{L} + \frac{\mathbf{1}_{\{x_0 + r \leq 2(L-1), y \geq r - L + 1\}}}{L} \sum_{y_0 = x_0 + r - L + 1}^{L-1} g(y; x_0, y_0), \end{aligned} \quad (2.11a)$$

where

$$g(y; x_0, y_0) = (P_{\dot{x}-1, y} + P_{\dot{x}, y+1})/4$$

and $\dot{x} := y - r + (L - 1)$. Finally, the function P_{xy} is calculated through

$$P_{xy} = \tilde{P}_{xy} + \mathbf{1}_{\{x \geq 1\}} \tilde{P}_{-x, y} + \mathbf{1}_{\{y \leq L-2\}} \tilde{P}_{x, 2(L-1)-y} + \mathbf{1}_{\{x \geq 1; y \leq L-2\}} \tilde{P}_{-x, 2(L-1)-y}. \quad (2.11b)$$

where

$$\tilde{P}_{xy} = \frac{4}{\dot{R}^2} \sum_{j=1}^{\dot{R}-1} \sum_{k=1}^{\dot{R}-1} \frac{\dot{h}_{jk}(x, y) \dot{h}_{jk}(x_0, y_0)}{1 - \cos(j\dot{\theta}) \cos(k\dot{\theta})} \quad (2.11c)$$

and

$$\begin{aligned} \dot{r} &= r - (L-1), & \dot{R} &= 2(L-1-\dot{r}) = 2(L-1) - r, & \dot{\theta} &= \frac{\pi}{\dot{R}} = \frac{\pi}{2(L-1) - r}, \\ \dot{h}_{jk}(x, y) &:= \sin(j\dot{\theta}(y-\dot{r}+x)) \sin(k\dot{\theta}(y-\dot{r}-x)), & & & \dot{r}+x &\leq y \leq L-1. \quad \diamond \end{aligned}$$

The proof of Proposition 2.2.1 is given in Appendix 2.D. We are now in a position to compute the expected number of jumps $\mathbb{E}[M_i]$ needed to transfer a message from the first to the i^{th} node, $i = 2, \dots, I$. Conditioning on the independent positions of the nodes in the i^{th} and $(i+1)^{\text{th}}$ segment at jump M_{i-1} gives

$$\begin{aligned} \mathbb{E}[M_i - M_{i-1}] &= \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} \mathbb{E} \left[M_i - M_{i-1} | x_i(M_{i-1}) = x; x_{i+1}(M_{i-1}) = y \right] \\ &\quad \cdot P(x_i(M_{i-1}) = x) \cdot P(x_{i+1}(M_{i-1}) = y). \end{aligned}$$

In steady state $P(x_{i+1}(M_{i-1}) = y) = 1/L$ and $\mathbb{E}[M_i - M_{i-1} | x_i(M_{i-1}) = x; x_{i+1}(M_{i-1}) = y]$ depends only on the positions of the two nodes and not on i . From (2.10) we have $\mathbb{E}[M_i - M_{i-1} | x_i(M_{i-1}) = x; x_{i+1}(M_{i-1}) = y] = N_{2L,r}(x, y+L)$. This gives

$$\mathbb{E}[M_i - M_{i-1}] = \frac{1}{L} \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} N_{2L,r}(x, y+L) \cdot P(x_i(M_{i-1}) = x). \quad (2.12)$$

The last term on the right hand side can be found by conditioning on the position of the node one step earlier. This gives the recursive relationship

$$P(x_i(M_{i-1}) = y) = \sum_{x=0}^{L-1} q(x, y) \cdot P(x_{i-1}(M_{i-2}) = x), \quad \begin{aligned} y &\in \{0, \dots, L-1\}, \\ i &= 2, \dots, I-1. \end{aligned} \quad (2.13)$$

These last couple of results are ideally suited for matrix implementation and are summarized in the next proposition.

Proposition 2.2.2 (Distribution relay location and expected number of jumps)

Let the visit time be exponentially distributed and let \mathbf{Q} and \mathbf{N} be matrixes with entry (x, y)

equal to $q(x, y)$ and $N_{2L,r}(x, y)$ respectively, i.e.,

$$\mathbf{Q} = \begin{pmatrix} q(0, 0) & \cdots & q(L-1, 0) \\ \vdots & \ddots & \vdots \\ q(0, L-1) & \cdots & q(L-1, L-1) \end{pmatrix}, \quad (2.14a)$$

$$\mathbf{N} = \begin{pmatrix} N_{2L,r}(0, L) & \cdots & N_{2L,r}(0, 2L-1) \\ \vdots & \ddots & \vdots \\ N_{2L,r}(L-1, L) & \cdots & N_{2L,r}(L-1, 2L-1) \end{pmatrix}, \quad (2.14b)$$

where $q(\cdot, \cdot)$ and $N_{2L,r}(\cdot, \cdot)$ can be calculated with the help of Proposition 2.2.1 and Proposition 2.1.1, respectively. Let $\mathbf{1}$ be a column vector of ones and define $\boldsymbol{\xi}_y$ as a column vector filled with zeros except for the y^{th} element which is equal to one, $y \in \{0, \dots, L-1\}$. Let \mathbf{x}^T be the transpose of a vector \mathbf{x} .

The pdf of the location where the i^{th} node receives the message is

$$P(x_i(M_{i-1}) = y) = \frac{1}{L} \mathbf{1}^T \mathbf{Q}^{i-1} \boldsymbol{\xi}_y, \quad y \in \{0, \dots, L-1\}, \quad i \in \{1, 2, \dots, I\}. \quad (2.15)$$

The expected number of jumps for the i^{th} transfer is given by

$$\mathbb{E}[M_i - M_{i-1}] = \frac{1}{L^2} \mathbf{1}^T \mathbf{Q}^{i-1} \mathbf{N} \mathbf{1}, \quad (2.16)$$

which leads to the expected number of jumps to transfer a message over I segments:

$$\mathbb{E}[M_{I-1}] = \frac{1}{L^2} \mathbf{1}^T \left(\sum_{i=0}^{I-2} \mathbf{Q}^i \right) \mathbf{N} \mathbf{1}. \quad (2.17)$$

The expected transfer time needed for this is given by

$$\mathbb{E}[T] = \frac{\mathbb{E}[S]}{2} \mathbb{E}[M_{I-1}]. \quad (2.18)$$

The location where the node receives the message converges rapidly to a distribution $\boldsymbol{\pi}$ which satisfies

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{Q} \quad (2.19)$$

This gives

$$\boldsymbol{\pi} = \mathbf{1}^T (\mathbf{Q} + \mathbf{E} - \mathbf{I})^{-1} \quad (2.20)$$

where \mathbf{E} is an $L \times L$ matrix of ones and \mathbf{I} is an $L \times L$ matrix of ones on the diagonal and zeros elsewhere. Assuming convergence, the average number of jumps for one relay is

$$\mathbb{E}[M_\infty - M_{\infty-1}] = \frac{1}{L} \boldsymbol{\pi} \mathbf{M} \mathbf{1}. \quad \diamond$$

Proof. By iterating (2.13) with $P(x_1(0) = x) = 1/L$, we see that (2.15) immediately follows. By definition of matrix multiplications (2.16) follows from (2.12) and (2.15). The rest of the equations follow from standard matrix manipulation arguments. ■

Once the matrixes in (2.14) have been generated, a number of performance measures can easily be calculated. This will be discussed in the next section.

2.2.1 Numerical results for a chain of nodes

We have used Proposition 2.2.1 to determine the mapping $y \rightarrow P(x_2(M_1) = y)$, the pdf of the location of node \mathbf{X}_2 when the relay with \mathbf{X}_1 occurs, $y \in \{0, \dots, L - 1\}$. This mapping is plotted in Figure 2.6 for different values of the starting position of node \mathbf{X}_1 ($x_1(0) = 5, 10, 15, 20$) and for $L = 30$ states per segment and $r = 35$. It is interesting to observe that $P(x_2(M_1) = y)$ is uniform in $[0, x_1(0)]$. This is easily explained by the fact that if \mathbf{X}_2 is located in $\{0, \dots, r - L\}$, then the transfer took place at time zero. The pdf of \mathbf{X}_2 at that moment was uniformly distributed. Each peak corresponds to the most likely value of y in $\{0, \dots, L\}$ where node \mathbf{X}_2 will be located at jump M_1 . This value is approximately given by $y = x_1(0) + r$.

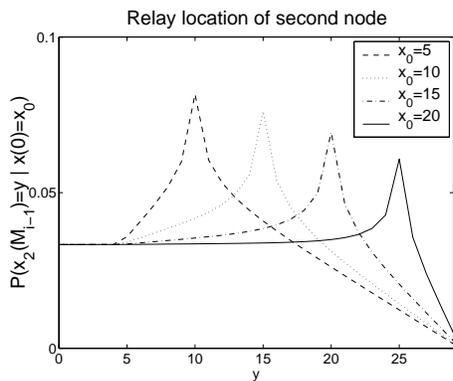


Figure 2.6: Mapping $x \rightarrow P(x_2(M_1) = y)$ (pdf of location of node \mathbf{X}_2 at the relay epoch) when node \mathbf{X}_1 is at position $x_0 \in \{5, 10, 15, 20\}$ at jump $n = 0$. Here $L = 30$ states per segment and $r = 35$.

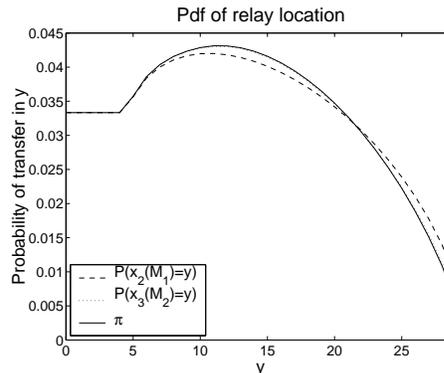


Figure 2.7: Mapping $x \rightarrow P(x_i(M_{i-1}) = y)$ (pdf of the starting location of node i after the $(i - 1)^{th}$ jump), for $i \in \{2, 3, 100\}$. Here $L = 30$ states per segment and $r = 35$.

The mapping $y \rightarrow P(x_i(M_{i-1}) = y)$ is shown in Figure 2.7 for $i \in \{2, 3, \infty\}$. Here $r = 35$, the starting position of the first node is taken from the stationary (uniform) distribution, and there are $L = 30$ states per segment. It is worth observing that the starting positions converge very rapidly (already the starting positions for $i = 2$ and $i = 3$ are very close to each other and the functions for $i = 3$ and π can hardly be distinguished from each other). This phenomenon can also be seen in Figure 2.8 where the expected number of jumps to

perform the i -th relay is shown.

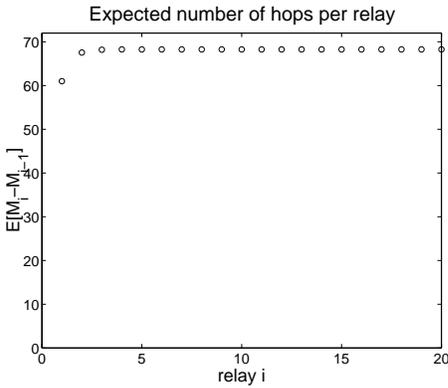


Figure 2.8: Expected number of jumps per relay. Here $L = 30$ states and $r = 35$.

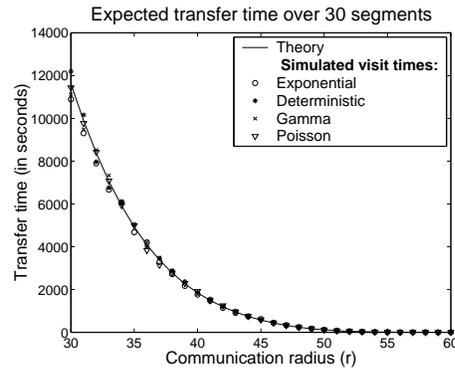


Figure 2.9: Expected transfer time over 30 segments with $L = 30$ states per segment.

Figure 2.9 displays the expected time to transfer a message over 30 segments, with $L = 30$ states per segment and $r = 35$. In addition, the transfer time—based on simulations—is shown for various visit time distributions. This figure once again confirms something which we have seen earlier: the expected transfer time (or number of jumps) is (almost) insensitive to the underlying visit time distribution.

2.3 Extensions to the model

In this chapter the message delay over a one-dimensional network was analysed for mobiles which move as random walkers in adjacent segments. The extension of this theory to more than two mobiles per segment, or with leakage from one domain to the next, does not seem to be mathematically tractable. The reason for this is the following. If there are N mobiles in a segment then their positions needs to be mapped to a single N -dimensional random walker. So far no problem. The model starts becoming more complex though when one has to take into account the positions of the N -dimensional random walker which correspond to any two nodes being in each others transmission range. For two one-dimensional random walkers this resulted in a diagonal line (see Figure 2.10). For three nodes this leads to three intersecting planes in a three-dimensional space. On top of this one has to take into account the (reflecting) border conditions for each of the mobiles. Although writing down these conditions is still feasible, the “real” problems lie in keeping track of which nodes have or have not received a copy of the message, the correlation between these nodes their positions, and finding an expression for the function which corresponds to the expected transfer time and which satisfies all of the necessary conditions.

A similar problem arises for N two-dimensional random walkers. In this case the positions of the mobiles can be mapped to a single $2N$ -dimensional random walker. Since the area in which a node can communicate with another node is given by a circle, it means

that with the method of images the corresponding boundary conditions are no longer in the simple form of a square (since the “mirrors” are located on the outer edge of the circles). Once again, finding a function which corresponds to the expected message delay, while keeping track of which mobile have a copy of the message, does not seem to be feasible.

2.4 Concluding remarks

In this chapter closed-form expressions were derived for the relay locations and for the transfer times (expected number of jumps). In addition, it was found that all of the expressions show insensitivity towards the underlying visit time distribution. For the transfer of a message over a one-dimensional network it was found that the transfer time grows linearly with the number of segments. These results are in agreement with those obtained in Chapter 1 where nodes do not move as Random Walkers but as Brownian motions.

The formulae derived are of use not only for the message delay between nodes which move as random walkers, but they can also be used to determine the time until two messages travelling through a sensor network are near to one another.

Besides message delay, the question of power control is also central in ad hoc networking. Ongoing research is concerned with determining the minimum transmission range that will ensure communication between mobiles (within a certain probability) before the battery power runs out, and with introducing utility functions into the model.

As mentioned earlier, the problem with keeping track of which mobiles do or do not have a copy of the message in combination with their positions greatly increases the complexity and limits the extendibility of the theory presented in this paper. However, by assuming that $r \ll L$ and by working in (at least) two dimensions it has been found that the number of copies in the network can be decoupled from the positions of the copies. This greatly simplifies the analysis and has led to generic results and formulae which hold under a variety of mobility models. This is the topic of the next chapter. Oddly enough, the theory developed there can not be used for the analysis of one-dimensional mobility models and hence there remains a need to study the one- and the two-dimensional settings separately.

Acknowledgements

I would like to express my thanks to Alain Jean-Marie for the discussion we had about the proof of equation (2.5).

2.A The stationary distribution of a random walker

In this section the stationary distribution of a random walker is shown to be uniformly distributed.

If a random walker starts in state j , then the occupation probability of state k just after the n -th jump is given by [31, Formula (3.15), p.357]

$$p_{jk}(n) = \frac{1 - (p/q)}{1 - (p/q)^L} \left(\frac{p}{q}\right)^k + \frac{2^{n+1} p^{1+(n-j+k)/2} q^{(n+j-k)/2}}{L} \sum_{r=1}^{L-1} S_r, \quad (2.21)$$

with

$$S_r = \frac{\cos^n \frac{\pi r}{L} \left\{ \sin \frac{\pi r(j+1)}{L} - \left(\frac{q}{p}\right)^{1/2} \sin \frac{\pi r j}{L} \right\} \left\{ \sin \frac{\pi r(k+1)}{L} - \left(\frac{q}{p}\right)^{1/2} \sin \frac{\pi r k}{L} \right\}}{1 - 2\sqrt{pq} \cos \frac{\pi r}{L}}.$$

The stationary distribution $\pi(k)$ can be obtained from this expression by taking $n \rightarrow \infty$ in equation (2.21). By doing this the second term goes to zero which leaves

$$\pi(k) = \lim_{n \rightarrow \infty} p_{jk}(n) = \frac{1 - (p/q)}{1 - (p/q)^L} \left(\frac{p}{q}\right)^k. \quad (2.22)$$

This last expression is also given in [57, p.104]. If we take $p \rightarrow 1/2$ in (2.22) then we see that

$$\pi(k) = \frac{1}{L}, \quad (2.23)$$

i.e., *the stationary distribution of a symmetric random walker in one or two dimensions is uniformly distributed.*

2.B Proof of Proposition 2.1.1

Proof. Fix $x_0 \in \{0, \dots, L-1\}$ and $y_0 \in \{0, \dots, L-1\}$ such that $r \leq x_0 + r \leq y_0 \leq L-1$. An equivalent way to view the random walkers \mathbf{X} and \mathbf{Y} is to consider the two-dimensional (2D) random walker $\mathbf{Z} = \{z(n) = (x(n), y(n)), n \in \mathbb{Z}^*\}$ as shown in Figure 2.10. The circles in Figure 2.10 represent the states which the 2D random walker can be in. The arrows between the states indicate the neighbouring states the 2D random walker can jump to assuming exponential visit times. If the visit times are deterministic then the arrows point in the two diagonal directions (instead of vertically and horizontally) since both nodes jump at the same time. The boundary states 0 and $L-1$ for the random walkers \mathbf{X} and \mathbf{Y} in the one-dimensional case are represented in the two-dimensional case by reflecting boundaries

at $x = -1/2$, $x = L - 1/2$, $y = -1/2$, and $y = L - 1/2$. With this we see that $N_{L,r}(x_0, y_0)$ is equal to the expected number of jumps needed for the 2D random walker \mathbf{Z} to hit any of the states on the line $y = x + r$ (i.e., the black points in Figure 2.10), given the reflecting boundaries at $x = -1/2$ and $y = L - 1/2$ and that $z(0) = (x_0, y_0)$.

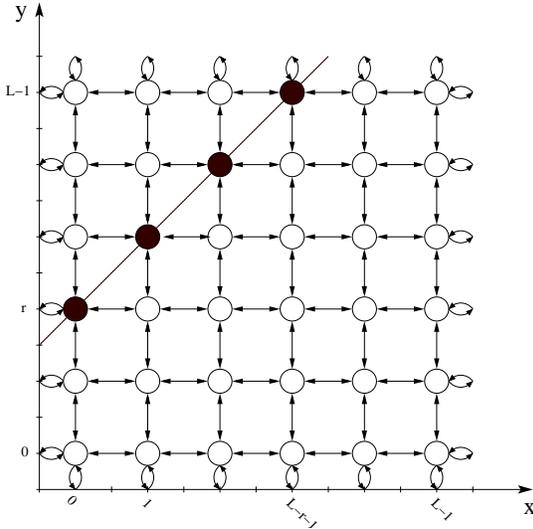


Figure 2.10: The positions of nodes \mathbf{X} and \mathbf{Y} are mapped to two dimensions by $\mathbf{Z} = \{z(N) = (x(N), y(N))\}$. \mathbf{X} and \mathbf{Y} are at a distance r of each other when they are located on the line $y = x + r$.

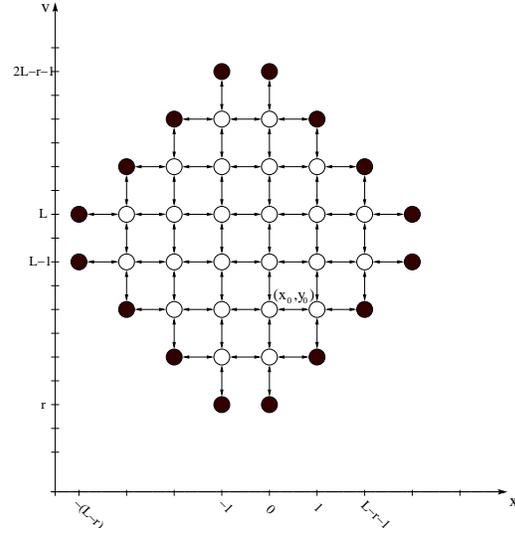


Figure 2.11: Since reflecting barriers at $x = -1/2$ and $y = L - 1/2$ act as mirrors, the method of images turns the problem into finding the expected number of jumps for a 2D random walker starting in (x_0, y_0) to hit one of the black states.

By using the classical method of images (see e.g., [68, p. 81]), the reflecting barriers at $x = -1/2$ and $y = L - 1/2$ can be seen as mirrors resulting in an extension of the state space. The expected number of jumps is then identical to the expected number of jumps needed to hit the boundary in the form of a square as shown in Figure 2.11, given that $z(0) = (x_0, y_0)$.

In order to facilitate readability with regards to sets of points, define the following sets

$$\mathcal{I} := \{x, y \in \mathbb{Z}^* : 0 \leq x \leq L-r-2; x+r+1 \leq y \leq L-1\},$$

$$\mathcal{I}^* := \{x, y \in \mathbb{Z}^* : -(L-r) < x < L-r-1; x+r < y < 2L-r-1-x; \\ r-1-x < y < x+2L-r\},$$

$$\mathcal{B} = \{x, y \in \mathbb{Z}^* : x+r = y; 0 \leq x \leq L-r-1\},$$

$$\mathcal{B}_2 = \{x, y \in \mathbb{Z}^* : x+r = y-2L; 0 \leq x \leq L-r-1\},$$

$$\mathcal{B}_3 = \{x, y \in \mathbb{Z}^* : -x-1+r = y; 0 \leq x \leq L-r-1\},$$

$$\mathcal{B}_4 = \{x, y \in \mathbb{Z}^* : -x-1+r = y-2L; 0 \leq x \leq L-r-1\},$$

$$\mathcal{B}^* = \mathcal{B} \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4.$$

The sets \mathcal{B} and \mathcal{B}^* are the boundaries which correspond, respectively, to the coloured states in Figures 2.10 and 2.11 (for $L = 6$ and $r = 2$). The set \mathcal{I} is the set of states to the left of \mathcal{B} in Figure 2.10, whereas \mathcal{I}^* is the set of interior states which are contained inside the square created by \mathcal{B}^* in Figure 2.11. Whenever the 2D random walk is taken on $\mathcal{I}^* \cup \mathcal{B}^*$ we shall refer to it as the *extended* random walk and denote all of the related sets with a star.

The remainder of this section is devoted to the derivation of the function $N_{L,r}(x_0, y_0)$, the expected time for a 2D random walker starting in state $(x_0, y_0) \in \mathcal{I} \cup \mathcal{B}$ to hit one of the absorbing states in \mathcal{B} , or equivalently, the expected time for the extended 2D random walker starting in state $(x_0, y_0) \in \mathcal{I} \cup \mathcal{B}$ to hit one of the absorbing states in \mathcal{B}^* . First of all, the function must satisfy the boundary conditions

$$N_{L,r}(x, y) = 0, \quad \text{for } (x, y) \in \mathcal{B}^*, \quad (2.24)$$

independent of the visit time distributions. For the time being assume the visit times are exponentially distributed, the deterministic case will be dealt with at a later stage. If the extended random walker starts in $(x, y) \in \mathcal{I}$, then the extended 2D random walker takes at least one jump and therefore the expected number of jumps to hit \mathcal{B}^* , $N_{L,r}(x, y)$, is equal to one plus the expected number of jumps starting in one of the states neighbouring (x, y) . After this jump, the process repeats itself (due to the memoryless property) but in a new state. Thus, for exponential visit times, $N_{L,r}(x, y)$ must also satisfy

$$N_{L,r}(x, y) = 1 + \frac{N_{L,r}(x-1, y) + N_{L,r}(x+1, y) + N_{L,r}(x, y-1) + N_{L,r}(x, y+1)}{4}, \quad (2.25)$$

for $(x, y) \in \mathcal{I}^*$.

In order to find the function which satisfies these conditions, notice that for $j, k \in \{0, 1, \dots, 2(L-r)\}$ each of the components

$$g_{jk}(x, y) := \sin(\theta j(y-r+x+1)) \sin(\theta k(y-r-x)), \quad \theta := \frac{\pi}{R} = \frac{\pi}{2(L-r)}$$

satisfies the boundary conditions of (2.24). By creating a function which is made up of these components we can obtain a function which is also a solution of (2.25). Assume that as a trial solution $N_{L,r}(x, y)$ takes the form

$$N_{L,r}(x, y) = \sum_{j=1}^{R-1} \sum_{k=1}^{R-1} U_{jk} g_{jk}(x, y), \quad (2.26)$$

with $R := 2(L-r)$ and U_{jk} an unknown—to be determined—function of j and k . By plugging this trial solution into (2.25) and bringing terms to the same side we see that U_{jk} must satisfy

$$\sum_{j=1}^{R-1} \sum_{k=1}^{R-1} U_{jk} \left[g_{jk}(x, y) - \frac{g_{jk}(x-1, y) + g_{jk}(x+1, y) + g_{jk}(x, y-1) + g_{jk}(x, y+1)}{4} \right] = 1. \quad (2.27)$$

With the trigonometric identity $\sin \alpha + \sin \beta = 2 \sin(\frac{\alpha+\beta}{2}) \cos(\frac{\alpha-\beta}{2})$ this can be rewritten as

$$\sum_{j=1}^{R-1} \sum_{k=1}^{R-1} U_{jk} \left(1 - \cos(j\theta) \cos(k\theta)\right) g_{jk}(x, y) = 1. \quad (2.28)$$

This gives a condition on U_{jk} that must be satisfied in order for the trial solution of (2.26) to hold. We shall come back to this expression after the derivation of a similar summation—through the means of a Discrete Sine Transform (DST)—which is also equal to one. As the two summations both equal one, the terms in each of the summations can be compared and an expression for U_{jk} is obtained.

Introduce a sequence of real numbers, f_m for $m = 1, \dots, R-1$. Its DST is given by [21, p. 126]

$$F_j = \frac{1}{R} \sum_{m=1}^{R-1} f_m \sin(mj\theta), \quad j = 1, \dots, R-1, \quad (2.29)$$

with its inverse transformation given by [21, p. 127]

$$f_m = 2 \sum_{j=1}^{R-1} F_j \sin(mj\theta), \quad m = 1, \dots, R-1.$$

Taking $f_m = 2 = f_n$ gives $4 = f_m f_n = 4 \sum_{j=1}^{R-1} \sum_{k=1}^{R-1} F_j F_k \sin(mj\theta) \sin(nk\theta)$. This is an expression similar to (2.28) but now with unknown F_j and F_k . However, F_j , and identically F_k , can be calculated by taking $f_m = 2$ in (2.29) leading to the DST of the constant two [21, pages 242 and 300]:

$$F_j = \frac{2}{R} \sum_{m=1}^{R-1} \sin(mj\theta) = \frac{\sin(j\theta)(1 - \cos(\pi j))}{2R \sin^2(j\theta/2)}.$$

Simplifying this using $2 \sin^2(\alpha) = 1 - \cos(2\alpha)$ [37, formula 1.317.1] gives

$$F_j = \begin{cases} 0 & \text{if } j \text{ is even,} \\ \frac{2 \sin(j\theta)}{R(1 - \cos(j\theta))} & \text{if } j \text{ is odd.} \end{cases} \quad (2.30)$$

Recapitulating, we have found F_j and F_k such that

$$\sum_{j=1}^{R-1} \sum_{k=1}^{R-1} F_j F_k \sin(mj\theta) \sin(nk\theta) = \sum_{j=1}^{R-1} F_j \sin(mj\theta) \sum_{k=1}^{R-1} F_k \sin(nk\theta) = 1,$$

for any $m, n = \{1, \dots, R-1\}$. As a specific case we can take $m = y-r+x+1$ and $n = y-r-x$ for any $(x, y) \in \mathcal{I}^*$ and compare this to (2.28). We find

$$U_{jk} = \frac{F_j F_k}{1 - \cos(j\theta) \cos(k\theta)}. \quad (2.31)$$

Thus the proposed trial solution in (2.26) with the coefficients U_{jk} given by (2.31) is the function that we set out to find as it satisfies (2.25) as well as the boundary conditions in (2.24). To the best of our knowledge this expression can not be simplified any further.

Now consider deterministic visit times. Whenever a random walker jumps, the other random walker also jumps. This implies that, in this case, $N_{L,r}(x, y)$ must satisfy

$$N_{L,r}(x, y) = 2 + \frac{N_{L,r}(x-1, y-1) + N_{L,r}(x-1, y+1) + N_{L,r}(x+1, y-1) + N_{L,r}(x+1, y+1)}{4}$$

for $(x, y) \in \mathcal{I}^*$. The two accounts for the fact that two jumps take place (one for each random walker). From here on the proof for the exponential visit times can be reused with the slight modification that (2.27) changes slightly and (2.28) then turns into

$$\sum_{j=1}^{R-1} \sum_{k=1}^{R-1} U_{jk} \left(1 - \frac{\cos(2j\theta) + \cos(2k\theta)}{2}\right) g_{jk}(u, v) = 2.$$

This leads to

$$U_{jk} = \frac{2F_j F_k}{1 - (\cos(2j\theta) + \cos(2k\theta))/2}. \quad \blacksquare$$

2.C Proof of Proposition 2.1.2

Proof. We start by taking exponential visit times. The proof for deterministic visit times is identical but with just a few minor modifications which are outlined at the end of this section.

Since the starting positions of the random walkers \mathbf{X} and \mathbf{Y} are uniform we have

$$\begin{aligned} \mathbb{E}[N_{L,r}] &= \frac{1}{L^2} \sum_{x_0=0}^{L-1} \sum_{y_0=0}^{L-1} \mathbb{E}[N_{L,r} | x(0) = x_0; y(0) = y_0] \\ &= \frac{1}{L^2} \sum_{x_0=0}^{L-r-2} \sum_{y_0=x_0+r+1}^{L-1} N_{L,r}(x_0, y_0) + \frac{1}{L^2} \sum_{x_0=r+1}^{L-1} \sum_{y_0=0}^{x_0-r-1} N_{L,r}(y_0, x_0) \\ &= \frac{2}{L^2} \sum_{x_0=0}^{L-r-2} \sum_{y_0=x_0+r+1}^{L-1} N_{L,r}(x_0, y_0), \end{aligned} \quad (2.32)$$

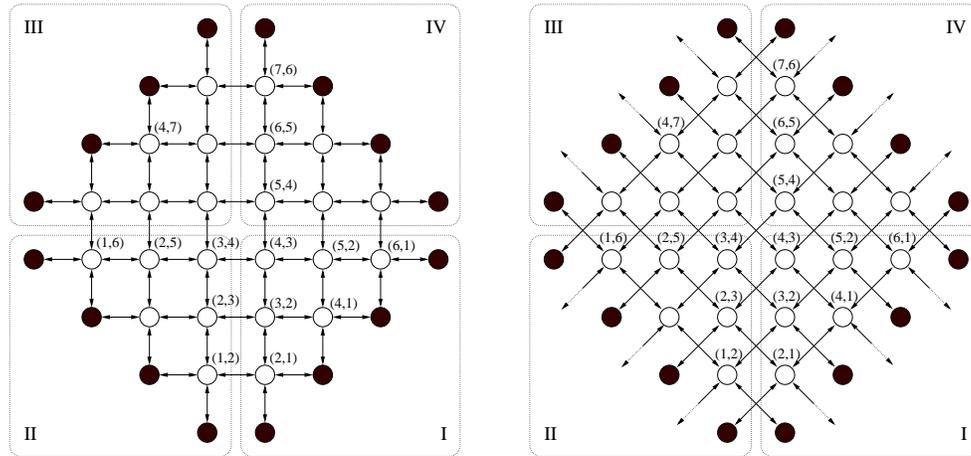
where $N_{L,r}(x_0, y_0)$ is given in Proposition 2.1.1. The last step follows from the interchangeability of x_0 and y_0 . Some insight can be gained into this function with the change of variables $u = y_0 - r + x_0 + 1$ and $v = y_0 - r - x_0$ (this shifts and rotates the coordinate system by 45 degrees). See Figure 2.12 for an example (with $R = 8$) of the labelling of the states in the new coordinate system (u, v) . The figure on the left is for exponential visit times and the figure on the right is for deterministic visit times. Equation (2.32) then becomes

$$\mathbb{E}[N_{L,r}] = \frac{2}{L^2} \left(\sum_{u=2}^{L-r} \sum_{\substack{v=1 \\ u+v \text{ odd}}}^{u-1} h(u, v) + \sum_{u=L-r+1}^{R-2} \sum_{\substack{v=1 \\ u+v \text{ odd}}}^{R-u-1} h(u, v) \right), \quad (2.33)$$

with

$$h(u, v) := \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{F_j F_k \sin(j\theta u) \sin(k\theta v)}{1 - \cos(j\theta) \cos(k\theta)}. \quad (2.34)$$

The dependence of the inner summations of (2.33) on u makes them difficult to evaluate. However, by rewriting the summation in such a way that they no longer depend on u it allows them to be evaluated and an approximate closed-form expression can be obtained. In order to do this, first note that the summations in (2.33) define a sum over a triangular region (quadrant *I* in Figure 2.12).



(a) Exponential visit times.

(b) Deterministic visit times.

Figure 2.12: The states with labels in the new coordinate system (u, v) with $R = 8$. The sum of the values of the points in *I* is equal to the sum of the values of the points in *II*, *III*, or *IV*.

By using the identity $\sin(ju\theta) = \sin(j(R-u)\theta)$ it turns out that $h(u, v)$ is symmetric in the sense that

$$h(u, v) = h(v, u) = h(R-u, R-v) = h(R-v, R-u), \quad (u, v) \in \mathcal{I}^*.$$

This symmetry, while considering the four quadrants *I*, *II*, *III*, and *IV* as shown in Figure 2.12, reveals that the summation over the points in *I* is equal to the summation over the points in *II*, *III*, or *IV*. Therefore the value of the summations in (2.33) is equal to a quarter of the summation over the entire area ($I + II + III + IV$). This gives

$$\mathbb{E}[N_{L,r}] = \frac{1}{2L^2} \left(\sum_{\substack{u=1 \\ u \text{ odd}}}^{R-1} \sum_{\substack{v=1 \\ v \text{ even}}}^{R-1} h(u, v) + \sum_{\substack{u=1 \\ u \text{ even}}}^{R-1} \sum_{\substack{v=1 \\ v \text{ odd}}}^{R-1} h(u, v) \right) = \frac{1}{L^2} \sum_{\substack{u=1 \\ u \text{ odd}}}^{R-1} \sum_{\substack{v=1 \\ v \text{ even}}}^{R-1} h(u, v),$$

where in the second step once again the symmetry $h(u, v) = h(v, u)$ was used. This expression can be simplified by changing the order of summation (this is allowed since the sums are finite) to reveal

$$\mathbb{E}[N_{L,r}] = \frac{1}{L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{F_j F_k}{1 - \cos(j\theta) \cos(k\theta)} \sum_{\substack{u=1 \\ u \text{ odd}}}^{R-1} \sin(ju\theta) \sum_{\substack{v=1 \\ v \text{ even}}}^{R-1} \sin(kv\theta).$$

The summations over the sin can be evaluated with the help of [37, Formulas 1.342.1 & 1.342.3]

$$\begin{aligned} \sum_{\substack{u=1 \\ u \text{ odd}}}^{R-1} \sin(j\theta u) &= \frac{\sin^2(j\pi)}{\sin(j\theta)} = \begin{cases} 0 & \text{if } j \text{ is even,} \\ \frac{1}{\sin(j\theta)} & \text{if } j \text{ is odd,} \end{cases} & (2.35) \\ \sum_{\substack{v=1 \\ v \text{ even}}}^{R-1} \sin(k\theta v) &= \frac{\sin(k\pi/2) \sin(k\pi/2 - k\theta)}{\sin(k\theta)} = \begin{cases} 0 & \text{if } k \text{ is even,} \\ \frac{\cos(k\theta)}{\sin(k\theta)} & \text{if } k \text{ is odd,} \end{cases} \end{aligned}$$

through which the first part of Proposition 2.1.2 is obtained,

$$\mathbb{E}[N_{L,r}] = \frac{4}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{\cos(k\theta)}{(1 - \cos(j\theta))(1 - \cos(k\theta))(1 - \cos(j\theta) \cos(k\theta))}. \quad (2.36)$$

Although an explicit expression the above summation is, to the best of our knowledge, not known, its scaling behaviour in R (or equivalently, in L and r) can be derived and will be shown to be equal to equation (2.5). To render the summation benign, we start by looking at the terms inside the summations. Bounds are then derived for these terms (with the help of Taylor expansions) after which the bounds for the summation are obtained.

We start by looking at the term $(1 - \cos(x))^{-1}$ and point out that it is largest around $x = 0$ (for $0 < x < \pi$). Consider the Taylor expansion [70, Table 4.2] of $\cos(x)$ around this point,

$$\cos(x) = 1 - x^2/2 + \underbrace{\sum_{k=2}^{\infty} \frac{x^{2k}(-1)^k}{(2k)!}}_{\varepsilon_x}. \quad (2.37)$$

Note that ε_x is positive for $0 < x < \pi$ and therefore $1 - x^2/2 < \cos(x) < 1$. Also, ε_x is bounded in value by

$$0 < \varepsilon_x = \frac{x^4}{4!} - \sum_{k=1}^{\infty} \frac{x^{4k+2}}{(4k+2)!} \left(1 - \frac{x^2}{(4k+4)(4k+3)}\right) < \frac{x^4}{4!}.$$

Using the Taylor expansion and the geometric series gives

$$\frac{1}{1 - \cos(x)} = \frac{1}{1 - (1 - x^2/2 + \varepsilon_x)} = \frac{2}{x^2} \left(\frac{1}{1 - 2\varepsilon_x/x^2} \right) = \frac{2}{x^2} \left(1 + \sum_{n \geq 1} (2\varepsilon_x/x^2)^n \right).$$

The last step is justified since $\frac{2\varepsilon_x}{x^2} < \frac{x^2}{12} < \frac{\pi^2}{12} < 1$. With the help of the above expression we find the following bounds when $0 < x < \pi$

$$\frac{2}{x^2} < \frac{1}{1 - \cos(x)} < \frac{2}{x^2} \left(1 + \sum_{n \geq 1} \left(\frac{x^2}{12} \right)^n \right) < \frac{2}{x^2} \left(1 + \sum_{n \geq 1} \left(\frac{\pi^2}{12} \right)^n \right) = \frac{2}{x^2} + C_1, \quad (2.38)$$

where $C_1 := \frac{1}{6 - \pi^2/2} \approx 0.94$.

Next consider the term $(1 - \cos(x)\cos(y))^{-1}$. This term contains two poles; one in $(x, y) = (0, 0)$ and the other in $(x, y) = (\pi, \pi)$. With the help of the Taylor expansion of the \cos around those points it can be seen that $(1 - \cos(x)\cos(y))^{-1}$ grows roughly with rate $2/(x^2 + y^2)$ and $2/((\pi - x)^2 + (\pi - y)^2)$. Making use of this knowledge we write

$$\frac{1}{1 - \cos(x)\cos(y)} = \frac{2}{x^2 + y^2} + \frac{2}{(\pi - x)^2 + (\pi - y)^2} + \beta_{xy},$$

where

$$\beta_{xy} = \frac{1}{1 - \cos(x)\cos(y)} - \frac{2}{x^2 + y^2} - \frac{2}{(\pi - x)^2 + (\pi - y)^2}. \quad (2.39)$$

By making a figure of β_{xy} (see Figure 2.13) for $0 \leq x, y \leq \pi$ it can quickly be seen that $0 < \beta_{xy}$ and that the maximum is $\lim_{x \rightarrow 0} \beta_{xx} = \lim_{x \rightarrow \pi} \beta_{xx} = \frac{\pi^2 - 3}{3\pi^2} := C_2 \approx 0.232$. This

gives (for $0 < x, y < \pi$)

$$\frac{2}{x^2+y^2} + \frac{2}{(\pi-x)^2 + (\pi-y)^2} < \frac{1}{1-\cos(x)\cos(y)} < \frac{2}{x^2+y^2} + \frac{2}{(\pi-x)^2 + (\pi-y)^2} + C_2, \quad (2.40)$$

from which we also obtain

$$\frac{2}{x^2+y^2} < \frac{1}{1-\cos(x)\cos(y)}. \quad (2.41)$$

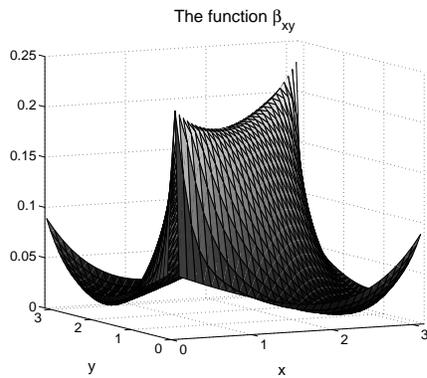


Figure 2.13: The function β_{xy} given by equation (2.39).

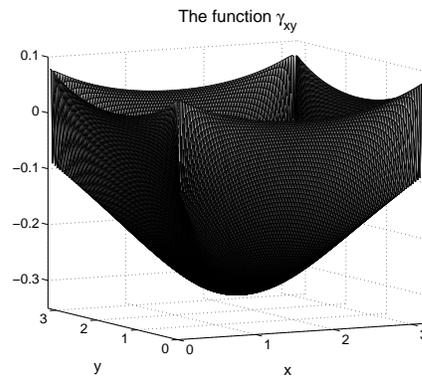


Figure 2.14: The function γ_{xy} given by equation (2.46).

This brings us one step closer to unveiling the bounds for the expected number of jumps. Continuing from (2.36), with the help of (2.37), (2.38), and (2.41), the expected number of jumps is bounded from below by

$$\begin{aligned} \mathbb{E}[N_{L,r}] &> \frac{4}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \left(1 - \frac{\pi^2 k^2}{2R^2}\right) \left(\frac{2R^2}{\pi^2 j^2}\right) \left(\frac{2R^2}{\pi^2 k^2}\right) \left(\frac{2R^2}{\pi^2(j^2 + k^2)}\right) \\ &= \frac{32R^4}{\pi^6 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \left(\frac{1}{j^2 k^2 (j^2 + k^2)} - \left(\frac{\pi^2}{2R^2}\right) \frac{1}{j^2 (j^2 + k^2)}\right). \end{aligned}$$

The summations in this lower bound are still dependent on R and the behaviour in R is therefore not obvious. To overcome this problem we shall extend the summations to infinity. To do this first observe that $1/j^2 - 1/R^2 \leq 0$ whenever $j \geq R$. This means that we

can decrease the bound by adding a couple of negative terms,

$$\begin{aligned} \mathbb{E}[N_{L,r}] &> \frac{32R^4}{\pi^6 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \left(\frac{1}{j^2 k^2 (j^2 + k^2)} - \left(\frac{\pi^2}{2R^2} \right) \frac{1}{j^2 (j^2 + k^2)} \right) \\ &+ \frac{32R^4}{\pi^6 L^2} \sum_{\substack{j=R \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \underbrace{\left(\frac{1}{j^2} - \frac{1}{R^2} \right)}_{\leq 0} \frac{1}{k^2 (j^2 + k^2)} \\ &+ \frac{32R^4}{\pi^6 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=R \\ k \text{ odd}}}^{\infty} \underbrace{\left(\frac{1}{k^2} - \frac{1}{R^2} \right)}_{\leq 0} \frac{1}{j^2 (j^2 + k^2)}. \end{aligned}$$

Collecting terms gives the lower bound:

$$\mathbb{E}[N_{L,r}] > \frac{32R^4}{\pi^6 L^2} \left(C_3 - \frac{2 + \pi^2}{2R^2} C_4 \right), \quad (2.42)$$

with $C_3 := \sum_{j \geq 1} \sum_{\substack{k \geq 1 \\ k \text{ odd}}} \frac{1}{j^2 k^2 (j^2 + k^2)} \approx 0.5279$ and

$$\begin{aligned} C_4 &:= \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=R \\ k \text{ odd}}}^{\infty} \frac{1}{k^2 (j^2 + k^2)} = \frac{1}{2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=R \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \left(\frac{1}{j^2} + \frac{1}{k^2} \right) \\ &= \frac{1}{2} \sum_{\substack{j \geq 1 \\ j \text{ odd}}} \sum_{\substack{k=R \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \left(\frac{j^2 + k^2}{j^2 k^2} \right) = \frac{1}{2} \left(\frac{\pi^2}{8} \right)^2, \end{aligned}$$

where in the last step $\sum_{k \geq 0} \frac{1}{(2k+1)^2} = \pi^2/8$ [37, Formula 0.234.2] was used.

Next is the derivation of an upper bound. Starting once again from (2.36) and using the upper bounds of (2.38) and (2.40) we find that the expected number of jumps is bounded from above by

$$\begin{aligned} \mathbb{E}[N_{L,r}] &< \frac{4}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \left(\frac{2R^2}{\pi^2 (j^2 + k^2)} + \frac{2R^2}{\pi^2 ((R-j)^2 + (R-k)^2)} + C_2 \right) \\ &\quad \times \left(\frac{2R^2}{\pi^2 j^2} + C_1 \right) \left(\frac{2R^2}{\pi^2 k^2} + C_1 \right) \end{aligned} \quad (2.43)$$

This summation can be simplified without raising the bound too much. For most of the terms behind the summation signs the indexes of the double summations can be extended to infinity leading to combinations of the constants C_3 and C_4 , or use can be made of

$\sum_{1 \leq k \leq R-1}^{k \text{ odd}} 1 = R/2$. One of the summations which should not blindly be extended to infinity is

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + k^2} < \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2} = \frac{\pi^2 R}{16}.$$

To bound a number of the other terms note that

$$\begin{aligned} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \frac{1}{j^2(R-j)^2} &= \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \frac{1}{R^2} \left(\frac{1}{j^2} + \frac{1}{(R-j)^2} + \frac{2}{jR} + \frac{2}{(R-j)R} \right) \\ &< \frac{2}{R^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \frac{1}{j^2} + \frac{4}{R^3} \left(1 + \sum_{j=1}^{R/2} \frac{1}{2j+1} \right). \end{aligned}$$

Using the asymptotic expansion of the harmonic numbers [87, page 184] this gives

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \frac{1}{j^2(R-j)^2} < \frac{\pi^2}{4R^2} + \frac{4}{R^3} \left(1 + \sum_{j=1}^{R/2} \frac{1}{2j} \right) < \underbrace{\frac{\pi^2}{4R^2} + \frac{4+2(\ln(R/2)+\gamma+1/R)}{R^3}}_{=:f(R)},$$

where $\gamma \approx 5.772$ is Euler's constant and $f(R)$ is a function which is $\mathcal{O}(1/R^2)$. One of the terms is then bounded by

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + (R-k)^2} \frac{1}{j^2} \frac{1}{k^2} < \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{(R-j)^2} \frac{1}{j^2} \frac{1}{k^2} < \frac{\pi^2}{8} f(R),$$

and another is bounded by

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + (R-k)^2} \frac{1}{j^2} < \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2} \frac{1}{j^2} < \frac{R}{2} f(R).$$

With the help of these expressions we can obtain upper bounds for each of the terms in equation (2.43).

$$\begin{aligned}
& \frac{32R^4 C_1}{\pi^6} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \frac{1}{j^2} \frac{1}{k^2} = \frac{32R^4}{\pi^6} C_3 \\
& \frac{16R^2 C_1}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \left(\frac{1}{j^2} + \frac{1}{k^2} \right) = \frac{16R^2}{\pi^4} C_4 = \mathcal{O}(R^2) \\
& \frac{8C_1^2}{\pi^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + k^2} \leq \frac{C_1^2 R}{2} = \mathcal{O}(R) \\
& \frac{32R^4}{\pi^6} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + (R-k)^2} \frac{1}{j^2} \frac{1}{k^2} \leq \frac{R^4}{4\pi^4} f(R) = \mathcal{O}(R^2) \\
& \frac{16R^2}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + (R-k)^2} \left(\frac{1}{j^2} + \frac{1}{k^2} \right) \leq \frac{16R^3}{\pi^4} f(R) = \mathcal{O}(R) \\
& \frac{8C_1^2}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{(R-j)^2 + (R-k)^2} = \frac{\pi^2 C_1^2}{16} = \mathcal{O}(R) \\
& \frac{16R^2 C_2}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2} \frac{1}{k^2} = \frac{16R^2 C_2}{\pi^4} \left(\frac{\pi^2}{8} \right)^2 = \mathcal{O}(R^2) \\
& \frac{8C_1 C_2}{\pi^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \left(\frac{1}{j^2} + \frac{1}{k^2} \right) = \frac{8C_1 C_2 R}{\pi^2} \left(\frac{\pi^2}{8} \right) = \mathcal{O}(R) \\
& \frac{4C_1^2 C_2}{R^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} 1 = \frac{4C_1^2 C_2}{R^2} \left(\frac{R}{2} \right)^2 = \mathcal{O}(R^2)
\end{aligned}$$

Combining all of this gives

$$\mathbb{E}[N_{L,r}] \leq \frac{32R^4}{\pi^6 L^2} C_3 + g(R/L),$$

where $g(x)$ is a function such that $\lim_{x \rightarrow \infty} g(x) = k$, where k is some constant greater than zero, i.e., $g(x) = \mathcal{O}(1)$. With the lower bound on the other side given by (2.42), we obtain the final expression for exponential visit times:

$$\mathbb{E}[N_{L,r}] = \frac{32R^4}{\pi^6 L^2} C_3 + \mathcal{O}(R^2/L^2). \quad (2.44)$$

The proof for deterministic visit times goes along identical lines but with a few minor differences. First of all (2.34) becomes

$$h(u, v) := \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{2F_j F_k \sin(j\theta u) \sin(k\theta v)}{1 - (\cos(2j\theta) + \cos(2k\theta))/2}.$$

Performing similar calculations from here on leads to

$$\mathbb{E}[N_{L,r}] = \frac{8}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{\cos(k\theta)}{(1 - \cos(j\theta))(1 - \cos(k\theta))(1 - (\cos(2j\theta) + \cos(2k\theta))/2)}, \quad (2.45)$$

which is (2.4b). Since $(1 - (\cos(2x) + \cos(2y))/2)^{-1}$ has a pole in $(x, y) = (0, 0)$ we can proceed as previously to show that

$$\frac{1}{x^2 + y^2} < \frac{1}{1 - (\cos(2x) + \cos(2y))/2},$$

which—for the same reasons as eq.(2.40)—leads to the same lower bound given in equation (2.42). There are three more poles in $(x, y) = (0, \pi)$, $(\pi, 0)$, and (π, π) for $0 < x, y < \pi$ which leads to the approximation

$$\frac{1}{1 - (\cos(2x) + \cos(2y))/2} = \frac{1}{x^2 + y^2} + \frac{1}{x^2 + (\pi - y)^2} + \frac{1}{(\pi - x)^2 + y^2} + \frac{1}{(\pi - x)^2 + (\pi - y)^2} + \gamma_{xy},$$

where

$$\gamma_{xy} := \frac{1}{1 - (\cos(2x) + \cos(2y))/2} - \frac{1}{x^2 + y^2} - \frac{1}{x^2 + (\pi - y)^2} - \frac{1}{(\pi - x)^2 + y^2} - \frac{1}{(\pi - x)^2 + (\pi - y)^2} \quad (2.46)$$

is shown in Figure 2.14 and is bounded by $\lim_{x \rightarrow 0} \gamma_{x0} = \frac{1}{2} - \frac{8}{\pi^2} < \gamma_{xy} < \lim_{x \rightarrow \pi/2} \gamma_{xx} = \frac{1}{3} - \frac{5}{2\pi^2} := C_5$ for $0 < x, y < \pi$. An upper bound for (2.45) is then given by

$$\begin{aligned} \mathbb{E}[N_{L,r}] < \frac{8}{R^2 L^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \left(\frac{R^2}{\pi^2(j^2 + k^2)} + \frac{R^2}{\pi^2(j^2 + (R-k)^2)} + \frac{R^2}{\pi^2((R-j)^2 + k^2)} \right. \\ \left. + \frac{R^2}{\pi^2((R-j)^2 + (R-k)^2)} + C_5 \right) \left(\frac{2R^2}{\pi^2 j^2} + C_1 \right) \left(\frac{2R^2}{\pi^2 k^2} + C_1 \right) \quad (2.47) \end{aligned}$$

Of the first five terms right behind the summation signs, the evaluation of the first and last terms have already been dealt with (but now with C_5 replaced by C_2). It remains to assess

the other three remaining terms, of which two are identical due to symmetry. First of all we have

$$\begin{aligned}
\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + (R-k)^2} \frac{1}{j^2 k^2} &< \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R/2} \frac{1}{j^2 + (R/2)^2} \frac{1}{j^2 k^2} + \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R/2} \frac{1}{j^2 + k^2} \frac{1}{j^2 (R/2)^2} \\
&< \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{4}{R^2} \frac{1}{j^2 k^2} + \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \frac{1}{j^2} \frac{4}{R^2} \\
&= \frac{4}{R^2} \left(\frac{\pi^4}{64} + C_4 \right) = \frac{3\pi^4}{32R^2}.
\end{aligned}$$

Another of the terms is

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2((R-j)^2 + k^2)} < \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2(R-j)^2} < \frac{R}{2} f(R),$$

and because of symmetry

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{k^2((R-j)^2 + k^2)} = \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{k^2(j^2 + k^2)} < C_4.$$

With the help of these expressions we obtain upper bounds for each of the expressions in (2.47). We have

$$\begin{aligned}
\frac{32R^4 C_1}{\pi^6} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \frac{1}{j^2} \frac{1}{k^2} &= \frac{32R^4}{\pi^6} C_3 \\
\frac{32R^4}{\pi^6} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + (R-k)^2} \frac{1}{j^2} \frac{1}{k^2} &\leq \frac{3R^2}{\pi^2} = \mathcal{O}(R^2) \\
\frac{32R^4}{\pi^6} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + k^2} \frac{1}{j^2} \frac{1}{k^2} &\leq \frac{3R^2}{\pi^2} = \mathcal{O}(R^2) \\
\frac{32R^4}{\pi^6} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{(R-j)^2 + (R-k)^2} \frac{1}{j^2} \frac{1}{k^2} &\leq \frac{R^4}{4\pi^4} f(R) = \mathcal{O}(R^2) \\
\frac{32R^2}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2} \frac{1}{k^2} &\leq \frac{32R^2 C_5}{\pi^4} \left(\frac{\pi^2}{8} \right)^2 = \mathcal{O}(R^2)
\end{aligned}$$

and also

$$\begin{aligned}
 & \frac{16R^2C_1}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} \frac{1}{k^2} = \frac{R^2C_1}{8} = \mathcal{O}(R^2) \\
 & \frac{16R^2C_1}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + (R-k)^2} \frac{1}{k^2} \leq \frac{8R^3C_1}{\pi^4} f(R) = \mathcal{O}(R) \\
 & \frac{16R^2C_1}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{(R-j)^2 + k^2} \frac{1}{k^2} = \frac{16R^2C_1C_4}{\pi^4} = \mathcal{O}(R^2) \\
 & \frac{16R^2C_1}{\pi^4} \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + (R-k)^2} \frac{1}{k^2} \leq \frac{8R^3C_1}{\pi^4} f(R) = \mathcal{O}(R) \\
 & \frac{16C_1C_5}{\pi^2} \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{k^2} = C_1C_2R = \mathcal{O}(R)
 \end{aligned}$$

Furthermore,

$$\begin{aligned}
 & \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + (R-k)^2} = \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + k^2} \\
 & = \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{(R-j)^2 + (R-k)^2} \\
 & = \sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \frac{1}{j^2 + k^2} \\
 & < \sum_{\substack{j=1 \\ j \text{ odd}}}^{\infty} \sum_{\substack{k=1 \\ k \text{ odd}}}^{\infty} \frac{1}{j^2 + k^2} = \frac{R\pi^2}{16} = \mathcal{O}(R),
 \end{aligned}$$

and

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{R-1} \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} 1 = \frac{R^2}{4} = \mathcal{O}(R^2).$$

This gives us

$$\mathbb{E}[N_{L,r}] = \frac{32R^4}{\pi^6 L^2} + g_2(R/L),$$

where $g_2(x)$ is a function with the property $\lim_{x \rightarrow \infty} g_2(x) = k_2$, where k_2 is some constant bigger than zero. In particular, $g_2(x)$ is $\mathcal{O}(1)$. Together with the lower bound in equation (2.42) this gives, just as for exponential visit times,

$$\mathbb{E}[N_{L,r}] = \frac{32R^4}{\pi^6 L^2} + \mathcal{O}\left(\left(\frac{L-r}{L}\right)^2\right). \quad \blacksquare$$

2.D Proof of Proposition 2.2.1

Proof. Let $x(n)$ and $y(n)$ be the relative positions of nodes \mathbf{X} and \mathbf{Y} at the n^{th} jump (here n is not the individual, but the sum of the number of jumps of the two nodes). Let N be the first time the nodes can communicate, i.e.,

$$N = \inf\{n \in \mathbb{Z}^* : x(n) + r \geq y(n) + L\}, \quad L \leq r \leq 2L.$$

Observe that $N = 0$ if $x(0) + r \geq y(0) + L$. By conditioning on the starting positions we find

$$\begin{aligned} P(y(N) = y | x(0) = x_0) &= \sum_{y_0=0}^{x_0+r-L} P(y(N) = y | x(0) = x_0; y(0) = y_0) P(y(0) = y_0) \\ &\quad + \sum_{y_0=x_0+r-L+1}^{L-1} P(y(N) = y | x(0) = x_0; y(0) = y_0) P(y(0) = y_0) \\ &= \frac{1}{L} \sum_{y_0=0}^{x_0+r-L} \mathbf{1}_{\{y=y_0\}} + \frac{1}{L} \sum_{y_0=x_0+r-L+1}^{L-1} P(y(N) = y | x(0) = x_0; y(0) = y_0) \\ &= \frac{\mathbf{1}_{\{y \leq x_0+r-L\}}}{L} + \frac{\mathbf{1}_{\{x_0+r \leq 2L-1, y \geq r-L\}}}{L} \sum_{y_0=x_0+r-L+1}^{L-1} \sum_{n=1}^{\infty} f_y^{(n)}, \end{aligned} \quad (2.48)$$

where

$$f_y^{(n)} = P(y(N) = y; N = n | x(0) = x_0; y(0) = y_0)$$

is the probability that the two nodes can communicate for the first time in the n^{th} step with node \mathbf{Y} in state $y(n) = y$. In (2.48) it is best to include the indicator function to reflect the fact that $f_y^{(n)} = 0$ outside the indicated values of x_0 and y .

It remains to solve $f_y^{(n)}$. We shall do this with the help of the probability

$$p_{x,y,n} := P\left(x(n) = x; y(n) = y; N > n | x(0) = x_0; y(0) = y_0; N > n-1\right),$$

that the two nodes starting in x_0 and y_0 are in states $x(n) = x$, respectively, $y(n) = y$ in the n^{th} step and that they do not communicate in step n . In particular, this means that $p_{\dot{x},y,n} = 0$ for $\dot{x} := y - r + L$, any y and any n . Equivalently, the positions of the two nodes can be viewed as a 2D random walker $\mathbf{Z} = \{z(n) = (x(n), y(n))\}$ and then $p_{x,y,n}$ represents the probability that the 2D random walker is in (x, y) in the n^{th} step, given that the random walker started in (x_0, y_0) and that it has not yet been absorbed by any of the absorbing states $\{(x, y) : x + r = y + L\}$.

It is now possible to express $f_y^{(n)}$ in terms of the position of the 2D random walker one step earlier. If $y(n) = y$, then necessarily $x(n) = \dot{x}$ with $\dot{x} := y - r + L$. Hence at time $n-1$ this means that $(x(n-1), y(n-1)) = (\dot{x} - 1, y)$ or $(x(n-1), y(n-1)) = (\dot{x}, y + 1)$. This gives

$$f_y^{(n)} = \begin{cases} p_{\dot{x},y+1,n-1}/4 & \text{if } y = r - L \\ (p_{\dot{x}-1,y,n-1} + p_{\dot{x},y+1,n-1})/4 & \text{if } r - L < y < L - 1 \\ p_{\dot{x}-1,y,n-1}/4 & \text{if } y = L - 1 \end{cases} \quad (2.49)$$

Although it not written for the sake of readability, this equation still depends on (x_0, y_0) . It remains to find $p_{x,y,n}$. Because of the possibility that the random walkers can jump back to the same state (at the outermost states), it is easier to first solve the related probability

$$\tilde{p}_{x,y,n} := P(\tilde{x}(n) = x; \tilde{y}(n) = y; N > n \mid \tilde{x}(0) = x_0; \tilde{y}(0) = y_0; N > n - 1),$$

of the extended 2D random walker and to then express $p_{x,y,n}$ in terms of these probabilities. The latter is easy to do since the mirror images contribute to the original probability as follows

$$p_{x,y,n} = \tilde{p}_{x,y,n} + \tilde{p}_{-x-1,y,n} + \tilde{p}_{x,2L-1-y,n} + \tilde{p}_{-x-1,2L-1-y,n}. \quad (2.50)$$

To give an expression for $\tilde{p}_{x,y,n}$ we will first establish the requirements which uniquely define $\tilde{p}_{x,y,n}$. We will then give the function $\tilde{p}_{x,y,n}$ which satisfies these requirements.

By definition, $\tilde{p}_{x_0,y_0,0} = 1$ and $\tilde{p}_{x,y,0} = 0$ if $x \neq x_0$ or $y \neq y_0$. Furthermore, since the boundaries are absorbing states it is required that $\tilde{p}_{x,y,n} = 0$ for $(x, y) \in \tilde{\mathcal{N}}$. Finally, if at step n the extended random walker is in (r, s) , then the next step takes it to one of the four neighbours $(r \pm 1, s), (r, s \pm 1)$. This gives the relationship

$$\tilde{p}_{x,y,n+1} = \frac{\tilde{p}_{x-1,y,n} + \tilde{p}_{x+1,y,n} + \tilde{p}_{x,y-1,n} + \tilde{p}_{x,y+1,n}}{4}, \quad n \in \mathbb{Z}^*, \quad (2.51)$$

The function $\tilde{p}_{x,y,n}$ which satisfies these requirements is given by

$$\tilde{p}_{x,y,n} = \frac{4}{R^2} \sum_{j=1}^{2(L-r)-1} \sum_{k=1}^{2(L-r)-1} \left[\cos(j\theta) \cos(k\theta) \right]^n \dot{g}_{jk}(x, y) \dot{g}_{jk}(x_0, y_0), \quad (2.52)$$

with $\dot{r} = r - L$ and

$$\dot{g}_{jk}(x, y) := \sin\left(\frac{\pi j(y - \dot{r} + x + 1)}{2(L - \dot{r})}\right) \sin\left(\frac{\pi k(y - \dot{r} - x)}{2(L - \dot{r})}\right), \quad r + x \leq y \leq L - 1.$$

First of all, it satisfies the boundary conditions ($\tilde{p}_{xy_n} = 0$ for $(x, y) \in \mathcal{B}$). Secondly, $\tilde{p}_{x_0, y_0, 0} = 1$ follows from $\sum_{j=1}^R \sin^2(\pi j u / R) = 0$ [37, equation 1.351.1]. Finally, to show it satisfies (2.51), we make use of the same relationship as used to go from (2.27) to (2.28), namely

$$\dot{g}_{jk}(x-1, y) + \dot{g}_{jk}(x+1, y) + \dot{g}_{jk}(x, y-1) + \dot{g}_{jk}(x, y+1) = 4 \cos(j\theta) \cos(k\theta) \dot{g}_{jk}(x, y).$$

Putting (2.52) into (2.51) gives

$$\begin{aligned} & \tilde{p}_{x, y, n+1} - \frac{\tilde{p}_{x-1, y, n} + \tilde{p}_{x+1, y, n} + \tilde{p}_{x, y-1, n} + \tilde{p}_{x, y+1, n}}{4} \\ &= \frac{4}{R^2} \sum_{j=1}^{2(L-\dot{r}-1)} \sum_{k=1}^{2(L-\dot{r}-1)} \left[\cos(j\theta) \cos(k\theta) \right]^n \dot{g}_{jk}(x_0, y_0) \times \\ & \quad \times \left[\cos(j\theta) \cos(k\theta) \dot{g}_{jk}(x, y) - \frac{\dot{g}_{jk}(x-1, y) + \dot{g}_{jk}(x+1, y) + \dot{g}_{jk}(x, y-1) + \dot{g}_{jk}(x, y+1)}{4} \right] \\ &= 0 \end{aligned}$$

for $(x, y) \in \mathcal{I}$. The proposition follows by combining (2.49)-(2.52), changing the order of summation, and evaluating the geometric summation over N . \blacksquare

Message Delay in Two-dimensional Networks

Contents

3.1	Introduction	64
3.2	The stochastic model	66
3.3	Applications	75
3.4	Large networks	84
3.5	Concluding Remarks	85
3.A	Proof of Lemma 3.2.1	86
3.B	Derivation of the time until the next event	88

In this chapter a stochastic model is introduced that accurately models the message delay in a mobile ad hoc network (MANET) where nodes can relay messages for each other. The model has only two input parameters: the number of nodes and the intensity of a finite number of homogeneous and independent Poisson processes modeling instances when any pair of nodes come within transmission range of one another. Closed-form expressions are obtained for the Laplace-Stieltjes transform of the message delay, defined as the time needed to transfer a message between a source and a destination. From this result, we derive the expected message delay in closed-form as well as its asymptotic expansion for large networks. The probability distribution of the number of copies of the message at the time the message is delivered is also computed. These calculations are carried out for two relay protocols, the two-hop relay and the unrestricted relay protocols. Despite its simplicity, the model is able to accurately predict the performance of both relay protocols for a number of mobility models (random waypoint, random direction and random walker mobility models), as shown by simulations.

Note: Part of the material in this chapter has been published as an extended abstract in the ACM SIGMETRICS 2005 proceedings [44] and will appear in the proceedings of PERFORMANCE 2005 [69]. It is also available as an INRIA research report [43].

3.1 Introduction

In MANET a mobile node (or simply a node) can only send data directly to another node if both nodes are within transmission range of one another or *in contact*. Two nodes are said to be within transmission range of one another if the distance between them does not some threshold R .

The fact that two nodes are in contact is of course not enough to ensure a success transmission of a message. Many phenomena may occur during the transmission and cause it to fail (interferences due to other transmissions, physical obstacles, power problems, etc.). Message relaying is a technique that facilitates the delivery of a message by using intermediary nodes to forward the message.

Routing protocols using relay nodes [61, 63, 89] have recently been proposed that increase the message delivery ratio in mobile ad hoc networks. These protocols operate on a *store-carry-forward* mode to take advantage of node mobility to improve node connectivity, and ultimately the message throughput. When information is available (node movement, node position, etc.) these protocols may use it in a static [89] or in a dynamic [61] way. The concept of relay nodes can also be used in the case when no information on the nodes is available [63].

Evaluating the performance of relay protocols (message delivery ratio, message latency, throughput, etc.) is a difficult task due to the inherent complexity of mobile ad hoc networks, particularly the random nature of both the movement of the nodes and of the demand (traffic). The performance of mobile ad hoc networks are in general studied via lengthy and complex simulations, for a limited number of mobility models, including the random waypoint mobility model [22, 55] or the random direction mobility model [11, 46] which are described in, respectively, sections 3.3.1.1 and 3.3.1.2.

In this chapter a simple stochastic model is introduced to evaluate the performance of relay protocols for mobile ad hoc networks. The model has only two input parameters: the number of nodes in the network and the intensity (λ) of some identical and independent Poisson processes. In particular, the model does not require the distribution of the stationary distribution of the location of the nodes as input.

These processes model instances, called *meeting times*, at which any pair of nodes come within transmission range of one another. Transmissions between two nodes can only take place at meeting times and are assumed to be instantaneous. The latter assumption models the situation where the transmission time of a message is very small with respect to the time needed for two nodes to meet. Therefore, the random nature of a MANET is captured in the model through a finite number of these independent and homogeneous Poisson processes.

The selection of the intensity λ will be discussed in Sections 3.3 and 3.4 and is the topic of Chapter 4.

The model is used to characterize the message delay between two arbitrary nodes—hereafter called the *source node* and the *destination node*—for two relay protocols and for three mobility models. The two relay protocols are the *two-hop relay* protocol and the *unrestricted relay* protocol.

In the two-hop relay protocol the source node may forward the message to all the nodes it meets along its route, including of course the destination node. Any node which has received the message from the source node may only forward it to the destination node.

In the unrestricted relay protocol the source node may forward a message to all the nodes it meets (as in the two-hop relay protocol), but in this protocol any node that carries the message may in turn forward the message to all the nodes *it* encounters, along *its* trajectory.

The three mobility models that are considered in this chapter are the random waypoint, the random direction, and the two-dimensional random walker mobility models. All three models and their mathematical properties will be described in Section 3.3.1.

The characterization of the message delay in MANET has already received some attention. In [71] it is shown that, under the two-hop relay protocol, the expected message delay is of the order $T_p(n)n$ for the random waypoint mobility model on a sphere (where n is the number of nodes per unit area and $T_p(n)$ is the transmission time of a message). It is shown in the same paper that the expected message delay is of the order $T_p(n) \log^2(n)$ when nodes execute independent Brownian motions on a sphere. In Chapter 1 (or in [41]) the expected message delay is computed for a one-dimensional network topology, where the nodes move in adjacent segments according to independent and reflected Brownian motions.

This chapter is organized as follows: the stochastic model is introduced in Section 3.2.1, then we compute in Section 3.2.2 the Laplace-Stieltjes Transform (LST) of the message delay (Theorem 3.2.1). In this theorem the distribution is also obtained of the number of copies of the message at the time the message is delivered to the destination node. In Theorem 3.2.2 the expected message delay is given in a closed-form expression. From this, an asymptotic expansion of the expected message delay for a large number of nodes derived. These calculations are done for the two relay protocols.

In Section 3.3, the expected message delay and the distribution of the number of copies of the message found in Section 3.2 are compared to results obtained by simulations. Simulations have been carried out for each of the six combinations of the two relay protocols and the three mobility models. The simulation results are very close to the analytical results. We observed discrepancies only when the node transmission range is large with respect to the size of the area in which the nodes move.

The model assumptions have been validated in Section 3.3 in the absence of interferences (a situation that will typically occur when the communication radius of the nodes is small with respect to the area in which the nodes move, and the node density is also small). One way to incorporate interferences into the model is to thin the meeting time sequences: with some (independent) probability p (resp. $1 - p$) a transmission occurring at a meeting time will be a success (resp. failure). Due to the fact that a thinned Poisson process is again a Poisson process, it is enough to replace λ by λp , with p the probability that a communication fails due to interferences. On-going research will be devoted to this issue.

On the other hand, one can also argue that the communication radius of the nodes must be small enough so that interferences remain at an acceptable level. It has been shown

in [45] [72, Lemma 1] that the transmission range of the nodes should be of the order $1/\sqrt{N}$ for the two-hop relay protocol, in order to maintain a constant capacity per node (with N the number of nodes per unit area). In Section 3.4 it is shown how for a transmission range that is of the order of $1/\sqrt{N}$ the model can be used to compute the expected message delays for large networks and for the two relay protocols considered.

A word on the notation: given a function $g(N)$, one writes $f(N) = \mathcal{O}(g(N))$ if $|f(N)/g(N)|$ is bounded from above as $N \rightarrow \infty$ and $f(N) = o(g(N))$ if $f(N)/g(N) \rightarrow 0$ as $N \rightarrow \infty$.

3.2 The stochastic model

Consider a network with $N + 1$ identical mobile nodes. There is a *single message* to be delivered by a source node to a destination node. Intermediary nodes can be used as relay nodes. The goal is to determine the distribution of the message delay and the distribution of the number of copies of the message at the time the message is delivered to the destination node.

First the model is introduced (Section 3.2.1) and then in Section 3.2.2 it is used to evaluate the performance (message delay, number of copies) of the two-hop relay and the unrestricted relay protocols.

3.2.1 Definition of the model

An analytical model that would carefully take into account the main features of a MANET (transmission range, mobility pattern, interferences, fading, etc.) would be mathematically intractable. Instead, a model is proposed where the impact of these features is captured through a single parameter (the parameter λ , see below).

Let $0 \leq t_{i,j}(1) < t_{i,j}(2) < \dots$ be the successive *meeting times* between nodes i and j ($i \neq j$). Define $\tau_{i,j}(n) := t_{i,j}(n+1) - t_{i,j}(n)$, the n -th *inter-meeting time* between nodes i and j .

Transmissions between two nodes may only take place at meeting times and are assumed to be *instantaneous*. The latter assumption covers the situation where the transmission time of a message between two nodes is negligible with respect to the node inter-meeting times.

Assume that *if* a transmission takes place between node i and node j (at some meeting time $t_{i,j}(n)$) then it will be *successful*. Assume that it is node i that carries the message just before time $t_{i,j}(n)$. Under the two-hop relay protocol node i will transmit (a copy of) the message to node j at time $t_{i,j}(n)$ if i is the source node or if j is the destination node. Under the unrestricted relay protocol node i will always transmit the message to node j at time $t_{i,j}(n)$ regardless of the identity of node j .

Throughout this chapter the following assumption will hold for each relay protocol:

- (A) the processes $\{t_{i,j}(n), n \geq 1\}$, $1 \leq i, j \leq N+1$, $i \neq j$, are mutually independent and homogeneous Poisson processes with rate $\lambda > 0$. Equivalently stated, the random variables (rvs) $\{\tau_{i,j}(n)\}_{i,j,n}$ are mutually independent and exponentially distributed with mean $1/\lambda$.

Introduce:

- T_2 (resp. T_U), the message delay under the two-hop (resp. unrestricted) relay protocol, defined as the time needed to send the message (or a copy of the message) from the source to the destination;
- $N_2 \in \{1, 2, \dots, N\}$ (resp. $N_U \in \{1, 2, \dots, N\}$), the number of copies of the message in the network (including the original message but excluding the message at the destination node) at the time the message is delivered to the destination node.

For $\theta \geq 0$ let

$$T_2^*(\theta) := E[e^{-\theta T_2}], \quad T_U^*(\theta) := E[e^{-\theta T_U}]$$

be the LST of T_2 and T_U , respectively.

3.2.2 Performance of relay protocols

Theorem 3.2.1 gives, for each relay protocol, the LST of the message delay and the distribution of the number of copies.

Theorem 3.2.1 (LST of message delay)

Under the two-hop relay protocol

$$T_2^*(\theta) = \sum_{i=1}^N i \frac{(N-1)!}{(N-i)!} \left(\frac{\lambda}{\lambda N + \theta} \right)^i \quad (3.1)$$

and

$$P(N_2 = i) = \frac{i}{N^i} \frac{(N-1)!}{(N-i)!}, \quad i = 1, \dots, N. \quad (3.2)$$

Under the unrestricted relay protocol

$$T_U^*(\theta) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^i \frac{\lambda j(N+1-j)}{\lambda j(N+1-j) + \theta} \quad (3.3)$$

and

$$P(N_U = i) = \frac{1}{N}, \quad i = 1, \dots, N, \quad (3.4)$$

that is, the number of copies is uniformly distributed over $\{1, \dots, N\}$. \diamond

Proof. For both the two-hop and the unrestricted protocols the proof is based on modeling the number of copies in the network as an absorbing finite-state Markov chain. The transition rates of these Markov chains will differ for each protocol.

For each protocol the Markov chain takes its values in $\{1, 2, \dots, N-1, N, A\}$. The Markov chain is in state $i = 1, 2, \dots, N$ when there are i copies of the message in the network including the original message, and it is in state A when the message has been delivered to the destination node. Note that states $1, 2, \dots, N$ are transient states and A is an absorbing state.

A separate proof is provided for (3.1)-(3.2) and (3.3)-(3.4).

Proof of (3.1) and (3.2):

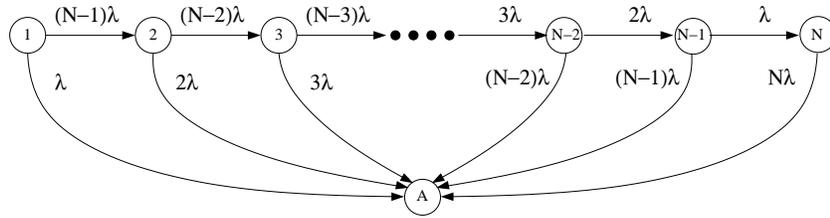


Figure 3.1: Two-hop relay: transition diagram of the Markov chain representing the number of copies

The transition diagram of the Markov chain corresponding to the two-hop relay protocol is given in Figure 3.1. Recall that under the two-hop relay protocol only the source node distributes copies of the message to nodes that come within its transmission range. Therefore, when there are i copies in the network, then either a new copy is sent to the $N - i$ nodes which do not have a copy yet, which occurs at the rate $\lambda(N - i)$ and triggers a transition from i to $i + 1$, or one of these i copies reaches the destination node, which occurs at the rate λi and triggers a transition from i to A . This explains the transition diagram in Figure 3.1.

The transition from i to $i + 1$ occurs with the probability $(N - i)\lambda / ((N - i)\lambda + i\lambda) = 1 - i/N$, and the transition from i to A occurs with the complementary probability $i\lambda / ((N - i)\lambda + i\lambda) = i/N$.

The sojourn time S_i in state $i = 1, 2, \dots, N$ is exponentially distributed with intensity λN (the sum of transition rates out of state i). Moreover S_1, \dots, S_N are mutually independent random variables.

By conditioning on the state of the Markov chain just before it enters state A , or equivalently by conditioning on the number of copies N_2 just before the message hits its

destination, we have

$$\begin{aligned} T_2^*(\theta) &= \sum_{i=1}^N \mathbb{E}[e^{-\theta T_2} | N_2 = i] P(N_2 = i) \\ &= \sum_{i=1}^N \mathbb{E}[e^{-\theta \sum_{j=1}^i S_j} | N_2 = i] P(N_2 = i). \end{aligned} \quad (3.5)$$

As mentioned earlier, $1 - j/N$ (resp. j/N) is the probability of jumping from state j to state $j + 1$ (resp. from state j to state A). Therefore,

$$P(N_2 = i) = \frac{i}{N} \prod_{j=1}^{i-1} \left(1 - \frac{j}{N}\right) = \frac{i}{N^i} \frac{(N-1)!}{(N-i)!}, \quad (3.6)$$

which establishes (3.2).

When in state $j = 1, 2, \dots, N$, the Markov chain can either enter state $j + 1$ after a time $S_{j,1}$ that is exponentially distributed with intensity $(N - j)\lambda$, or enter state A after a time $S_{j,2}$, independent of $S_{j,1}$, which is exponentially distributed with intensity $j\lambda$. Observe that $S_j = \min\{S_{j,1}, S_{j,2}\}$. Moreover

$$\begin{aligned} P[S_{j,1} < x | S_{j,1} < S_{j,2}] &= P[S_{j,2} < x | S_{j,1} > S_{j,2}] \\ &= P(S_j < x) \\ &= 1 - \exp(-\lambda x) \end{aligned} \quad (3.7)$$

as a consequence of the exponential distribution. Therefore,

$$\mathbb{E}[e^{-\theta \sum_{j=1}^i S_j} | N_2 = i] = \mathbb{E}[e^{-\theta(\sum_{j=1}^{i-1} S_{j,1} + S_{i,2})} | S_{j,1} < S_{j,2}, \dots, S_{i-1,1} < S_{i-1,2}, S_{i,1} > S_{i,2}] \quad (3.8)$$

From (3.7), (3.8) and the fact that the rvs $\{S_{j,k}\}_{j=1, \dots, N, k=1,2}$ are mutually independent, we find

$$\mathbb{E}[e^{-\theta \sum_{j=1}^i S_j} | N_2 = i] = \prod_{j=1}^i \mathbb{E}[e^{-\theta S_j}] = \left(\frac{\lambda N}{\lambda N + \theta}\right)^i. \quad (3.9)$$

Putting (3.5), (3.6) and (3.9) together yields

$$T_2^*(\theta) = \sum_{i=1}^N i \frac{(N-1)!}{(N-i)!} \left(\frac{\lambda}{\lambda N + \theta}\right)^i,$$

which proves (3.1).

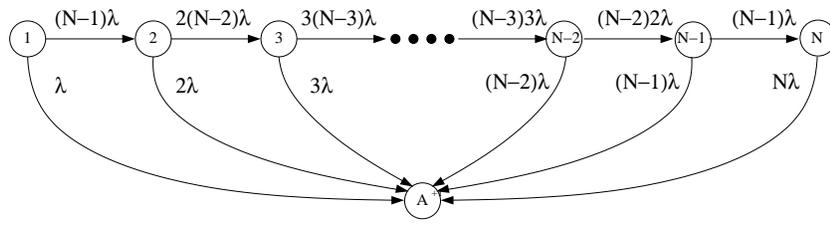


Figure 3.2: Unrestricted relay: transition diagram of the Markov chain representing the number of copies

Proof of (3.3) and (3.4):

The transition diagram of the Markov chain associated with the unrestricted relay protocol is displayed in Figure 3.2. Under this protocol, each node which has a copy of the message is allowed to distribute it to a node which does not have a copy and which comes within its transmission range. Therefore, when there are i copies of the message in the network a new copy is created at the rate $\lambda i(N - i)$ (transition from i to $i + 1$) and one of these i copies reaches the destination node at the rate λi (transition from i to A), as depicted on Figure 3.2.

The chain jumps from state i to state $i + 1$ with probability $(N - i)/(N + 1 - i)$ and it jumps from state i to state A with probability $1/(N + 1 - i)$. The sojourn time \tilde{S}_i in state i is exponentially distributed with intensity $\lambda i(N + 1 - i)$ (obtained as the sum of the transition rates going out state i).

By conditioning on the number of copies N_U , we have

$$T_U^*(\theta) = \sum_{i=1}^N \mathbb{E}[e^{-\theta \sum_{j=1}^i \tilde{S}_j} | N_U = i] P(N_U = i)$$

with

$$P(N_U = i) = \frac{1}{N + 1 - i} \prod_{j=1}^{i-1} \frac{N - j}{N + 1 - j} = \frac{1}{N},$$

which proves (3.4).

Similarly to (3.9) we have

$$\mathbb{E}[e^{-\theta \sum_{j=1}^i \tilde{S}_j} | N_U = i] = \prod_{j=1}^i \mathbb{E}[e^{-\theta \tilde{S}_j}] = \prod_{j=1}^i \frac{\lambda j(N + 1 - j)}{\lambda j(N + 1 - j) + \theta},$$

so that

$$T_U^*(\theta) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^i \frac{\lambda j(N + 1 - j)}{\lambda j(N + 1 - j) + \theta},$$

which proves (3.3). ■

Theorem 3.2.2 gives the expected message delay for each relay protocol. This result shows that for each protocol the expected message delay is a linear function of the expected inter-meeting time $1/\lambda$.

Theorem 3.2.2 (First two moments message delay)

Under the two-hop relay protocol, the expectation of the message delay is given by

$$\mathbb{E}[T_2] = \frac{1}{\lambda N} \sum_{i=1}^N \frac{i^2(N-1)!}{(N-i)! N^i} \quad (3.10a)$$

$$= \frac{1}{\lambda} \left(\sqrt{\frac{\pi}{2N}} + \mathcal{O}\left(\frac{1}{N}\right) \right). \quad (3.10b)$$

The second moment and the variance of this quantity are given by

$$\mathbb{E}[T_2^2] = \frac{1}{\lambda^2 N^2} \sum_{i=1}^N \frac{i^2(i+1)(N-1)!}{N^i(N-i)!} = \frac{1}{\lambda^2 N} \left(2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \right), \quad (3.11)$$

$$\text{Var}(T_2) = \frac{1}{\lambda^2 N} \left(\frac{4-\pi}{2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \right). \quad (3.12)$$

Under the unrestricted relay protocol, the expected message delay is given by

$$\mathbb{E}[T_U] = \frac{1}{\lambda N} \sum_{i=1}^N \frac{1}{i} \quad (3.13a)$$

$$= \frac{1}{\lambda N} \left(\log(N) + \gamma + \mathcal{O}\left(\frac{1}{N}\right) \right), \quad (3.13b)$$

where $\gamma \approx 0.57721$ is Euler's constant. The second moment is given by

$$\mathbb{E}[T_U^2] = \frac{1}{N\lambda^2} \sum_{i=1}^N \left[\left(\sum_{k=1}^i \frac{1}{k(N+1-k)} \right)^2 + \sum_{k=1}^i \left(\frac{1}{k(N+1-k)} \right)^2 \right] \quad (3.14)$$

Proof. The proof will make use of the following lemma which is proven in Appendix 3.A (for the sake of readability).

Lemma 3.2.1 *For large N*

$$\sum_{i=1}^N \frac{i^2 N!}{(N-i)! N^i} = N^{3/2} \sqrt{\frac{\pi}{2}} + \mathcal{O}(N), \quad (3.15a)$$

and

$$\sum_{i=1}^N \frac{i^3 N!}{(N-i)! N^i} = 2N^2 + \mathcal{O}(N^{3/2}). \quad (3.15b)$$

◇

Since $\mathbb{E}[T_2] = -\left. \frac{dT_2^*(\theta)}{d\theta} \right|_{\theta=0}$, the first part of $\mathbb{E}[T_2]$ can be derived at once from (3.1). Equation (3.10b) follows quickly with the help of (3.15a).

In the same way the second moment of the expected message delay under the two-hop relay protocol can be derived by taking $\mathbb{E}[T_2^2] = \left. \frac{\partial^2 T_2^*(\theta)}{\partial \theta^2} \right|_{\theta=0}$. We have

$$\frac{\partial^2 T_2^*(\theta)}{\partial \theta^2} = \sum_{i=1}^N i \frac{(N-1)!}{(N-i)!} \left(\frac{\lambda}{\lambda N + \theta} \right)^i \left(\frac{i(i+1)}{(\lambda N + \theta)^2} \right),$$

and so

$$\begin{aligned} \mathbb{E}[T_2^2] &= \left. \frac{\partial^2 T_2^*(\theta)}{\partial \theta^2} \right|_{\theta=0} \\ &= \frac{1}{\lambda^2 N^2} \sum_{i=1}^N \frac{i^2 (i+1) (N-1)!}{N^i (N-i)!}. \end{aligned}$$

With the help of the Lemma 3.2.1 it follows that

$$\mathbb{E}[T_2^2] = \frac{1}{\lambda^2 N^3} \left(2N^2 + \mathcal{O}(N^{3/2}) \right) = \frac{1}{\lambda^2 N} \left(2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \right).$$

This means that the variance is given by

$$\begin{aligned} \text{Var}(T_2) &= \frac{1}{\lambda^2 N} \left(2 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \right) - \frac{1}{\lambda^2} \left(\frac{\pi}{2N} + \mathcal{O}\left(\frac{1}{N^{3/2}}\right) \right) \\ &= \frac{1}{\lambda^2 N} \left(\frac{4-\pi}{2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \right). \end{aligned}$$

Similarly, we find by differentiating (3.3) with regards to θ , and then by setting $\theta = 0$, that

$$\begin{aligned}
 \mathbb{E}[T_U] &= \frac{1}{\lambda N} \sum_{i=1}^N \sum_{j=1}^i \frac{1}{j(N+1-j)} \\
 &= \frac{1}{\lambda N(N+1)} \sum_{i=1}^N \sum_{j=1}^i \left(\frac{1}{j} + \frac{1}{N+1-j} \right) \\
 &= \frac{1}{\lambda N(N+1)} \sum_{j=1}^N \sum_{i=j}^N \left(\frac{1}{j} + \frac{1}{N+1-j} \right) \\
 &= \frac{1}{\lambda N(N+1)} \sum_{j=1}^N \left(\frac{1}{j} + \frac{1}{N+1-j} \right) (N+1-j) \\
 &= \frac{1}{\lambda N} \sum_{j=1}^N \frac{1}{j},
 \end{aligned}$$

which is (3.13a). This last summation is known as the harmonic numbers. Its asymptotic expansion is [70, p. 186]

$$\sum_{j=1}^N \frac{1}{j} = \log(N) + \gamma + \mathcal{O}\left(\frac{1}{N}\right),$$

where γ is Euler's constant. This gives (3.13b).

To derive the second moment of the expected message delay under the unrestricted relay protocol we make use of

$$\begin{aligned}
 \frac{\partial^2 T_U^*(\theta)}{\partial^2 \theta} &= \frac{1}{N} \sum_{i=1}^N \left[\left(\sum_{k=1}^i \frac{-1}{\lambda k(N+1-k) + \theta} \right)^2 + \sum_{k=1}^i \left(\frac{1}{\lambda k(N+1-k) + \theta} \right)^2 \right] \\
 &\quad \cdot \prod_{j=1}^i \frac{\lambda j(N+1-j)}{\lambda j(N+1-j) + \theta}.
 \end{aligned}$$

This gives the second moment of T_U by taking $\theta = 0$. ■

The next result gives the number of occurrences of the message at the time the message is delivered.

Corollary 3.2.1 (Expected number of copies) *The expected number of copies under*

the two-hop relay protocol is given by (cf. (3.2))

$$\mathbb{E}[N_2] = \frac{1}{N} \sum_{i=1}^N \frac{i^2}{N^i} \frac{N!}{(N-i)!}. \quad (3.16)$$

Hence (cf. (3.10a))

$$\mathbb{E}[N_2] = \lambda N \mathbb{E}[T_2],$$

so that

$$\mathbb{E}[N_2] = \sqrt{\frac{\pi N}{2}} + \mathcal{O}(1). \quad (3.17)$$

The expected number of copies under the unrestricted relay protocol is (cf. (3.4))

$$\mathbb{E}[N_U] = \frac{N+1}{2}. \quad \diamond$$

The relative performance of the two-hop relay and unrestricted relay protocols can be captured through the ratios $\mathbb{E}[T_U]/\mathbb{E}[T_2]$ and $\mathbb{E}[N_U]/\mathbb{E}[N_2]$ given by (cf. Theorem 3.2.2)

$$\frac{\mathbb{E}[T_U]}{\mathbb{E}[T_2]} = \frac{N \sum_{i=1}^N \frac{1}{i}}{\sum_{i=1}^N \frac{i^2}{N^i} \frac{N!}{(N-i)!}}$$

and (cf. Corollary 3.2.1)

$$\frac{\mathbb{E}[N_U]}{\mathbb{E}[N_2]} = \frac{N(N+1)}{2 \sum_{i=1}^N \frac{i^2}{N^i} \frac{N!}{(N-i)!}},$$

respectively. Note that both ratios are independent of λ .

By using the asymptotic expansions (3.10b), (3.13b) and (3.17), we see that

$$\frac{\mathbb{E}[T_U]}{\mathbb{E}[T_2]} \approx \frac{\log(N)}{\sqrt{N}} \sqrt{\frac{2}{\pi}}$$

and

$$\frac{\mathbb{E}[N_U]}{\mathbb{E}[N_2]} \approx \sqrt{\frac{N}{2\pi}},$$

for large N . In other words, for large N the expected message delay under the unrestricted relay protocol is approximately $\log(N)/\sqrt{\pi N/2}$ times smaller than under the two-hop relay protocol, while the expected number of copies is approximately $\sqrt{N/2\pi}$ times larger.

For instance, for $N = 10^3$ then $\mathbb{E}[T_U]/\mathbb{E}[T_2] \approx 0.17$ and $\mathbb{E}[N_U]/\mathbb{E}[N_2] \approx 12.6$.

3.3 Applications

This section is devoted to the application of the results in Section 3.2 to three different mobility models. It is structured as follows: the mobility models are presented in Section 3.3.1 and the simulation setting for each mobility model is introduced in Section 3.3.2. Through both intuitive reasoning and simulations it is shown in Section 3.3.3 that assumption (A) is reasonable when the transmission range is not too large relative to the surface area. Based on this observation, estimates are obtained for the meeting rate λ , for each mobility model and for various transmission ranges. With the help of these estimates and Theorem 3.2.2, the expected message delays predicted by the analytical model are computed for each mobility pattern, for both the two-hop relay and for the unrestricted relay protocols, and are compared to simulation results.

The accuracy of the model is demonstrated in Sections 3.3.4 and 3.3.5, where the expected message delay and the distribution of the number of copies obtained by the model are compared to simulation results.

3.3.1 Mobility models

Although the results in Section 3.2 hold regardless of the dimension of the space in which the nodes move, in the following we shall only apply them to three standard *two-dimensional* mobility models: the random waypoint mobility model (Section 3.3.1.1), the random direction mobility model (Section 3.3.1.2), and the two-dimensional random walker mobility model (Section 3.3.1.3).

3.3.1.1 Random waypoint mobility model

The random waypoint mobility model was first introduced in [56] and later given its name in [22]. It is commonly used for the simulation of mobile ad hoc networks [55, 22]. Its properties have been studied extensively [13, 14, 12] and explicit expressions for the spatial node distribution, in terms on an integral which has to be evaluated (numerically), are given in [51, 52]. An initialisation procedure for simulations is given in [65].

In the random waypoint mobility model each node is assigned an initial *waypoint* location in a given area (typically a square) and travels at a constant speed S to a destination waypoint chosen uniformly in this area. The speed S is chosen uniformly in (v_{min}, v_{max}) , independently of the initial location and destination. After reaching the destination, the node may pause for a random amount of time after which a new destination and a new speed are chosen, independently of all previous destinations, speeds, and pause times. The stationary distributions of location and speed in the random waypoint mobility model differ significantly from the uniform distribution. In particular, it has been observed that the stationary distribution of the location of a node is more concentrated near the center of the region in which the nodes move [14]. Also, v_{min} needs to be strictly positive to ensure that the average speed over time does not go to zero [86].

3.3.1.2 Random direction mobility model

In the random direction mobility model [11, 46] each node is assigned an initial direction α , speed $S \in [v_{\min}, v_{\max}]$ in km/hour and a finite travel time τ . The node then travels in the direction α for a duration τ and at speed S . When the travel time has expired new speed and travel times are chosen at random independently of each other and of all previous directions, speeds and travel times. In addition, a new angle is chosen for the change of direction of travel. Depending on the definition, the new travel direction can also be chosen independent of the previous travel direction. When a node reaches a boundary it is either reflected [8, 11] or the area wraps around so that the node reappears on the other side [11].

The stationary distributions of the location and direction have been shown to be uniform [8, 64] for arbitrary direction, speed and travel time distributions, irrespective of the boundaries being reflecting or wrap around. This is in contrast with the random waypoint mobility model where nodes are more likely to be concentrated near the center of the area. Another difference is that the minimum speed v_{\min} does not have to be strictly positive. The speed can be equal to zero since the node maintains a certain speed only for a limited amount of time.

Notice that random direction mobility model is a generalization of the random walker: by fixed the travel time and choosing only a vertical or horizontal direction of travel one obtains a random walker. Notice also that the random direction can be 'scaled' to obtain a Brownian motion by sending the travel time to zero.

3.3.1.3 Random walker mobility model

In the two-dimensional Random Walker Mobility Model each node moves as a random walker on a two-dimensional square lattice. The time is discrete and at each time step each node has a probability of 1/4 of hopping to a position above, below, to the left, or to the right of its current position. If the node is positioned on a boundary, then instead of hopping off the lattice it hops back to the same state. This movement can be seen as someone wandering at a constant speed from intersection to intersection through a city, where all of the streets are equally spaced and perpendicular to each other (Manhattan network). The stationary distribution of the location of a two-dimensional random walker on a square lattice is uniform over the area. This properties is a consequence of the fact that a two-dimensional random walker can be constructed from two independent one-dimensional random walkers, and that the stationary location of a symmetric random walk in one dimension is uniform (see Section 1.2).

3.3.2 Simulation Setting

The numerical results presented hereafter are based on simulation programs in which mobiles nodes move in a square of size $4 \times 4\text{km}^2$.

For the random waypoint mobility model special care must be taken in the initialization process of the simulator since the starting positions are not uniform. A good procedure for

doing this, and which we have used, is presented in [65]. The method consists of sampling the initial speeds and locations from their stationary distributions. Then, subsequent speeds and locations are sampled from the uniform distribution. The pause times are taken to be equal to zero.

Since the stationary distribution of the location of a node is uniform for both the random direction and the random walker mobility models, their implementation does not pose any difficulty.

For the random waypoint mobility model and the random direction mobility model, a speed (in km/h) was chosen uniformly in $[v_{min}, v_{max}] = [4, 10]$ km/h.

In the random direction mobility model, a node moves in a direction that is uniformly distributed in $[0, 2\pi)$, for an exponentially distributed amount of time (expressed in hours) with mean $1/4$, and at a speed that is uniformly distributed in $[4, 10]$ km/h, before the node chooses a new direction, travel time and speed.¹

For the random walker mobility model we assume that the streets are 80 meters apart and that the random walkers move at the constant speed² of one block per minute (this results in $51^2 = 2601$ states and a constant speed of 4.8 km/h).

We assume that there are no inferences and that the transmission of a message between two nodes (in contact) is instantaneous. The former assumption typically models a situation where the transmission radius is small with respect to the size of the area. The latter assumption typically models a situation where the message transmission time between two nodes is negligible with respect to the node inter-meeting times. With these assumptions it is easy to derive the time between consecutive events; details of this are given in Appendix 3.B.

In order to apply the results in Section 3.2 we need, for each mobility model, to check the validity of assumption (A) and to identify the parameter λ of the exponential inter-meeting time distribution.

3.3.3 Validation of the Poisson meeting times

For each mobility model and for various communication radii, the movement of two nodes have been simulated and estimations have been made of the distribution (Section 3.3.3.1) and the auto-covariance function (Section 3.3.3.2) of the inter-meeting times between these two nodes. The results, based on 100,000 observations and which are reported below, show that the Poisson assumption for the meeting time sequences is valid for all three mobility models and for a large range of communication radii.

¹As a small remark it is worth mentioning that with the right choice of parameters the random direction mobility model can be used to simulate random walkers.

²In the previous chapter it was shown that due to time averages the underlying visit time distribution is of little or no importance for the message delay, hence a constant visit time can be taken.

3.3.3.1 Inter-Meeting Time Distribution

Figure 3.3 displays, on a log-scale for the y axis, the complementary cumulative distribution function (complementary cdf) of the inter-meeting time between two nodes for each mobility model and for three different communication radii ($R = 50m, 100m, 250m$).

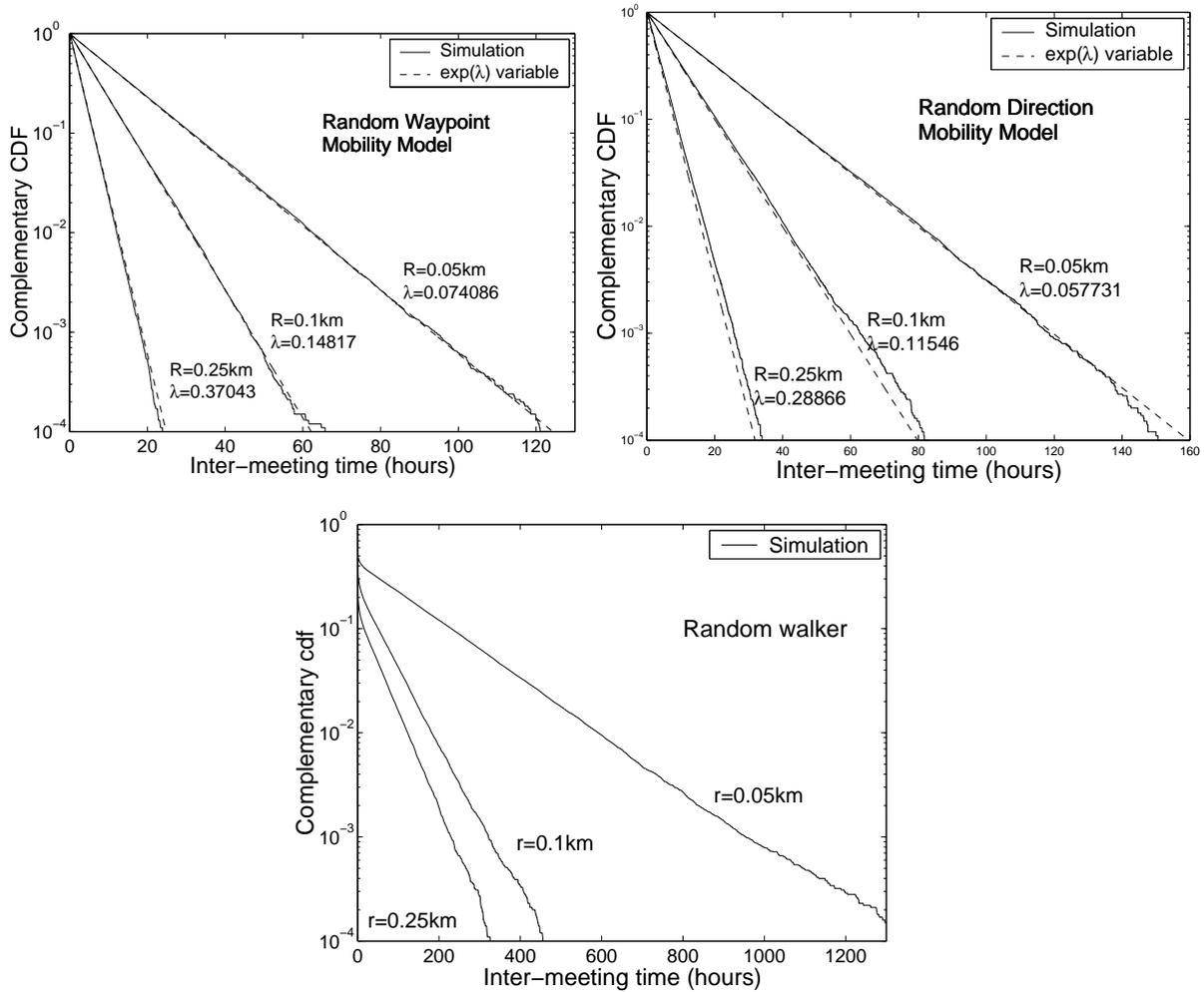


Figure 3.3: Complementary cdf of the inter-meeting time of two nodes for the random waypoint, the random direction, and the random walker mobility models.

For the random direction and the random waypoint mobility models, and for each communication radius, we have also plotted the complementary cdf of an exponential distribution (i.e. a straight line on a log-scale for the y axis) with intensity (i.e. slope) λ . We observe an excellent agreement between the estimated cdf (solid line) and the exponential cdf (dashed line) for the three different communication radii. Estimates for the value of λ for these two models will be derived in Chapter 4 and are given in Corollary 4.2.2.

For the random walker mobility model the situation is more complicated: the inter-meeting time is not exponential but there is an exponential tail which rapidly emerges.

The reason, we argue, is because random walkers have a tendency to “hang around” the same region (resulting in many small inter-meeting times). If, however, the two random walkers have wandered away from one another, then we find an exponential distribution for the time until they meet again. This explains why in [72] the first-meeting time (defined as the time between a random moment and the moment when two nodes meet) between two Brownian motions resembles an exponential distribution whereas the inter-meeting time does not.

Because of the quick emergence of the exponential tail for the random walker mobility, this movement pattern has been included into the analysis to see how robust the model is. An explicit expression for the exponential tail under the random walker mobility model is, to the best of our knowledge, not known and it is therefore obtained numerically as the complementary of the average first-meeting time obtained across all simulations.

The fact that, for each mobility model, the cdf of the inter-meeting distribution is well-approximated by an exponential distribution, at least for small to moderate transmission radii (with respect to the size of the area) finds its roots in the various independence assumptions placed on each mobility model. Indeed, nodes move independently of each other and future directions and speeds (and therefore locations) of a node are independent of past directions and speeds of this node. If we pick two mobile nodes at random at some stationary time, then there is a probability q that they will meet (in the sense of being within transmission range of one another) before the next change of direction of either node. At the next change of direction, because of the independent assumptions recalled above, the process repeats itself and there is a probability q that these nodes will meet before the next change of direction. This yields a geometric distribution for the number of changes of direction before both nodes meet. The exponential distribution pops up because the number of changes of direction is “linearly” related to the time traveled before the nodes meet.

The fact that inter-meeting times are exponentially distributed has already been observed for the Brownian motion and the random waypoint mobility model on a sphere [71]. This issue will be studied in more detail in Chapter 4. In particular, Theorem 4.2.1 provides the parameter of the exponential distribution and Corollary 4.2.2 gives a number of special cases for which simple closed-form expressions exist. Most notable is that λ has a linear relationship with the transmission radius (under the assumption that $R \ll L$). This is confirmed in Figure 3.4 where the estimate of λ based on simulations is plotted against the communication range R . Also shown in the figure are the theoretical values of λ obtained through Corollary 4.2.2. Not surprising, the value of λ is lower for nodes moving according to the random waypoint model than for the random direction mobility model since the nodes are more concentrated around the center of the region. The linear relationship between λ and R has also been observed for the random walkers, but for that mobility model the theoretical value of λ remains to be found.

Since the parameter λ describes the meetings time between two nodes, and interference was considered negligible, it means that λ is independent of the relaying protocol used.

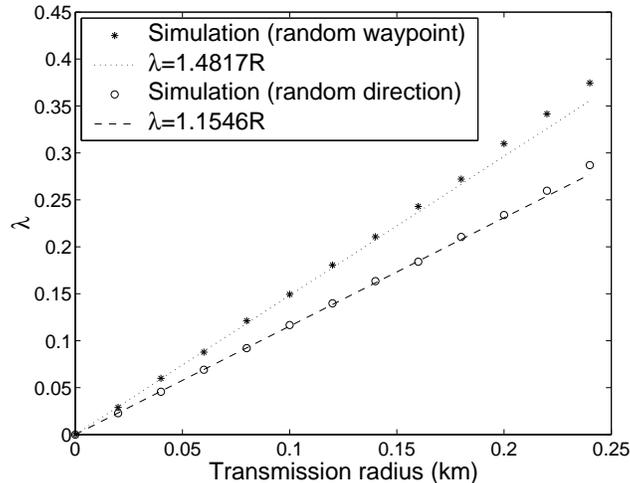


Figure 3.4: Relationship between the inter-meeting time intensity λ and the transmission radius R . Here $[v_{min}, v_{max}] = [4, 10]$ km/hour.

3.3.3.2 Independence of Inter-Meeting Times

Let $\{\tau(n)\}_n$ be the inter-meeting times between two given nodes. To check the assumption that the rvs $\{\tau(n)\}_n$ are mutually independent rvs, we have used the following classical estimator for the auto-correlation function of $\{\tau(n)\}_n$

$$\rho_m(h) = \frac{\gamma_m(h)}{\gamma_0(h)}, \quad h \geq 0,$$

where

$$\gamma_m(h) := \frac{1}{m} \sum_{n=1}^{m-h} (\tau(n+h) - \hat{\tau}^{(m)}) (\tau(n) - \hat{\tau}^{(m)})$$

is an estimator of the auto-covariance function, with $\hat{\tau}^{(m)} = (1/m) \sum_{n=1}^m \tau(n)$ the sample mean for m observations.

If the rvs $\{\tau(n)\}_n$ are mutually independent then their autocorrelation function is equal to zero for all $h \geq 1$.

The mapping $h \rightarrow \rho_m(h)$ corresponding to the random waypoint mobility model is plotted in Figure 3.5 for $m = 100,000$ and $R = 0.25km$. The autocorrelation functions corresponding to other values of R ($R = 0.05km$, $R = 0.1km$) and/or to the random direction mobility model and the random walker mobility model are not displayed since they are identical to the results in Figure 3.5. From these results we conclude that the assumption that the inter-meeting times between two nodes are mutually independent rvs is a reasonable assumption.

In conclusion, the results reported in Sections 3.3.3.1 and 3.3.3.2 validate the assumption that the meeting time process between two given nodes is a Poisson process for all three

mobility models and for small to moderate communication radii (with respect to the size of the area in which the nodes move).

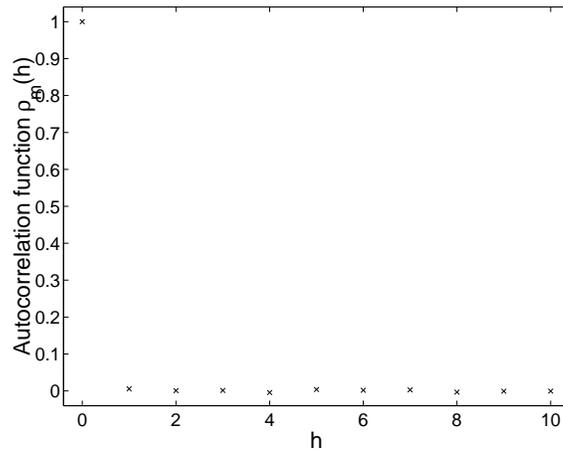


Figure 3.5: Autocorrelation function of inter-meeting times for the random waypoint mobility model with $R = 0.25\text{km}$.

3.3.4 Expected Message Delay

For the three mobility models introduced in Section 3.3.1 and for three communication radii ($R = 0.05\text{km}$, 0.1km , 0.25km), Figures 3.6-3.7 display the expected message delays obtained both through simulations and by the analytical model as a function of the number of nodes. Results for the two-hop (resp. unrestricted) relay protocol are given in Figure 3.6 (resp. Figure 3.7).

These results demonstrate the ability of the analytical model to predict the expected message delay under both the two-hop relay protocol and the unrestricted relay protocol for different mobility patterns, across any number of nodes and communication radii.

3.3.5 Distribution of Number of Copies

Figures 3.8-3.10 compare the distribution of the number of copies at message delivery time obtained through simulations (represented by bars in the figures) and by the analytical model (solid lines), under both relay protocols and for 40 nodes (i.e., $N = 39$).

Results for the two-hop relay protocol are displayed in Figures 3.8-3.9 for the random waypoint mobility model and the random walker mobility model, respectively (results for the random direction mobility models are identical to that of the random waypoint mobility model and have not been displayed). We observe that for all three mobility models the fit is quite good when $R = 50\text{m}$ and that it deteriorates as R increases (although the results are still acceptable for $R = 100\text{m}$ for the random waypoint mobility model and the random direction mobility model).

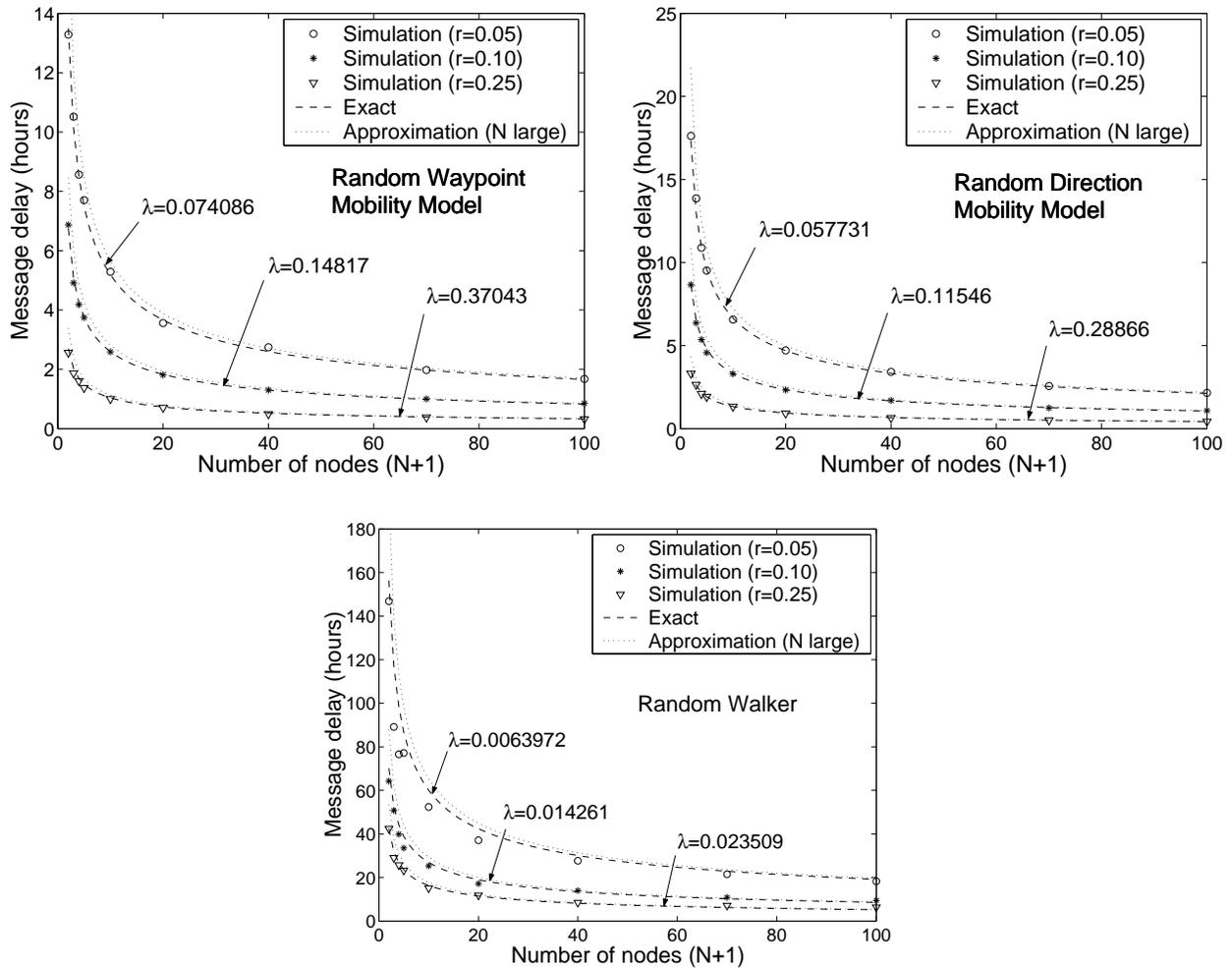


Figure 3.6: Expected message delay as a function of the number of nodes: the two-hop relay protocol

Results for the unrestricted relay protocol are reported in Figure 3.10. Recall that for this protocol the number of copies is uniformly distributed in the analytical model, namely, $P(N_U = i) = 1/39 = 0.0256$ for all $i = 1, \dots, 39$ (see Theorem 3.2.1). Results are displayed for each mobility model, each for a different transmission range. We can see that in all cases the distribution of the number of copies is very close to the uniform distribution.

These results give a good indication that our model, despite its genericness, is able to capture the main features of the interaction of the mobility models and the relay protocols.

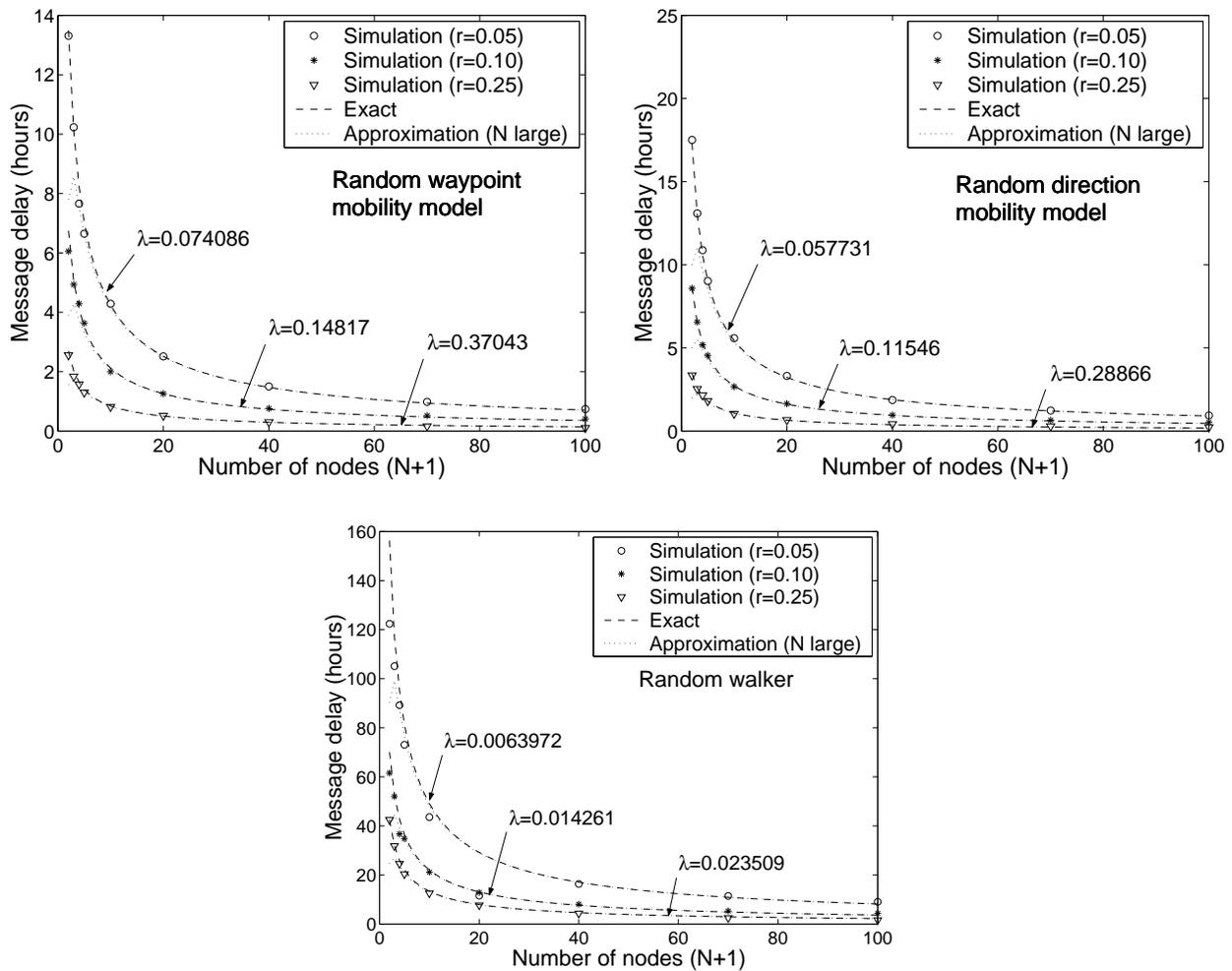


Figure 3.7: Expected message delay as a function of the number of nodes: the unrestricted relay protocol

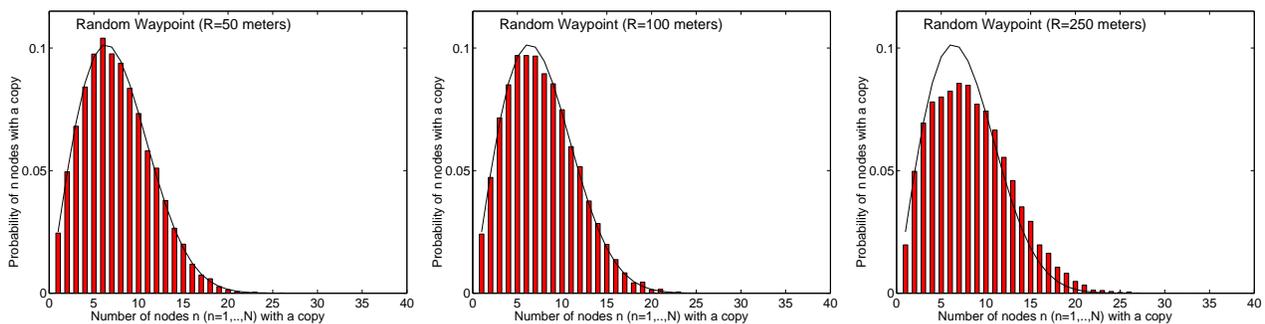


Figure 3.8: Distribution of the number of copies: the two-hop relay protocol under the random waypoint mobility model

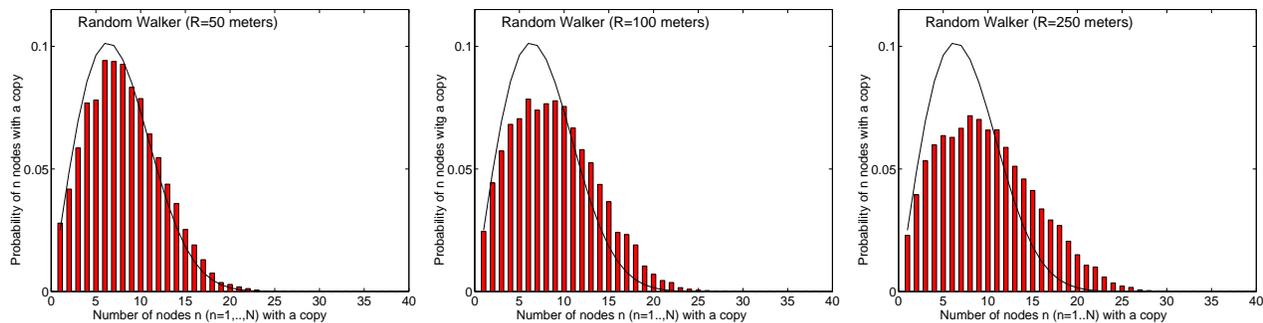


Figure 3.9: Distribution of the number of copies: the two-hop relay protocol under the random walker mobility model

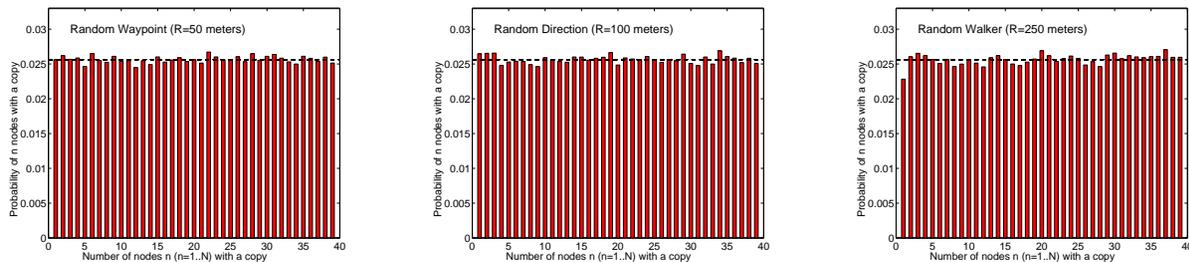


Figure 3.10: Distribution of the number of copies: the unrestricted relay protocol

3.4 Large networks

We have observed in Section 3.3 that, for the three mobility models considered in this chapter, the inter-meeting time intensity (λ) is well approximated by a linear function of the transmission range R . This approximation is valid as long as R is not “too large” with respect to the size of the area in which the nodes move.

On the other hand, when the number of nodes increases R should decrease to prevent interferences from becoming excessive.

Putting these two observations together, yields

$$\lambda = \mathcal{O}(R(N)),$$

where $R(N)$, the transmission range for a network with N nodes, is a decreasing function of N .

Introducing this behaviour of λ in Theorem 3.2.2 immediately gives the following:

Corollary 3.4.1 *For large N*

$$\mathbb{E}[T_2] = \mathcal{O}\left(\frac{1}{\sqrt{N}R(N)}\right) \quad \text{and} \quad \mathbb{E}[T_U] = \mathcal{O}\left(\frac{\log(N)}{NR(N)}\right). \quad \diamond$$

If we choose $R(N) = \mathcal{O}(1/\sqrt{N})$ in order to keep interference at an acceptable level (it is shown in [72, Lemma 1] that $R(N) = \mathcal{O}(1/\sqrt{N})$ in order to achieve a constant capacity per node with the two-hop relay protocol), then

$$\mathbb{E}[T_2] = \mathcal{O}(1) \quad \text{and} \quad \mathbb{E}[T_U] = \mathcal{O}\left(\frac{\log(N)}{\sqrt{N}}\right),$$

by using Corollary 3.4.1.

Alternatively, one may want to find the function $R(N)$ so that the expected message delay for unrestricted relaying is $\mathcal{O}(1)$ as the number of nodes becomes large. This is achieved when $R(N) = \mathcal{O}(\log(N)/N)$ leading to $\mathbb{E}[T_2] = \mathcal{O}(\sqrt{N}/\log(N))$.

Remark 3.4.1 *Let T_R be the message delay under the two-hop relay protocol. Recall that in this protocol a message is relayed instead of copied [45] [71] [72]. Similar to the analysis conducted in Section 3.2, it can be shown that $\mathbb{E}[T_R] = \lambda^{-1}(1 + 1/N - 1/N^2) = \lambda^{-1}(1 + \mathcal{O}(1/N))$. Since $\lambda = \mathcal{O}(r(N))$ for the random waypoint model, and with the scaling $r = \mathcal{O}(1/\sqrt{N})$, we find that the expected message delay under the two-hop relay protocol is $\mathcal{O}(\sqrt{N})$, just as was found in [71] but for nodes moving on a sphere.³*

3.5 Concluding Remarks

In the first three chapters of this thesis the message delay was studied for a number of different movement patterns. Some of these results which followed had to be calculated separately for each mobility model whereas a number of other results were generic of nature. The two most noteworthy generic results are the exponential inter-meeting times for $R \ll L$ and Theorem 3.2.1 for the LST of the message delay. In the next chapter it will be studied for which class of mobility models these last two important results hold and in particular the value λ will be derived for the random direction and the random waypoint mobility model.

An important reason for studying which class of movement patterns result in exponential first-meeting times lies in the complexity of the first-meeting times. As Chapters 1 and 2 show, obtaining this measure for two relatively simple movement patterns in one direction already proves to be quite challenging. Obtaining exact closed-form results in two dimensions and for more than two mobiles does not seem feasible, hence the need to see when the exponential approximations hold.

The generic model of this chapter can be used to evaluate and compare the performance of different routing protocols for message delivery in MANET for a wide range of mobility models. Future research will deal with the study of message delivery within a certain timeframe and for nodes which are not always turned on.

³The Brownian motion mobility model was studied in [72], whereas in this paper we consider the random walker mobility model. To go from the latter movement to the former care must be taken in terms of the limit and the metric under consideration (expected delay).

In this work it was assumed that there was a single message and that communications between other nodes do not interfere with the transfer of that single message. The next step requires the study of simultaneous communications that take the queueing at the relay nodes into account.

3.A Proof of Lemma 3.2.1

Proof. We start with proving (3.15a). Define $A(N) := \sum_{i=1}^N \frac{i^2 N!}{(N-i)! N^i}$.

If it were not for the presence of the factor i^2 in $A(N)$, then this quantity would be the Ramanujan Q-distribution [70, page 188], also known as the birthday function. This function often shows up in the analysis of algorithms.

The derivation of the approximation (3.15a) follows that of the Ramanujan Q-distribution approximation [70, Proposition 4.8]. We now outline it.

Let $i_0 := \lfloor N^{3/5} \rfloor$. This implies that $i_0^2/N \rightarrow \infty$ as $N \rightarrow \infty$ and $i_0 = o(N^{2/3})$. We have

$$A(N) = \sum_{i=1}^{i_0} \frac{i^2 N!}{(N-i)! N^i} + B(N),$$

with $B(N) := \sum_{i=i_0+1}^N \frac{i^2 N!}{(N-i)! N^i}$.

$B(N)$ is an exponentially small function of N , in the sense that $B(N)$ is $\mathcal{O}(1/N^a)$ for any $a > 0$. The proof of this result goes as follows. It is shown in the proof of Proposition 4.8 in [70] that $C(N) := \sum_{i=i_0+1}^N \frac{N!}{(N-i)! N^i}$ is an exponentially small quantity. On the other hand, $B(N) \leq N^2 C(N)$, from which we conclude that $B(N)$ is exponentially small since the product of an exponentially small quantity and any polynomial in N remains an exponentially small quantity [70, Exercise 4.10, p. 158].

Therefore,

$$A(N) = \sum_{i=1}^{i_0} \frac{i^2 N!}{(N-i)! N^i} + \Delta(N),$$

where $\Delta(N)$ represents a function which is exponentially small.

For any integer i that is $o(N^{2/3})$ it is shown in [70, Proposition 4.4] that

$$\frac{N!}{(N-i)! N^i} = e^{-i^2/(2N)} \left(1 + \mathcal{O}\left(\frac{i}{N}\right) + \mathcal{O}\left(\frac{i^3}{N^2}\right) \right). \quad (3.18)$$

Since $i = o(N^{2/3})$ whenever $1 \leq i \leq i_0$, we deduce from (3.18) that

$$A(N) = \sum_{i=1}^{i_0} i^2 e^{-i^2/(2N)} \left(1 + \mathcal{O}\left(\frac{i}{N}\right) + \mathcal{O}\left(\frac{i^3}{N^2}\right) \right) + \Delta(N).$$

By applying the Euler-MacLaurin summation [70, Theorem 4.2] to the functions $x^3 e^{-x^2/2}$ and $x^5 e^{-x^2/2}$ we find that (see [70, Exercise 4.69] for similar results)

$$\sum_{i=1}^{i_0} i^2 e^{-i^2/(2N)} \mathcal{O}\left(\frac{i}{N}\right) = \mathcal{O}(N)$$

and

$$\sum_{i=1}^{i_0} i^2 e^{-i^2/(2N)} \mathcal{O}\left(\frac{i^3}{N^2}\right) = \mathcal{O}(N),$$

respectively. Hence,

$$A(N) = \sum_{i=1}^{i_0} i^2 e^{-i^2/(2N)} + \mathcal{O}(N). \tag{3.19}$$

By noting that $i^2 e^{-i^2/(2N)}$ is exponentially small for $i > i_0$ we can add all terms for $i > i_0$ into the right hand side of (3.19), which gives

$$A(N) = \sum_{i \geq 1} i^2 e^{-i^2/(2N)} + \mathcal{O}(N). \tag{3.20}$$

The above summation is the summation of the function $Nx^2 e^{-x^2/2}$ at regularly spaced points with step $1/\sqrt{N}$. Another application of the Euler-MacLaurin formula [70, Proposition 4.2] yields

$$\begin{aligned} \sum_{i \geq 1} i^2 e^{-i^2/(2N)} &= N^{3/2} \int_0^\infty x^2 e^{-x^2/2} dx + \mathcal{O}(N) \\ &= N^{3/2} \sqrt{\frac{\pi}{2}} + \mathcal{O}(N), \end{aligned} \tag{3.21}$$

so that

$$A(N) = N^{3/2} \sqrt{\frac{\pi}{2}} + \mathcal{O}(N)$$

from (3.20) and (3.21), which concludes the proof of the first part of the lemma.

The proof (3.15b) is done in a similar way to shown that

$$\begin{aligned} \sum_{i \geq 1} \frac{i^3 N!}{(N-i)! N^i} &= \sum_{i \geq 1} i^3 e^{-i^2/(2N)} + \mathcal{O}(N^{3/2}) \\ &= N^2 \int_0^\infty x^3 e^{-x^2/2} dx + \mathcal{O}(N^{3/2}) \\ &= 2N^2 + \mathcal{O}(N^{3/2}). \end{aligned}$$

■

3.B Derivation of the time until the next event

For the purpose of numerical analysis an event-driven simulation program was written. An event can be:

- the change of direction or position of a node;
- two nodes coming within range of each other;
- two nodes getting out of range of each other.

The calculation of the time until one of the last two events happens is done the next paragraph.

Calculation of the time until two nodes meet.

Since the simulation program is event-driven it is necessary to calculate the time between the change of direction of a node and the time of a possible meeting with another node before its next change of direction. To derive this time take a look at Figure 3.11.

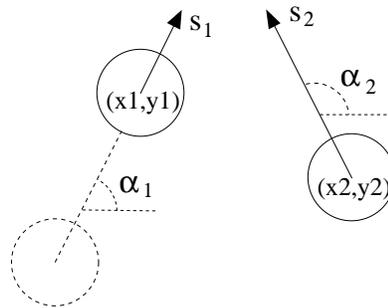


Figure 3.11: Illustration of the meeting of two nodes.

Assume the second nodes changes direction at time $t = 0$. Let the position of node $i = 1, 2$ at that moment be given by (x_i, y_i) , the direction by α_i , and the speed by s_i . The position of node i at time $t \geq 0$ is

$$x_i(t) = x_i + t \cdot s_i \cdot \cos(\alpha_i), \quad (3.22a)$$

$$y_i(t) = y_i + t \cdot s_i \cdot \sin(\alpha_i). \quad (3.22b)$$

The mobiles meet or separate from each other at time t if the distance between them is exactly R . These times are given by the solution to $R = \sqrt{(x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2}$.

Filling in (3.22) gives

$$\begin{aligned}
R^2 &= \left(x_1 + t \cdot s_1 \cdot \cos(\alpha_1) - x_2 + t \cdot s_2 \cdot \cos(\alpha_2) \right)^2 + \left(y_1 + t \cdot s_1 \cdot \sin(\alpha_1) - y_2 + t \cdot s_2 \cdot \sin(\alpha_2) \right)^2 \\
&= \left((x_1 - x_2) + t(s_1 \cos(\alpha_1) - s_2 \cos(\alpha_2)) \right)^2 + \left((y_1 - y_2) + t(s_1 \sin(\alpha_1) - s_2 \sin(\alpha_2)) \right)^2 \\
&= (x_1 - x_2)^2 + (y_1 - y_2)^2 \\
&\quad + 2t \left((x_1 - x_2)(s_1 \cos(\alpha_1) - s_2 \cos(\alpha_2)) + (y_1 - y_2)(s_1 \sin(\alpha_1) - s_2 \sin(\alpha_2)) \right) \\
&\quad + t^2 \left((s_1 \cos(\alpha_1) - s_2 \cos(\alpha_2))^2 + (s_1 \sin(\alpha_1) - s_2 \sin(\alpha_2))^2 \right).
\end{aligned}$$

Solving this second-degree polynomial in t gives

$$\begin{aligned}
t_1 &= \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \\
t_2 &= \frac{-b + \sqrt{b^2 - 4ac}}{2a},
\end{aligned}$$

with

$$\begin{aligned}
a &:= h_1^2 + h_2^2, & h_1 &:= s_1 \cos(\alpha_1) - s_2 \cos(\alpha_2), \\
b &:= 2(x_1 - x_2)h_1 + 2(y_1 - y_2)h_2, & h_2 &:= s_1 \sin(\alpha_1) - s_2 \sin(\alpha_2), \\
c &:= (x_1 - x_2)^2 + (y_1 - y_2)^2 - R^2.
\end{aligned}$$

Here t_1 and t_2 , respectively, give the moments in time when the mobiles come into contact with one another and when they loose contact with each other. If

- $b^2 - 4ac < 0$: the two nodes never meet;
- $t_1 < 0$ and $t_2 < 0$: their hypothetical meeting and separation took place in the past;
- $t_1 < 0$ and $t_2 \geq 0$: they are in contact at time $t = 0$ and will separate at time t_2 ;
- $t_1 \geq 0$ and $t_2 \geq 0$: they will meet at time t_1 and separate at time t_2 .

Finally, the times t_1 and t_2 can be compared to the time until the next change of direction of either node, and then the appropriate event can be scheduled.

λ : The Mysterious Parameter

Contents

4.1 Introduction	92
4.2 The values for λ	93
4.3 Examples	98
4.4 Implications for the message delay	98
4.5 Concluding remarks	100
4.A Proof of Theorem 4.2.1.	101
4.B Proof of Proposition 4.2.1.	103
4.C Proof of Proposition 4.2.2.	104
4.D Probability of contact at a random moment	105

In this chapter it is proven that the first-meeting times and the inter-meeting times between two nodes moving under the random direction and the random waypoint mobility model are exponentially (λ) distributed. This result is shown under the hypothesis that there is no interference and that the fixed transmission range (r) is small compared to the length (L) of the square area the nodes move on, i.e., $r \ll L$. The parameter λ is identified for each of the models and in the special case where the speeds of the two nodes are constant and identical a simple expression reveals itself. The obtained expressions are validated against numerical results.

The obtained expressions can be used in the previous chapter where—assuming exponential (λ) inter-meeting times—the message delay was expressed as a function of λ and the number of nodes.

4.1 Introduction

The *first-meeting* time is defined as the time, starting from a random moment in time, until two nodes meet. The *inter-meeting* time is defined as the time that passes between two consecutive meetings. It is assumed that each node has the same transmission range r .

An interesting property emerges for the random direction and the random waypoint mobility models in two-dimensions: the first-meeting and the inter-meeting times are (approximately) exponentially distributed as long as the fixed transmission radius is small compared to the length (L) of the square surface area the nodes travel on, i.e., $r \ll L$. This is best seen in Figure 4.1 where the complementary cdf of the first meeting times is shown for both mobility models, $L = 4$ km, and for three different transmission radii ($r = 50$, $r = 100$, and $r = 250$ meters). The speeds at each change of direction are chosen from $[v_{min}, v_{max}] = [4, 10]$ km/hour. The variation in the tails are artefacts of simulations. The figures for the inter-meeting times are virtually identical and are given in Figure 4.5 (the reason for this is given in Section 4.4).

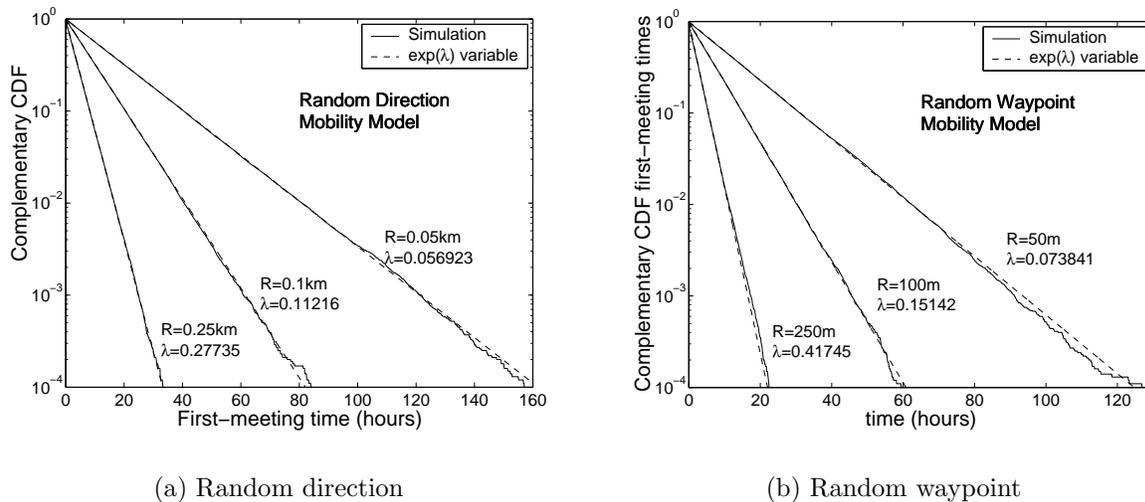


Figure 4.1: Complementary cdf of the first-meeting time.

Note the log-scale on the vertical axis. Alongside each of the simulated values is also a straight line which corresponds to an $\exp(\lambda)$ distribution, where $\lambda = 1/\text{mean}(\text{First-meeting times})$ obtained through the simulations. The variances in the tails are artifacts of simulations.

In the next section the values of λ will be derived for both of the mobility models, whereafter in Section 4.3 examples will be given. The obtained expressions are then linked in Section 4.4 to the results of Chapter 3 after which concluding remarks are given in Section 4.5.

4.2 The values for λ

The following theorem, which holds for *any* mobility model with the stated assumptions, is central to this chapter and provides us with a basis for the estimates of λ for the random direction and the random waypoint mobility models. The proof is given in Appendix 4.A.

Theorem 4.2.1 (First-meeting time distribution) *Let two nodes move independently of each other in a square of size $L \times L$ with speeds V_1 and V_2 . Let $\mathbb{E}[V_*]$ be the average relative speed between the two nodes, and let $\pi(x, y)$ be the distribution of the node location in steady-state. If the transmission range $r \ll L$ and the location of a node at time t is independent of its location at time $t + \Delta_t$, for some small Δ_t , then the first-meeting time (F) between the nodes is exponentially distributed with parameter λ , where λ is given by*

$$\lambda \approx 2r\mathbb{E}[V_*] \int_0^L \int_0^L \pi^2(x, y) dx dy. \quad (4.1)$$

The average time until the two nodes meet is

$$\mathbb{E}[F] = \frac{1}{\lambda} \approx \frac{1}{2r\mathbb{E}[V_*] \int_0^L \int_0^L \pi^2(x, y) dx dy}. \quad \diamond$$

Although the assumption in the above theorem at first glance seems very strong, for mobility patterns which “move around” fast enough it provided excellent approximations. For example, consider the random waypoint mobility model: knowing a node’s current position provides us some information on where the node may have been located a little moment earlier, but it provides us very little, almost no, information about the positions where the node has previously changed direction. Hence we see that there is an independence property over time.

In order to apply Theorem 4.2.1 to the random waypoint and the random direction mobility models we first need to know two measures, namely the integral over the square of the spatial node distribution and the average relative speed, $\mathbb{E}[V^*]$. These two quantities are summarized in the following propositions for which the proofs are forwarded to Appendixes 4.B and 4.C.

Proposition 4.2.1 *Let $\pi(x, y)$ be the pdf of the spatial node distribution of a node moving in a square of size $L \times L$. We have*

$$\int_0^L \int_0^L \pi^2(x, y) dx dy = 1/L^2, \quad (4.2a)$$

for nodes moving according to the random direction mobility model and

$$\int_0^L \int_0^L \pi^2(x, y) dx dy \approx 1.3683/L^2, \quad (4.2b)$$

for nodes moving according to the random waypoint mobility model. The constant $\omega \approx 1.3683$ shall be referred to as the waypoint constant . \diamond

Proposition 4.2.2 (Distribution of the relative speed)

There are two nodes both travelling in a straight line and let their direction of travel both be uniformly distributed. If the nodes are travelling with speeds $V_1 = v_1$ and $V_2 = v_2$ then the cdf of their relative speed is given by

$$P(V_* \leq v_* | V_1 = v_1; V_2 = v_2) = \begin{cases} 0, & \text{if } v_* \leq |v_1 - v_2|, \\ \frac{1}{\pi} \arccos\left(\frac{v_1^2 + v_2^2 - v_*^2}{2v_1 v_2}\right), & \text{if } |v_1 - v_2| < v_* < v_1 + v_2, \\ 1, & \text{if } v_* \geq v_1 + v_2. \end{cases}$$

The conditional pdf follows from this by taking the derivative:

$$f_{V_*}(v_* | V_1 = v_1; V_2 = v_2) = \begin{cases} \frac{v_*}{\pi v_1 v_2 \sqrt{1 - \left(\frac{v_1^2 + v_2^2 - v_*^2}{2v_1 v_2}\right)^2}}, & \text{if } |v_1 - v_2| < v_* < v_1 + v_2, \\ 0, & \text{otherwise.} \end{cases} \quad (4.3)$$

If the at each change of direction the nodes do not pause and their speed is selected uniformly from $(v_{min}, v_{max}]$ then the pdf of the average relative speed under the random direction mobility model is

$$f_{V_*} = \frac{1}{(v_{max} - v_{min})^2} \int_{v_{min}}^{v_{max}} \int_{v_{min}}^{v_{max}} f_{V_*}(v_* | V_1 = v_1; V_2 = v_2) dv_1 dv_2,$$

whereas for the random waypoint mobility model, the pdf of the average relative speed is given by

$$f_{V_*}(v_*) = \frac{1}{\ln^2\left(\frac{v_{max}}{v_{min}}\right)} \int_{v_{min}}^{v_{max}} \int_{v_{min}}^{v_{max}} \frac{f_{V_*}(v_* | V_1 = v_1; V_2 = v_2)}{v_1 v_2} dv_1 dv_2. \quad (4.4)$$

Both the pdf and the expected value,

$$\mathbb{E}[V_*] = \int_0^{2v_{max}} f_{V_*}(v_*) dv_* \quad (4.5)$$

can be calculated numerically. \diamond

Notice in particular that Proposition 4.2.2—with pause times equal to zero—applies to the random direction mobility model in a square and to the random waypoint mobility model in a circle¹. It can also be used as a very good approximation for the random walker in a square since there the direction of the nodes is not far from the uniform [48].

As an example, Figure 4.2(a) shows the pdf of the relative speed for both mobility models with $[v_{min}, v_{max}] = [4, 10]$ km/hour. In this case, the expected relative speed under the

¹Due to symmetry arguments the direction of travel is uniformly distributed.

random direction mobility model is $\mathbb{E}[V_*] \approx 9.2$ km/hour and under the random waypoint mobility model it is $\mathbb{E}[V_*] \approx 8.7$ km/hour.

From the figure we see that there is merely a slight difference between the two mobility models. However, as v_{min} gets closer to zero the difference between the two models becomes more distinct. To see this, Figure 4.2(b) displays the pdf of the relative speed when $[v_{min}, v_{max}] = [0.3, 10]$ km/hour. The average relative speed at the moment of contact is $\mathbb{E}[\tilde{V}_*] \approx 7.4$ km/hour for the random direction mobility model and $\mathbb{E}[\tilde{V}_*] \approx 4.4$ km/hour for the random waypoint mobility model. Here we see that under the latter movement pattern the nodes become “trapped” in the slower speeds.

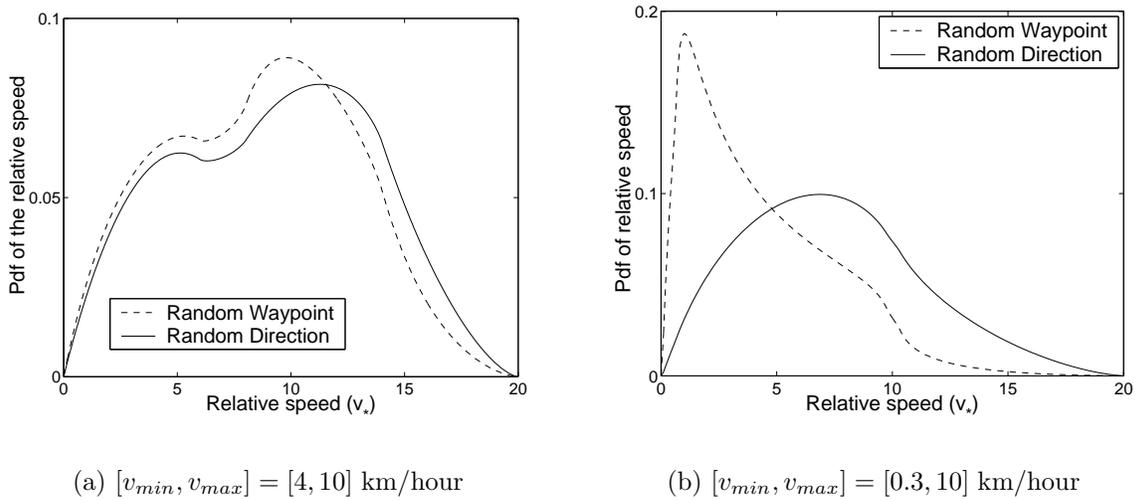


Figure 4.2: Pdf of the relative speed in two dimensions.

At the other extreme, when the minimum speed is equal to the maximum speed ($v_{min} = v_{max}$) a simple expression unveils itself and the two mobility models converge to the same distribution as shown in the following corollary.

Corollary 4.2.1 (Relative speed for nodes travelling at a constant speed)

For two nodes travelling at the same speed the cdf of the relative speed—for both the random direction and the random waypoint model—is given by

$$P(V_* \leq v_* | V_1 = v = V_2) = \frac{2}{\pi} \arcsin\left(\frac{v_*}{2v}\right). \quad (4.6)$$

The distribution function is

$$f_{V_*}(v_* | V_1 = v = V_2) = \frac{1}{\pi v \sqrt{1 - (v_*/2v)^2}} = \frac{2}{\pi \sqrt{4v^2 - v_*^2}}, \quad (4.7)$$

and the average relative speed is

$$\mathbb{E}[V_* | V_1 = v = V_2] = \frac{4v}{\pi}. \quad (4.8)$$

◇

Proof : For $V_1 = v = V_2$ the first part of proposition 4.2.2 becomes

$$P(V_* \leq v_* | V_1 = v = V_2) = \frac{1}{\pi} \arccos \left(1 - 2 \left(\frac{v_*}{2v} \right)^2 \right).$$

With the help of $\arccos(-x) + \arccos(x) = \pi$ and [37, equation 1.626.2] $\arccos(2x^2 - 1) = 2 \arccos(x)$ this gives

$$P(V_* \leq v_* | V_1 = v = V_2) = 1 - \frac{2}{\pi} \arccos \left(\frac{v_*}{2v} \right) = \frac{2}{\pi} \arcsin \left(\frac{v_*}{2v} \right),$$

where in the last step the property $\arccos(\alpha) + \arcsin(\alpha) = \pi/2$ was used. The distribution function and the expectation are easy to derive from this. ■

Interestingly, the average relative speed which is given in (4.8) is larger than v , the (average) speed of each node.² In one dimension the situation is different: there it can be shown that the average relative speed is equal to v .

We are now in a position to summarize all of the previous results into the following corollary.

Corollary 4.2.2 (First-meeting times: random waypoint and random direction)

The first-meeting time for the random direction mobility model for $r \ll L$ is approximately exponentially (λ_{RD}) distributed, where λ_{RD} is

$$\lambda_{RD} \approx \frac{2r \mathbb{E}[V_*]}{L^2}. \quad (4.9a)$$

The expected time until two nodes meet is given by

$$\mathbb{E}[F_{RD}] \approx \frac{L^2}{2r \mathbb{E}[V_*]}. \quad (4.9b)$$

Likewise, for the random waypoint mobility model we have

$$\lambda_{RW} \approx \frac{2\omega r \mathbb{E}[V_*]}{L^2}, \quad (4.10a)$$

$$\mathbb{E}[F_{RW}] \approx \frac{L^2}{2\omega r \mathbb{E}[V_*]}, \quad (4.10b)$$

²In addition—although not proven here—the average relative at the moment when two nodes meet is given by $\frac{\pi}{2}v$, which is larger than the average relative speed ($\frac{4}{\pi}v$) at a random moment in time.

where $\omega \approx 1.3683$ is the Waypoint constant.

The average relative speed can be calculated with the help of equations (4.3) and (4.5). If the speeds of both nodes are identical and constant, i.e., $V_1 = v = V_2$, then

$$\lambda_{RD} \approx \frac{8rv}{\pi L^2}, \quad \mathbb{E}[F_{RD}] = \frac{\pi L^2}{8rv}, \quad (4.11a)$$

$$\lambda_{RW} \approx \frac{8\omega rv}{\pi L^2}, \quad \mathbb{E}[F_{RW}] = \frac{\pi L^2}{8\omega rv}. \quad (4.11b)$$

◇

Proof: The first set of equations follow by filling in Proposition 4.2.1 into Theorem 4.2.1. Equation (4.11) is a direct consequence of Corollary 4.2.1. ■

Remark: A simple example why the distributions of the first-meeting times can never be truly exponential is because at a random moment in time there is a non-zero probability (but very small since $r \ll L$) that two nodes are in contact with one another (i.e., $P(F = 0) > 0$). For the interested reader this probability is given in Appendix 4.D.

An application of this Corollary has already been given in Figure 3.4 on page 80 for $[v_{min}, v_{max}] = [4, 10]$ km/hour. There the values of λ follow from $\mathbb{E}[V_*] \approx 9.3$ for the random direction and $\mathbb{E}[V_*] \approx 8.7$ for the random waypoint mobility model.

A consequence of the above corollary is the following.

Corollary 4.2.3 (Comparison between random waypoint and random direction)

If the expected relative speed is the same for the random direction and the random waypoint mobility model,³ then the parameter λ is a factor $\omega \approx 1.4$ higher for the random waypoint model. This implies that the expected first-meeting time, the expected inter-meeting time, and the expected message delay are a factor ω lower for the random waypoint model. This result holds for both the two-hop and the unlimited relay protocol.

Similarly, if the average relative speed is a factor ω higher for the random direction mobility model, then the expected first-meeting times and the expected message delay for both models are the same.

Proof: It will be shown in Section 4.4 that the first-meeting times and the inter-meeting times are almost identical to one another if $r \ll L$. The remainder is then a direct consequence of combining Corollary 4.2.2 with Theorem 3.2.2. ■

³This is the case if both nodes travel at the same constant speed ($V_1 = v = V_2$). See Corollary (4.2.2).

4.3 Examples

As examples of Corollary 4.2.2, Figures 4.3 and 4.4 show the complementary cdf for, respectively, constant speeds ($V_1 = V_2 = 10$ km/hour) and speeds chosen uniformly in $[v_{min}, v_{max}] = [4, 10]$ km/hour.

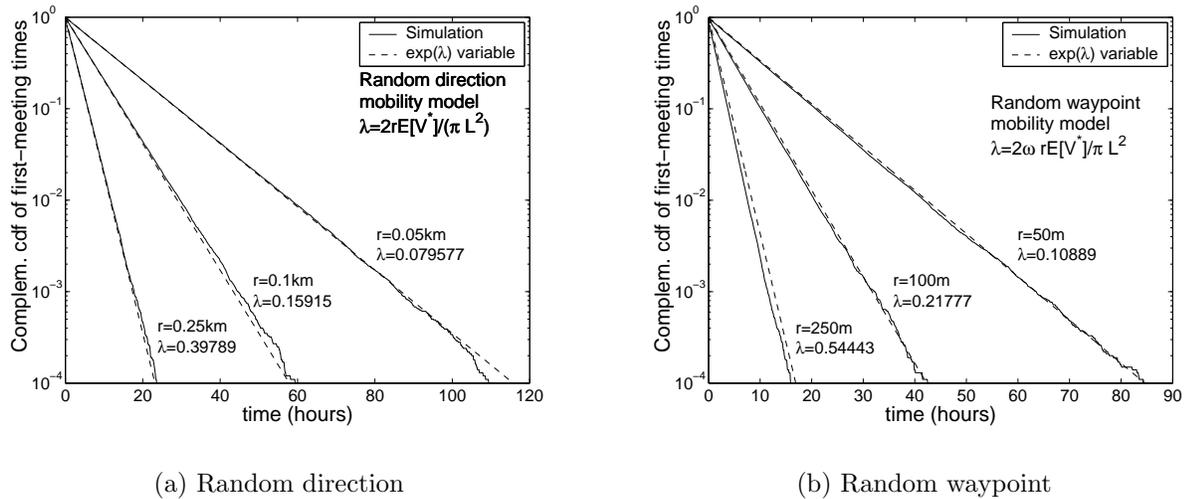


Figure 4.3: Complementary cdf of the first-meeting times for $V_1 = V_2 = v = 10$ km/hour.

Alongside each set of simulated first-meeting times is shown an exponential (λ) distribution, where λ is obtained through Corollary 4.2.2. In all cases there is an excellent match.

4.4 Implications for the message delay

In the previous chapter the Laplace-Stieltjes transform and the first two moments of the message delay were given for the case where the inter-meeting times between nodes for a Poisson process. In this section we compare the first-meeting times to the inter-meeting times and from that conclude that the results from this chapter can be applied to the previous chapter.

The inter-meeting time between two nodes is defined as the time that passes between the moment when two nodes come in contact and the moment in time when they come into contact again. This is equivalent to saying that the inter-meeting time is the contact time between two nodes plus the time until the two nodes meet again⁴. However, for movements which are very dynamic we may expect that the positions of each node, after they were in contact, quickly converge to the steady-state node location. In other words, we expect

⁴Sometimes the latter time is referred to as the inter-meeting time.

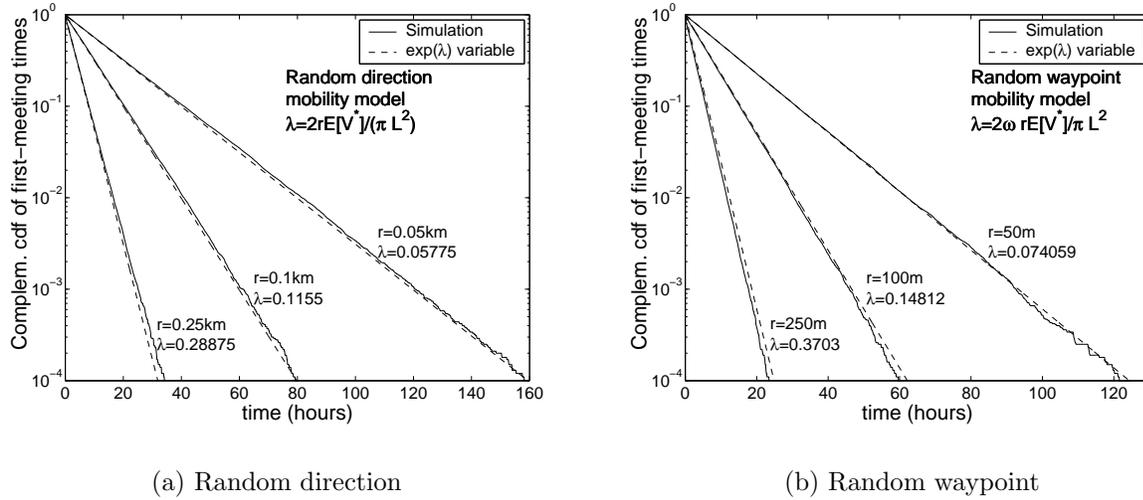


Figure 4.4: Complementary cdf of the first-meeting times if the speeds of both nodes are chosen uniformly in $[v_{min}, v_{max}] = [4, 10]$ km/hour.

the inter-meeting times and the first-meeting times will resemble one another. Similarly, if we have a Poisson process then the time between events and the time between a random moment and the next event are both exponentially distributed with the same parameter.

The quickest method of checking this is by showing in Figure 4.5 the complementary cdf of the inter-meeting times for the two movement patterns studied throughout this chapter. Once again $[v_{min}, v_{max}] = [4, 10]$ km/hour and an $exp(1/4)$ travel time is taken for the random direction mobility model. Alongside the simulated values is shown an $exp(\lambda)$ distribution with λ obtained through Corollary 4.2.2.

As can be seen from Figure 4.5, the complementary cdf of the inter-meeting time for both of the mobility models is exponentially distributed. In addition, as was expected, the estimates provided for λ for the first-meeting times are in excellent agreement with the inter-meeting times. Hence we can apply all of the results obtained in this chapter to the previous chapter.

One of the results which can be used with the previous chapter is that if the relative speeds are the same for both models—a situation that arises if the nodes all travel at the same constant speed—then the time until two nodes meet is a factor $\omega \approx 1.4$ smaller under the random walker than under the random direction mobility model. This in turn implies that the message delay is a factor ω smaller for the random waypoint model, *irrespective* of the relay protocol used! This improvement in performance is due to the nodes travelling regularly over a central region. Hence we have a direct quantification of the improvement for relaying with hotspots due to the different mobility patterns.

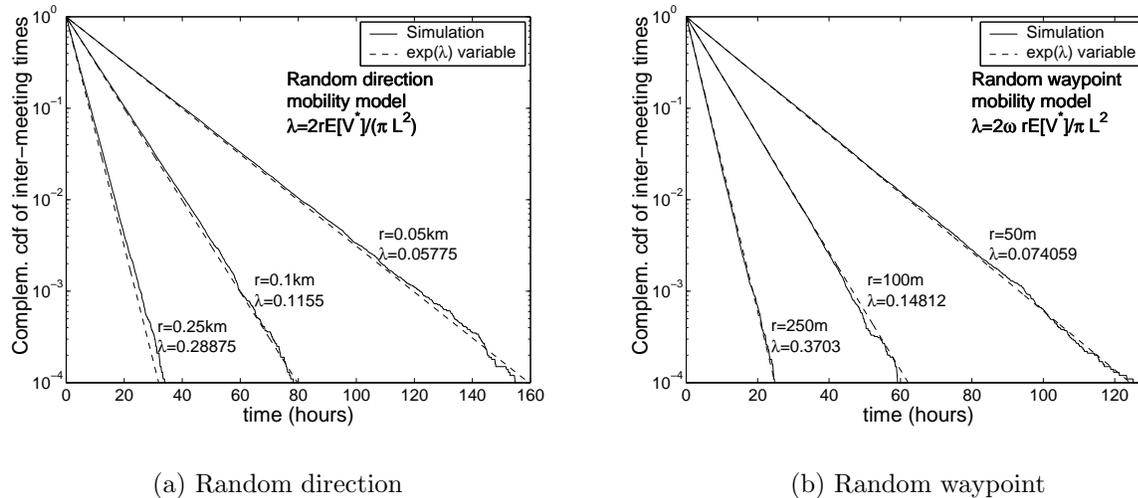


Figure 4.5: Complementary cdf of the inter-meeting time.

4.5 Concluding remarks

In this chapter it was shown that under certain conditions the first-meeting times of the two movement patterns presented—the random direction and the random waypoint mobility models—can be fitted by an exponential distribution. With this property it was shown in Chapter 3 that the message delay can be described by a simple formula. On the basis of this a number of questions come to mind which will be addressed in future research. These questions are:

- How does the first-meeting time distribution change when the transmission range (r) is not much smaller than the length (L) of the square area the nodes move on?
- Which other (types of) movement patterns have exponential first-meeting times? In the previous chapter it was shown numerically that the random walker also has this characteristic. However, for that movement pattern the nodes do not “move around” fast enough for Theorem 4.2 to be applied and hence therefore it’s value for λ remains an open question.
- How do varying transmission ranges affect the results? For example, there could be groups of nodes each with a different (fixed) transmission range, or each node could have a stochastic transmission range which varies over time.

Aside from these issues many optimization questions can be addressed by combining Theorem 3.2.1 and Theorem 4.2. For example, which relay strategy and/or transmission range must be employed to ensure a successful transmission in 95% of the cases within a certain amount of time? How can the battery power of a node be used optimally?

Another issues which requires further study is the comparison of the message delay and network consumption by having different mobility models.

In this first part of the thesis the relatively simple fixed radius model was used. Although not part of this thesis, the interference should be included into the models and the numerical studies. It is expected that the same results and conclusions hold as long as $r \ll L$ and the number of nodes per unit area is not too large.

In other words, there are still a lot of interesting problems which have to be solved....

4.A Proof of Theorem 4.2.1.

Proof: The proof consists of modelling the meeting of two nodes as a geometric variable with some probability p of success and then taking the limit to derive the exponential distribution. The probability p depends on the speeds and the positions of the two nodes. We shall derive this probability by conditioning on the position (x_1, y_1) of the first node and by deriving the probability p_{x_1, y_1} that within the next Δ_t time the two nodes will meet each other. Without much impact we can assume that there is no change of direction during Δ_t .

Note that the probability p_{x_1, y_1} is dependent on the area (A) that node one covers in Δ_t time, i.e., the highlighted area in Figure 4.6. Also note that the area of A is given by $2rV_1\Delta_t$ (ignoring border effects).

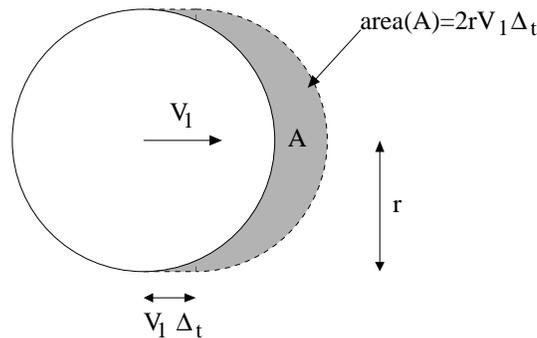


Figure 4.6: Area covered by the first node during a time Δ_t .

First we will take the position of the second node each Δ_t units of time from the stationary distribution. Afterwards the speed of the second node will be taken into account.

Assuming the second node is in its stationary position it means that the probability that the second node is located in area A is

$$p_{x_1, y_1} = \int \int_A \pi(x, y) dx dy.$$

For small r the points $\pi(x, y)$ in A can be approximated by $\pi(x_1, y_1)$ to give

$$p_{x_1, y_1} \approx 2r \cdot V_1 \cdot \Delta_t \cdot \pi(x_1, y_1).$$

Unconditioning on the position of the first node leads to the probability p that within the next Δ_t time the two nodes meet each other:

$$\begin{aligned} p &= \int_0^L \int_0^L p_{x_1, y_1} \pi(x_1, y_1) dx_1 dy_1 \\ &\approx 2r \cdot V_1 \cdot \Delta_t \int_0^L \int_0^L \pi^2(x_1, y_1) dx_1 dy_1. \end{aligned}$$

Taking the limit $\Delta_t \rightarrow 0$ gives an exponential distribution with parameter

$$\lambda \approx 2rV_1 \int_0^L \int_0^L \pi^2(x_1, y_1) dx_1 dy_1.$$

Now consider the situation where the second node is mobile with speed $V_2 = v_2$. In this case we should not look at the area the first node covers, but instead we should look at the positions the second node has to be in such it will end up in the area the first node will cover, i.e., the area the second node has to be in such that they meet in the next Δ_t units of time. As shown in Figure 4.7, the area A depends on V_* , the relative speed between the nodes⁵.

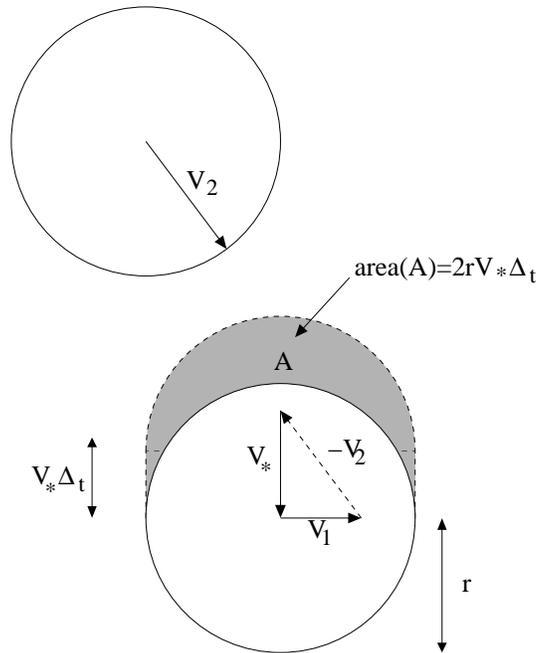


Figure 4.7: Position the second node has to be in for the nodes to meet.

In particular, if the first node is in (x_1, y_1) then the probability of a “success” is given

⁵Hint: think of this process from the viewpoint of the first node. In the case of Figure 4.7 the first node “sees” the second node coming straight down at it (since the horizontal speeds are identical). The only positions the second node can start in such that the two nodes meet are the ones directly above the first node.

by

$$\begin{aligned} p_{x_1, y_1} &= \int \int_A \pi(x_2, y_2) dx_2 dy_2 \\ &\approx 2r \cdot V_* \cdot \Delta_t \cdot \pi(x_1, y_1), \end{aligned}$$

since the area of A is $2r \cdot V_* \cdot \Delta_t$. Integrating over all starting positions of the first node give

$$\begin{aligned} p &= \int_0^L \int_0^L p_{x_1, y_1} \pi(x_1, y_1) dx_1 dy_1 \\ &\approx 2r \cdot V_* \cdot \Delta_t \cdot \int_0^L \int_0^L \pi^2(x_1, y_1) dx_1 dy_1. \end{aligned}$$

Taking the limit $\Delta_t \rightarrow 0$ finally leads to the exponential distribution with parameter

$$\lambda \approx 2rV_* \int_0^L \int_0^L \pi^2(x_1, y_1) dx_1 dy_1. \quad \blacksquare$$

4.B Proof of Proposition 4.2.1.

Proof: The expression for nodes moving according to the random direction mobility model is easy to obtain as it is known [64] that the spatial node distribution is uniform, i.e., $\pi(x, y) = 1/L^2$.

Unlike the one-dimensional case, there is unfortunately not a simple expression for the steady-state location of nodes under the random waypoint mobility model in two dimensions. Close approximations exist [13] and at best the spatial node distribution over any convex region (which includes of course a rectangle, square, circle, box, or sphere) is expressed as an integral which can be evaluated numerically [51] [52].

Summarizing the results in [51] which allow us to calculate $\pi(x, y)$ numerically: the steady-state location of nodes on a square of size $L \times L$ is given by

$$\pi(x, y) = \frac{1}{\bar{l}L^5} \int_0^\pi g(x, y, \phi) g(x, y, \phi + \pi) (g(x, y, \phi) + g(x, y, \phi + \pi)) d\phi,$$

where $\bar{l} \approx 0.521405$ is the average path length in a unit square and

$$g(x, y, \phi) = \begin{cases} \frac{L-y}{\sin \phi}, & \text{if } \arctan \frac{L-y}{L-x} \leq \phi < \frac{\pi}{2} + \arctan \frac{x}{L-y}, \\ \frac{-x}{\cos \phi}, & \text{if } \frac{\pi}{2} + \arctan \frac{x}{L-y} \leq \phi < \pi + \arctan \frac{y}{x}, \\ \frac{-y}{\sin \phi}, & \text{if } \pi + \arctan \frac{y}{x} \leq \phi < \frac{3\pi}{2} + \arctan \frac{L-x}{y}, \\ \frac{a-x}{\sin \phi}, & \text{if } 0 \leq \phi \arctan \frac{a-y}{L-x} \text{ or } \frac{3\pi}{2} + \arctan \frac{L-x}{y} \leq \phi < 2\pi. \end{cases}$$

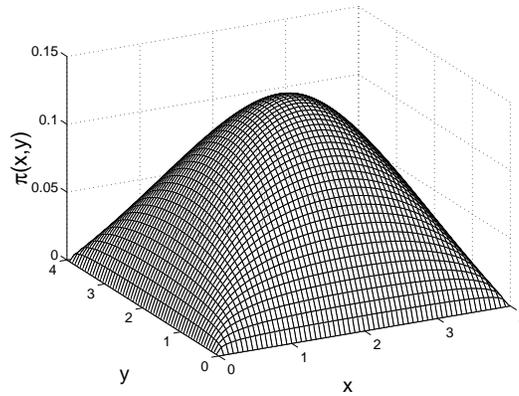


Figure 4.8: Spatial node distribution under the random waypoint mobility model.

As an example the stationary distribution of the node location in a 4×4 square is given in Figure 4.8. The integral over the square of $\pi(x,y)$ was derived numerically from these values.

4.C Proof of Proposition 4.2.2.

Proof: Let $V_1 = v_1$ and $V_2 = v_2$ be the speeds of the two nodes and let V_* be the resulting relative speed between them due to the difference $\Gamma \in (-\pi, \pi]$ in their direction, see Figure 4.9.

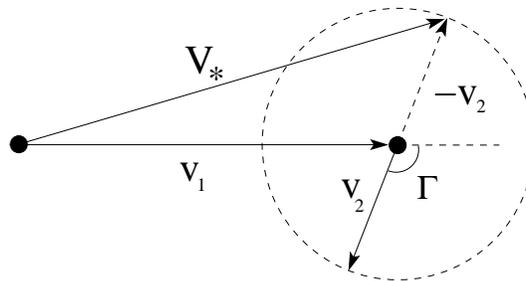


Figure 4.9: Combining the speeds $V_1 = v_1$, $V_2 = v_2$, and Γ (the difference in directions) leads to the relative speed V_* .

Notice that the relative speed is given by

$$\begin{aligned} V_*^2 &= (v_1 + v_2 \cos(\pi - \Gamma))^2 + (v_2 \sin(\pi - \Gamma))^2 \\ &= v_1^2 + v_2^2 - 2v_1v_2 \cos(\Gamma). \end{aligned}$$

This relationship allows us to derive the cdf since

$$\begin{aligned}
 P(V_* \leq v_* | V_1 = v_1; V_2 = v_2) &= P(v_1^2 + v_2^2 - 2v_1v_2 \cos(\Gamma) \leq v_*^2 | V_1 = v_1; V_2 = v_2) \\
 &= P(\cos(\Gamma) \geq \frac{v_1^2 + v_2^2 - v_*^2}{2v_1v_2} | V_1 = v_1; V_2 = v_2) \\
 &= P(|\Gamma| \leq \arccos\left(\frac{v_1^2 + v_2^2 - v_*^2}{2v_1v_2}\right) | V_1 = v_1; V_2 = v_2).
 \end{aligned}$$

The change from \geq to \leq in the last line is because $\arccos(x)$ is a decreasing function in x . Note that if the directions of travel of each node is uniformly distributed then Γ is also uniformly distributed. This leads to equation (4.3).

Next we need the pdf of the speed of a node at a random moment in time in order to uncondition (4.3). If the speed of a node is selected uniformly from $[v_{min}, v_{max}]$, then the speed of a node under the random direction mobility model is uniformly distributed and so $f_{V_i}(v_i) = 1/(v_{max} - v_{min})$.

For the pdf of the random waypoint mobility model a little more work is required. If the speed, S , of a node at a change of direction is selected uniformly from $[v_{min}, v_{max}]$, then $f_S(s) = 1/(v_{max} - v_{min})$ and for $i = 1, 2$ we have $f_{V_i}(v_i) = f_S(s)C/v_i$. Here C is a constant to ensure the probability sums to one and the division by v_i is necessary because at a random time the node only maintains a certain speed v_i for a fraction $1/v_i$ of time. Since $1 = \int_{v_{min}}^{v_{max}} f_{V_i}(v_i) = \frac{\ln(v_{max}/v_{min})}{v_{max} - v_{min}} C$ we obtain $C = \frac{v_{max} - v_{min}}{\ln(v_{max}/v_{min})}$ and

$$f_{V_i}(v_i) = \frac{1}{v_i \ln\left(\frac{v_{max}}{v_{min}}\right)}, \quad i = 1, 2.$$

(This gives the well-known result $\mathbb{E}[V_i] = (v_{max} - v_{min})/\ln\left(\frac{v_{max}}{v_{min}}\right)$)

Unconditioning (4.3) then leads to (4.4). To the best of the author's knowledge, the evaluation of the resulting integrals do not lead to a simple closed-form expression. ■

4.D Probability of contact at a random moment

In this section we derive the probability that from a random moment in time two nodes are in contact with one another. For this purpose define F as the first-meeting time and

$$C_{x,y,r} := \{(a, b) \in [0, L]^2 : \sqrt{(x-a)^2 + (y-b)^2} \leq r\},$$

i.e., $C_{x,y,r}$ is the set of points which are at most a distance r from the point (x, y) .

This allows us to write out the probability that (given the position of one of the nodes) two node are within each others vicinity as

$$P(F = 0 | (X_1, Y_1) = (x, y)) = \int \int_{C_{x,y,r}} \pi(x_2, y_2) dx_2 dy_2.$$

For the points $(x_2, y_2) \in C_{x,y,r}$ and a small r the approximation $\pi(x_2, y_2) \approx \pi(x, y)$ can be used. This results in

$$P(F = 0 | (X_1, Y_1) = (x, y)) \approx \int \int_{C_{x,y,r}} \pi(x, y) dx_2 dy_2 = \pi r^2 \pi(x, y)$$

and

$$\begin{aligned} P(F = 0) &= \int_0^L \int_0^L P(F = 0 | (X_1, Y_1) = (x, y)) \cdot \pi(x, y) dx dy \\ &\approx \pi r^2 \int_0^L \int_0^L \pi^2(x, y) dx dy \end{aligned}$$

For the random direction mobility model this gives

$$P(F = 0) \approx \frac{\pi r^2}{L^2}, \tag{4.12a}$$

and for the random waypoint mobility model (c.f., equation (4.2b))

$$P(F = 0) \approx \frac{\pi \omega r^2}{L^2}, \tag{4.12b}$$

where $\omega \approx 1.368$ is the Waypoint contact.

Since $\pi r^2/L^2 \approx 0.012$ for $r = 0.25\text{km}$ and $L = 4\text{km}$ it follows that $P(F > 0) \geq 0.98$. This explains why the complementary cdf in Figure 4.1 and Figures 4.3-4.5 appear to start from one on the y-axis. ■

Polling Systems with Correlated Switchover Times

An Alternating-Priority Server with Correlated Switchover Times

Contents

5.1	Introduction and Motivation	110
5.2	Model Description	111
5.3	Exhaustive/Exhaustive Service System	114
5.4	Exhaustive/Gated Service System	124
5.5	Examples	135
5.6	Concluding Remarks	143
5.A	Proof of Theorem 5.3.2.	145
5.B	Proof of Theorem 5.3.3.	147
5.C	Alternative proof of Theorem 5.3.4.	150
5.D	Proof of Theorem 5.4.3.	150
5.E	List of Notation	157

In this second part of the thesis a single server queueing system is analysed in which service is alternated between two queues and the server requires a (finite) switchover time to switch from one queue to the other. The distinction from classical results is that the sequence of switchover times from each of the queues need not be i.i.d. nor independent from each other; each sequence is merely required to form a stationary ergodic sequence. With the help of stochastic recursive equations explicit expressions are derived for a number of performance measures, most notably for the average delay of a customer and the average queue lengths under different service disciplines. With these expressions a comparison is made between the service disciplines and the influence of correlation is studied. Finally, through a number of examples it is shown that the correlation can significantly increase the

mean delay and the average queue lengths indicating that the correlation between switchover times should not be ignored. This has important implications for communication systems in which a common communication channel is shared amongst various users and where the time between consecutive data transfers is correlated (for example in ad-hoc networks).

Note: A summary of the material presented in this chapter is published in [40]. A longer version almost identical to this chapter has been submitted to QUESTA and is available as an INRIA research report [39].

5.1 Introduction and Motivation

So far only few explicit results have been known in queueing theory for systems whose evolution is described by general stationary ergodic processes. One line of research that allows one to handle stationary ergodic sequences is based on identifying measures that are insensitive to correlations. For example, the probability of finding a G/G/1 queue non-empty is just the ratio between the expected service time and the expected inter-arrival time of customers (which follows directly from Little's Law). The expected cycle duration in a polling system (under fairly general condition) too, depends on the inter-arrival, service and vacation times only through their expectations under general stationary ergodic assumptions (see e.g., [7]). An example of performance measures that depend on the whole distribution of service times but is insensitive to correlations is the growth rate of the number of customers or of the sojourn time in a (discriminatory) processor sharing queue in overload [6, 54]. Other insensitivity results on bandwidth sharing in a network can be found in [18, 17].

The polling models studied here do not exhibit insensitivity. Approximating correlated vacations by independent ones can result in large errors in the performance, see e.g., [90] in the context of Bluetooth. To study these systems use is made of stochastic recursions equations (SRE) introduced in [3] which extend branching processes with migration on one hand, and linear stochastic recursive equations¹ on the other. It has already been shown in [3] that vector valued SRE can be used to describe some embedded processes appearing in polling models.²

In this chapter one dimensional SRE are identified which are used in order to compute the expected waiting times and queue lengths in a system with two queues where a single server alternates between two queues and requires switch-over times (modeled as vacations) to move from one queue to the other. Two systems are studied: one in which both queues are served exhaustively and one in which one queue is served exhaustively and the other according to the gated discipline. In the exhaustive service discipline the server serves a queue until it is empty before switching to the next queue. In the gated service discipline a gate is closed in front of a queue when the server arrives at that queue. The server then serves only the customers that were present at the moment the gate was closed, after which

¹Linear SRE have already been used to study the impact of correlation of the loss process on TCP throughput [5].

²SRE have also been used recently to study the infinite server queue with correlated arrivals [4].

it moves on to the next queue. The customers who arrived after the gate was closed have to wait until the next round before being served. The analytical results obtained are then used to study numerically the impact of correlated switchover times on the performance, as well as the difference in performance due to the service discipline used.

The polling system studied in this paper, but without the correlation, has been used in the past [75] to model communication systems in which transmission between two stations can take place only in one direction at a time³. The position of the server then corresponds to the direction data is travelling in. A similar situation arises in ad hoc networks; there is a common channel which needs to be shared amongst various users. The more users there are, the longer one has to wait before being able to capture the channel necessary to (re)transmit data. In particular, if one has to wait a long time before being able to transmit data, then it is very likely that there are many users around and that the next time one has to wait once again for a long period of time. For this reason the correlation of the number of users over time in an ad hoc network inherently introduces correlation between the waiting (switchover) times, and this in turn leads to an increase in the mean delay and queue lengths.

The remainder of this document is structured as follows. In section 5.2 the polling system is described in more detail and the notation and some formulas are established. Section 5.3 (respectively section 5.4) is entirely devoted to the analysis of the exhaustive/exhaustive (respectively exhaustive/gated) queueing system. This is done by first deriving a SRE in section 5.3.1 (5.4.1) which leads to the derivation of a number of performance measures in section 5.3.2 (5.4.2), of which most notably the expected waiting times and queue lengths in section 5.3.2.6 (5.4.2.7). The performance measures are then used in the examples of section 5.5 to show the influence of correlated switchover times. Finally, conclusions are given in section 5.6. To aid the reader, a list of notation is given in appendix 5.E.

5.2 Model Description

We examine the polling of two queues, i.e., one queue is served after which the other queue is served. No limit is specified for the length of either queue. After serving queue i ($i = 1, 2$) for the n -th time, the server requires a switchover time of duration $V_{n,i}$. Assume all $V_{n,i}$ have the same distribution as V_i ($V_{n,i} \sim V_i$), where V_i is assumed to form a general distribution with first and second moment v_i and $v_i^{(2)}$, and with variance $\delta_i^2 := v_i^{(2)} - v_i^2$, $i = 1, 2$. Let $R := v_1 + v_2$ and $\Delta^2 := \delta_1^2 + \delta_2^2$. The sequences of switchover times are assumed to be stationary ergodic instead of the usual i.i.d., and possibly dependent on each other. This implies that there can be a correlation between the switchover times of the two queues and/or within the sequence of switchover times for each queue. The arrival of customers at queue i is Poisson with rate λ_i and the service times are nonnegative, i.i.d. random variables with (finite) first and second moments for queue i given by, respectively, b_i and $b_i^{(2)}$. The *load* at queue i is $\rho_i := \lambda_i b_i$ and the system is stable [78, page 280] if and only if the overall *load* $\rho := \rho_1 + \rho_2 < 1$,

³One can also think of a narrow bridge with cars travelling over the bridge from one direction at a time.

which we assume throughout. Furthermore, we will continuously assume that the queues are operating under stationary regime.

Introduce the covariance functions ($i = 1, 2$)

$$c_i(n) = \mathbb{E}[V_{0,i}V_{n,i}] - \mathbb{E}[V_{0,i}]\mathbb{E}[V_{n,i}], \quad n \in \mathbb{N}, \quad (5.1)$$

$$c_{12}(n) = \mathbb{E}[V_{0,1}V_{n,2}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,2}], \quad n \in \mathbb{Z}. \quad (5.2)$$

Note that $c_{12}(n)$ is defined for $n \in \mathbb{Z}$. With this convention it is not necessary to work with $c_{21}(n) := \mathbb{E}[V_{0,2}V_{n,1}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{n,1}]$, since under stationary regime it follows that $c_{21}(n) = \mathbb{E}[V_{0,2}V_{n,1}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{n,1}] = \mathbb{E}[V_{-n,2}V_{0,1}] - \mathbb{E}[V_{-n,2}]\mathbb{E}[V_{0,1}] = c_{12}(-n)$. In particular, if for each queue the sequence of switchover times is uncorrelated, then $c_i(0) = \delta_i^2$ and $c_i(n) = 0$, for $n \in \mathbb{N}$. If there is no correlation between the switchover times of the two queues, then $c_{12}(n) = 0$, for $n \in \mathbb{Z}$.

Because of the assumption of the queues operating under stationary regime (5.1) and (5.2) can be rewritten as

$$\mathbb{E}[V_{0,i}V_{n,i}] = v_i^2 + c_i(n), \quad i = 1, 2, \quad n \in \mathbb{N}, \quad (5.3a)$$

$$\mathbb{E}[V_{0,1}V_{n,2}] = v_1v_2 + c_{12}(n), \quad n \in \mathbb{Z}. \quad (5.3b)$$

In order to establish the SRE, let $\mathcal{D}_{n,i}(N)$ be the duration of the busy period in the i^{th} queue, initiated by N customers waiting in that queue when the server arrives at that queue for the n^{th} time. Similarly, let $\mathcal{N}_{n,i}(T)$ be the number of customers arriving at queue i during a period of time T during the server's n^{th} visit to queue i .

Let us now establish a number of formulas which will be used throughout. First recall that if D_i is a random sequence of i.i.d. random variables, independent of a random variable N , and $\mathbb{E}[D_i] = d$, $\mathbb{E}[D_i^2] = d^{(2)}$, and

$$\tau(N) := \sum_{i=1}^N D_i, \quad (5.4)$$

then

$$\begin{aligned} \mathbb{E}[\tau^2(N)] &= \sum_{n=1}^{\infty} \mathbb{E} \left[\sum_{i=1}^n D_i \cdot \sum_{i=1}^n D_i \mid N = n \right] P(N = n) \\ &= \sum_{n=1}^{\infty} n \left(\mathbb{E}[D_1^2] + (n-1)(\mathbb{E}[D_1])^2 \right) P(N = n) \\ &= d^2 \mathbb{E}[N^2] + (d^{(2)} - d^2) \mathbb{E}[N]. \end{aligned} \quad (5.5)$$

Similarly,

$$\begin{aligned} \mathbb{E}[\mathcal{N}_{n,i}^2(T)] &= \int_0^{\infty} \mathbb{E}[\mathcal{N}_{n,i}^2(t) \mid T = t] dT(t) = \int_0^{\infty} \left((\lambda_i t)^2 + \lambda_i t \right) dT(t) \\ &= \lambda_i^2 \mathbb{E}[T^2] + \lambda_i \mathbb{E}[T]. \end{aligned} \quad (5.6)$$

Next we proceed in a similar manner to obtain the second moment of the busy period generated by Y customers initially in the system. First recall that $\mathcal{D}_{n,i}(1)$ is a single busy period initiated by a single customer in an M/G/1 queue with Poisson arrivals with rate λ_i and general service time with first and second moments b_i and $b_i^{(2)}$ respectively. The first two moments of a single busy period initiated by a single customer are given by [57, equations 5.141 and 5.142]

$$d_i := \mathbb{E}[\mathcal{D}_{n,i}(1)] = \frac{b_i}{1 - \rho_i}, \quad (5.7a)$$

$$d_i^{(2)} := \mathbb{E}[\mathcal{D}_{n,i}^2(1)] = \frac{b_i^{(2)}}{(1 - \rho_i)^3}, \quad i = 1, 2. \quad (5.7b)$$

The busy period, $D_{n,i}(Y)$, generated by Y customers is the sum of Y independent single busy periods, each with distribution $D_{n,i,k} \sim D_{n,i}(1)$. Hence the second moment is

$$\mathbb{E}[\mathcal{D}_{n,i}^2(Y)] = \mathbb{E}\left[\sum_{k=1}^Y \mathcal{D}_{n,i,k}\right]^2 = \mathbb{E}\left[\sum_{k=1}^Y \mathcal{D}_{n,i}(1)\right]^2 = d_i^2 \mathbb{E}[Y^2] + (d_i^{(2)} - d_i^2) \mathbb{E}[Y]. \quad (5.8)$$

By taking $Y = \mathcal{N}_{n,i}(T)$ we obtain ($i = 1, 2$)

$$\begin{aligned} \mathbb{E}[\mathcal{D}_{n,i}^2(\mathcal{N}_{n,i}(T))] &= d_i^2 \mathbb{E}[\mathcal{N}_{n,i}^2(T)] + (d_i^{(2)} - d_i^2) \mathbb{E}[\mathcal{N}_{n,i}(T)] \\ &= \lambda_i^2 d_i^2 \mathbb{E}[T^2] + d_i^{(2)} \lambda_i \mathbb{E}[T]. \end{aligned} \quad (5.9)$$

5.3 Exhaustive/Exhaustive Service System

5.3.1 Introduction

We start by examining the exhaustive polling of two queues, i.e., one queue is served until it is empty after which the other queue is served until emptied. Consider the system at the moment the server starts serving the first queue for the n^{th} time with $L_{n,1}^*$ customers waiting in the queue. From here on the following steps take place (see Figure 5.1 for a visual representation of this decomposition):

- *Exhausting the first queue.* The $L_{n,1}^*$ customers in the first queue require a busy period duration of $D_{n,1} := \mathcal{D}_{n,1}(L_{n,1}^*)$ to exhaust.
- *Switching to the second queue.* After serving the first queue the server requires a switchover time of $V_{n,1}$ units of time.
- *Exhausting the second queue.* In the time needed to switch from the second to the first queue ($V_{n-1,2}$), to exhaust the first queue ($D_{n,1}$), and to switch back to the second queue ($V_{n,1}$), there have been $L_{n,2}^* := \mathcal{N}_{n-1,2}(V_{n-1,2}) + \mathcal{N}_{n,2}(\mathcal{D}_{n,1}(L_{n,1}^*) + V_{n,1})$ customers arriving at the second queue. It requires $D_{n,2} := \mathcal{D}_{n,2}(L_{n,2}^*)$ units of time to empty this queue.
- *Switching back to the first queue.* After serving the second queue the server requires a switchover time of $V_{n,2}$ units of time.

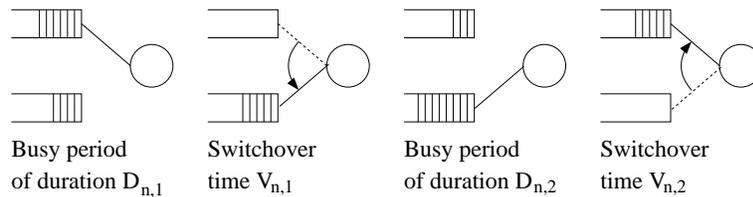


Figure 5.1: Decomposition of the n^{th} cycle into busy periods ($D_{n,1}$ and $D_{n,2}$) and switchover times ($V_{n,1}$ and $V_{n,2}$).

After this the process starts over again and a new *cycle* begins. Hence the n^{th} cycle is made up of

$$C_n = D_{n,1} + V_{n,1} + D_{n,2} + V_{n,2}.$$

The time between the server finishing work at queue i and returning to queue i in the next cycle is the *inter-visit* time $I_{n,i}$ and is given by

$$I_{n,1} = V_{n,1} + D_{n,2} + V_{n,2}, \quad (5.10a)$$

$$I_{n,2} = V_{n,2} + D_{n+1,1} + V_{n+1,1}. \quad (5.10b)$$

A SRE will be established for this quantity and we will see that it plays a central role for the derivation of the expected waiting times and queue lengths. The time $D_{n+1,i}$ spent at queue i in the $(n+1)^{th}$ cycle is related to the inter-visit time according to

$$D_{n+1,1} = \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})), \quad (5.11a)$$

$$D_{n+1,2} = \mathcal{D}_{n+1,2}[\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})))] \quad (5.11b)$$

The expectation is the sum of expected busy periods [57, p.217] and thus

$$\mathbb{E}[D_{n,1}] = \frac{b_1 \mathbb{E}[\mathcal{N}_{n,1}(I_{n,1})]}{1 - \rho_1} = \frac{\rho_1 \mathbb{E}[I_{n,1}]}{1 - \rho_1}. \quad (5.12a)$$

Using the stationarity and the divisibility ($\mathcal{N}_{n,2}(a+b) = \mathcal{N}_{n,2}(a) + \mathcal{N}_{n,2}(b)$) of the arrival process it can be shown that

$$\mathbb{E}[D_{n,2}] = \frac{\rho_2 \mathbb{E}[I_{n,2}]}{1 - \rho_2}. \quad (5.12b)$$

Since the busy periods are sums of service times, the divisibility property also holds for $\mathcal{D}_{n,i}$. This means that from (substitute 5.11b into 5.10a)

$$I_{n+1,1} = V_{n+1,1} + V_{n+1,2} + \mathcal{D}_{n+1,2}[\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})))],$$

$n \in \mathbb{N}$, we see a SRE (as presented and solved for stationary ergodic sequences in [3]) arising. Although the system is two dimensional (as there are two queues), the reduction to a one dimensional SRE is a key element in obtaining explicit formulas for the performance measures.

Theorem 5.3.1 (SRE for exhaustive/exhaustive system). *The inter-visit time of the first queue allows itself to be written as a one-dimensional SRE,*

$$I_{n+1,1} = \mathcal{A}_n(I_{n,1}) + \mathcal{B}_n \quad (5.13)$$

with

$$\mathcal{A}_n(\cdot) := \mathcal{D}_{n+1,2}(\mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(\cdot)))) \quad (5.14)$$

$$\mathcal{B}_n := V_{n+1,1} + V_{n+1,2} + \mathcal{D}_{n+1,2}(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1})). \quad (5.15)$$

Note that from (5.12a) and (5.12b) we have $\mathbb{E}[\mathcal{A}_n(I_{n,1})] = \alpha \mathbb{E}[I_{n,1}]$ where $\alpha := \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$.

◇

The existence of a stationary ergodic $I_{n,1}$ which satisfies (5.13) is given in the following theorem which is proven in Appendix 5.A.

Theorem 5.3.2 *If $\rho < 1$ then there exists a stationary ergodic regime $I_{n,1}^*$ which satisfies (5.13).* \diamond

5.3.2 Performance Measures

In the next couple of sections the first two moments of various performance measures will be derived. Starting with the first two moments of the intervisit time of the first queue, we see that all of the other quantities can easily be expressed in terms of the intervisit time. Most importantly, closed form expressions for the average waiting time and the mean queue length will be derived in the presence of correlated switchover times.

5.3.2.1 Intervisit Time

Taking the expectation of the square of (5.13) gives

$$\mathbb{E}[I_{n+1,1}^2] = \mathbb{E}[\mathcal{A}_n(I_{n,1})]^2 + \mathbb{E}[\mathcal{B}_n^2] + 2\mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n]. \quad (5.16)$$

This expression is central to the proof of the following theorem.

Theorem 5.3.3 (Intervisit time in exhaustive/exhaustive system). *Under the stationary regime the expected intervisit time of queue i is given by*

$$E[I_{n,i}] = \frac{R(1 - \rho_i)}{1 - \rho}, \quad \rho := \rho_1 + \rho_2, \quad i = 1, 2. \quad (5.17)$$

The second moment is given by

$$\beta \mathbb{E}[I_{n,1}^2] = \frac{R}{1 - \rho} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1 - \rho_1)^2} + \lambda_2 b_2^{(2)} \right) + \delta_1^2 + \left(1 - \frac{2\rho_2(1 - \rho)}{1 - \rho_1} \right) \delta_2^2 + \left(\frac{1 - \rho + 2\rho_1\rho_2}{1 - \rho} \right) R^2 + 2C \quad (5.18)$$

where

$$\alpha := \frac{\rho_1\rho_2}{(1 - \rho_1)(1 - \rho_2)}, \quad \beta := \frac{(1 - \rho)(1 - \rho + 2\rho_1\rho_2)}{(1 - \rho_1)^2}.$$

and

$$C := \sum_{j=1}^{\infty} \left[c_1(j) + c_2(j) + \frac{(1-\rho)^2}{\rho_1(1-\rho_1)} c_2(j) + \frac{1-\rho_2}{\rho_1} c_{12}(-j) + \frac{1-\rho_2(1-\alpha)}{\alpha} c_{12}(j-1) \right] \alpha^j \quad (5.19)$$

is the addition to the intervisit time due to the correlation between the switchover times. \diamond

Remark: By taking $\lambda_2 = 0$, one obtains an ordinary M/G/1 queue with correlated vacations where $\mathbb{E}[I_{n,1}^2]$ ($= v_1^{(2)} + v_2^{(2)} + 2v_1v_2$) no longer depends on the correlation (since $\alpha = 0$).

The proof is given in Appendix 5.B. On the basis of this theorem a number of other results quickly follow.

Corollary 5.3.1 (Intervisit Time With Uncorrelated Switchover Times).

If the successive switchover times in each queue are uncorrelated then $c_1(j) = c_2(j) = 0$ for $j \geq 1$. Furthermore, if the switchover times between the two queues are uncorrelated, then $c_{12}(j) = 0$ for $j \in \mathbb{Z}$. This leads to $C = 0$. The expression then agrees with the classical result mentioned in [28, equation (55)] or in [77, equation⁴ (4.36a)]. \diamond

5.3.2.2 Number of Customers Waiting When Server Arrives

The exhaustive nature of the server implies that the number of customers building up at queue i is exactly the number of customers that arrived at that queue during its intervisit time. Thus,

$$L_{n+1,1}^* = \mathcal{N}_{n,1}(I_{n,1}).$$

From this we immediately obtain

$$\mathbb{E}[L_{n+1,i}^*] = \lambda_i \mathbb{E}[I_{n,i}] = \frac{R\lambda_i(1-\rho_i)}{1-\rho} \quad (5.20)$$

as the expected length of the queue, under stationary regime, at the moment the server arrives at queue i ($i = 1, 2$). The second moment follows through squaring,

$$\mathbb{E}[(L_{n+1,1}^*)^2] = \mathbb{E}[\mathcal{N}_{n,1}^2(I_{n,1})] = \lambda_1^2 \mathbb{E}[I_{n,1}^2] + \lambda_1 \mathbb{E}[I_{n,1}],$$

which leads to

$$\mathbb{E}[(L_{n,i}^*)^2] := \lambda_i^2 \mathbb{E}[I_{n,i}^2] + \frac{R\lambda_i(1-\rho_i)}{1-\rho}. \quad (5.21)$$

⁴Instead of switchover times this work refers to reply intervals. Unfortunately, the definition or the indices in (4.36a) are not coherent for the case $N = 2$ as the indices of ones and two have to be interchanged (for δ_1 and δ_2 , as well as for r_1 and r_2) to obtain the correct result. The result is mentioned correctly in [28, equation (55)].

5.3.2.3 Duration of Busy Periods

The expected time per cycle, in steady state, for the server to work on queue i ($i = 1, 2$) is given by

$$\mathbb{E}[D_{n,i}] = \frac{R\rho_i}{1-\rho}. \quad (5.22)$$

This follows directly from $\mathbb{E}[D_{n,i}] = \mathbb{E}[\mathcal{D}_{n,i}(L_{n,i}^*)] = \frac{b_i \mathbb{E}[L_{n,i}^*]}{1-\rho_i}$.

Since $\mathbb{E}[D_{n,1}^2] = \mathbb{E}[\mathcal{D}_{n,1}^2(\mathcal{N}_{n,1}(I_{n,1}))]$, the second moment follows with (5.9) and is given by

$$\mathbb{E}[D_{n,1}^2] = \frac{\rho_1^2 \mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2} + \frac{R\lambda_1 b_1^{(2)}}{(1-\rho_1)^2(1-\rho)}. \quad (5.23)$$

5.3.2.4 Cycle time

By definition $C_{n+1,i} = D_{n+1,i} + I_{n+1,i}$. From this the expected duration of a cycle under stationary regime immediately follows:

$$\mathbb{E}[C_{n+1,i}] = \frac{R}{1-\rho}, \quad i = 1, 2. \quad (5.24)$$

Taking the expectation over the square produces

$$\mathbb{E}[C_{n+1,1}^2] = \mathbb{E}[D_{n+1,1}^2] + \mathbb{E}[I_{n+1,1}^2] + 2\mathbb{E}[D_{n+1,1}I_{n+1,1}].$$

The last term on the right hand side can be derived using (5.13) to give

$$\begin{aligned} \mathbb{E}[D_{n+1,1}I_{n+1,1}] &= \mathbb{E}[D_{n+1,1} \cdot (\mathcal{A}_n(I_{n,1}) + \mathcal{B}_n)] = \mathbb{E}[D_{n+1,1} \cdot \mathcal{A}_n(I_{n,1})] + \mathbb{E}[D_{n+1,1} \cdot \mathcal{B}_n] \\ &= \mathbb{E}[D_{n+1,1} \cdot \mathcal{D}_{n+1,2}(\mathcal{N}_{n+1,2}(D_{n+1,1}))] + \mathbb{E}[\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})) \cdot \mathcal{B}_n] \\ &= \frac{\rho_2}{1-\rho_2} \mathbb{E}[D_{n+1,1}^2] + \frac{\rho_1}{1-\rho_1} \mathbb{E}[I_{n,1}\mathcal{B}_n]. \end{aligned}$$

Plugging in the expressions for $\mathbb{E}[I_{n,1}\mathcal{B}_n]$ (equations (5.72) and (5.75)) reveals that

$$\begin{aligned} \mathbb{E}[C_{n+1,1}^2] &= \mathbb{E}[D_{n+1,1}^2] + \mathbb{E}[I_{n+1,1}^2] + \frac{2\rho_2}{1-\rho_2} \mathbb{E}[D_{n+1,1}^2] + \frac{2\rho_1}{1-\rho_1} \mathbb{E}[I_{n,1}\mathcal{B}_n] \\ &= \left(\frac{1+\rho_2}{1-\rho_2} \right) \mathbb{E}[D_{n+1,1}^2] + \mathbb{E}[I_{n,1}^2] + \frac{2\rho_1}{1-\rho_2} \left(\frac{R^2}{1-\rho} + \frac{\rho_2\delta_2^2}{1-\rho_1} \right) \\ &\quad + \frac{2}{\rho_2} \sum_{j=1}^{\infty} \alpha^j \left[\frac{c_1(j)+c_2(j)}{1-\rho_2} + \frac{(1-\rho)^2 c_2(j)}{\rho_1(1-\rho_1)(1-\rho_2)} + \frac{c_{12}(-j)}{\rho_1} + c_{12}(j) \left(1 + \frac{\alpha\rho_2}{1-\rho_2} \right) \right] \\ &\quad - 2c_2(1) + \frac{2(-c_{12}(-1) + \alpha c_{12}(0))}{1-\rho_2}. \end{aligned}$$

The expression for $\mathbb{E}[D_{n+1,1}^2] = \mathbb{E}[D_{n,1}^2]$ is given in equation (5.23) and leads to

$$\begin{aligned} \mathbb{E}[C_{n+1,1}^2] &= \left(\left(\frac{1+\rho_2}{1-\rho_2} \right) \left(\frac{\rho_1^2}{(1-\rho_1)^2} \right) + 1 \right) \mathbb{E}[I_{n+1,1}^2] + \frac{2\rho_1}{1-\rho_2} \left(\frac{R^2}{1-\rho} + \frac{\rho_2\delta_2^2}{1-\rho_1} \right) \\ &\quad + \left(\frac{1+\rho_2}{1-\rho_2} \right) \frac{R\lambda_1 b_1^{(2)}}{(1-\rho_1)^2(1-\rho)} - 2c_2(1) + \frac{2(-c_{12}(-1) + \alpha c_{12}(0))}{1-\rho_2} \\ &\quad + \frac{2}{\rho_2} \sum_{j=1}^{\infty} \alpha^j \left[\frac{c_1(j) + c_2(j)}{1-\rho_2} + \frac{(1-\rho)^2 c_2(j)}{\rho_1(1-\rho_1)(1-\rho_2)} + \frac{c_{12}(-j)}{\rho_1} + c_{12}(j) \left(1 + \frac{\alpha\rho_2}{1-\rho_2} \right) \right]. \end{aligned}$$

Remark: A cycle can also be defined as $C_{n,1} = I_{n,1} + D_{n+1,1}$. The first moment of the cycle time is then the same. However, the second moment is then different since

$$\begin{aligned} \mathbb{E}[C_{n,1}^2] &= \mathbb{E}[I_{n,1}^2] + \mathbb{E}[D_{n+1,1}^2] + 2\mathbb{E}[I_{n,1} \cdot \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1}))] \\ &= \mathbb{E}[I_{n,1}^2] + \lambda_1^2 d_1^2 \mathbb{E}[I_{n,1}^2] + \lambda_1 d_1^{(2)} \mathbb{E}[I_{n,1}] + 2\lambda_1 d_1 \mathbb{E}[I_{n,1}^2] \\ &= \frac{\mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2} + \frac{\lambda_1 b_1^{(2)} \mathbb{E}[I_{n,1}]}{(1-\rho_1)^3}. \end{aligned}$$

This last expression corresponds to [77, equation⁵ (4.23b)]. ■

5.3.2.5 Number Served per Cycle

To derive the first and second moments of the number of customers served per cycle, consider an $M/G/1$ queue with arrival rate λ_i , average service time b_i , and the second moment of the service time $b_i^{(2)}$. Then the expectation [57, equation (5.153)] and the variance [57, equation (5.154)] of the number of customers served in a single busy period are

$$\mathbb{E}[\Gamma_i] = \frac{1}{1-\rho_i}, \quad \text{Var}[\Gamma_i] = \frac{\rho_i(1-\rho_i) + \lambda_i^2 b_i^{(2)}}{(1-\rho_i)^3}.$$

Let $\mathcal{T}_{n,i}(N)$ be the number of customers served at queue i during the n^{th} cycle if there are N customers in the queue at the moment of polling. Since the number served is the sum of the number served during N busy periods, (5.4) tells us that the expected number of customers served, per cycle, at queue i is

$$\mathbb{E}[T_{n,i}] = \mathbb{E}[\mathcal{T}_{n,i}(L_{n,i}^*)] = \frac{\mathbb{E}[L_{n,i}^*]}{1-\rho_i} = \frac{\lambda_i R}{1-\rho_i}, \quad (5.25)$$

⁵There is in fact a factor λ_i missing in [77, formula (4.23b)]. It should read $\dots + \frac{\lambda_i[-(1-\rho_i) + \lambda_i^2 b_i^{(2)}] \sum_{k=1}^N r_k}{1 - \sum_{k=1}^N \rho_k}$.

for $i = 1, 2$. To derive the second moment note that

$$\begin{aligned}\mathbb{E}[T_{n,i}^2] &= \mathbb{E}[\mathcal{T}_{n,i}^2(L_{n,i}^*)] \\ &= (\mathbb{E}[\Gamma_i])^2 \mathbb{E}[(L_{n,i}^*)^2] + \text{Var}[\Gamma_{n,i}] \mathbb{E}[L_{n,i}^*] \\ &= \frac{\mathbb{E}[(L_{n,i}^*)^2]}{(1-\rho_i)^2} + \frac{R\lambda_i(\rho_i(1-\rho_i) + \lambda_i^2 b_i^{(2)})}{(1-\rho_i)^2(1-\rho)}.\end{aligned}$$

Using equation (5.21) gives

$$\mathbb{E}[T_{n,i}^2] = \frac{1}{(1-\rho_i)^2} \left(\lambda_i^2 \mathbb{E}[I_{n,i}^2] + \frac{R\lambda_i(1-\rho_i^2 + \lambda_i^2 b_i^{(2)})}{1-\rho} \right) \quad i = 1, 2. \quad (5.26)$$

5.3.2.6 Expected Waiting Time and Average Queue Length

Two of the most important performance measures, the expected waiting time and the average queue length, still remain to be given. The following theorem provides us with this result.

Theorem 5.3.4 (Expected waiting time and queue length for the exhaustive/exhaustive service discipline). *The expected waiting time (total time in system minus service time) of a customer going through queue i is decomposed of two parts, namely*

$$\mathbb{E}[W_{q,i}] = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{(1-\rho)\mathbb{E}[I_{n,i}^2]}{2R(1-\rho_i)}, \quad (5.27)$$

$i = 1, 2$. This gives

$$\begin{aligned}\mathbb{E}[W_{q,1}] &= \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{\lambda_1 \rho_2^2 b_1^{(2)} + \lambda_2 (1-\rho_1)^2 b_2^{(2)}}{2(1-\rho_1)(1-\rho)(1-\rho+2\rho_1\rho_2)} \\ &\quad + \frac{(1-\rho_1)R}{2(1-\rho)} + \left(\frac{\Delta^2}{2} - \frac{\rho_2(1-\rho)}{1-\rho_1} \delta_2^2 + C \right) \psi, \quad (5.28)\end{aligned}$$

where $\psi := \frac{1-\rho_1}{R(1-\rho+2\rho_1\rho_2)}$ and C (defined in (5.19)) is the increase in the expected waiting time due to correlated switchover times.

The average number of customers at queue i (in service and in the queue) follows directly from Little and is

$$\mathbb{E}[L_{s,i}] = \lambda_i \mathbb{E}[W_{q,i}] + \rho_i, \quad i = 1, 2. \quad \diamond$$

In the uncorrelated case ($C = 0$) this is in correspondence with [78, formula⁶ (3.12)].

Proof: For a customer arriving at queue i ($i = 1, 2$) the system behaves as an M/G/1 queue where the server goes on vacation as soon as the queue is empty. The random variable for the n^{th} "vacation" from the first queue is exactly $I_{n,i}$. Conditioning the waiting time in the queue on whether or not a customer arrives when the server is busy or on vacation produces

$$\mathbb{E}[W_{q,i}] = \frac{\mathbb{E}[I_{n,i}]}{\mathbb{E}[C_{n,i}]} \mathbb{E}[W_{q,i}|vac] + \frac{\mathbb{E}[D_{n,i}]}{\mathbb{E}[C_{n,i}]} \mathbb{E}[W_{q,i}|busy]. \quad (5.29)$$

A tagged customer that arrives during a vacation has to wait for the vacation to finish plus the time needed to serve the customers that arrived before him/her in the vacation. The expected remaining vacation time is $\mathbb{E}[I_{n,i}^2]/2\mathbb{E}[I_{n,i}]$ and the expected number of customer that arrived before the tagged customer is $\lambda_i \mathbb{E}[I_{n,i}^2]/2\mathbb{E}[I_{n,i}]$. This means that

$$\mathbb{E}[W_{q,i}|vac] = \frac{\mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]} (1 + \lambda_i b_i). \quad (5.30)$$

A tagged customer that arrives when the server is busy has to wait for the current customer in service to finish plus the expected time needed to serve the $L_{q,i}$ customers that arrived at (and still are in) the queue before the tagged customer did. This gives

$$\mathbb{E}[W_{q,i}|busy] = \frac{b_i^{(2)}}{2b_i} + b_i \mathbb{E}[L_{q,i}|busy]. \quad (5.31)$$

To obtain the number of customers in the queue, first realize that the expected waiting time of a customer in the system is $\mathbb{E}[W_{s,i}] = b_i + \mathbb{E}[W_{q,i}]$. Little [62] tells us that the expected number of customers, $L_{s,i}$, at the queue i (in service and in the queue) is

$$\mathbb{E}[L_{s,i}] := \lambda_i \mathbb{E}[W_{s,i}] = \rho_i + \lambda_i \mathbb{E}[W_{q,i}].$$

On the other hand,

$$\begin{aligned} \mathbb{E}[L_{s,i}] &= \frac{\mathbb{E}[D_{n,i}]}{\mathbb{E}[C_{n,i}]} \mathbb{E}[L_{s,i}|busy] + \frac{\mathbb{E}[I_{n,i}]}{\mathbb{E}[C_{n,i}]} \mathbb{E}[L_{s,i}|vac] \\ &= \rho_i (1 + \mathbb{E}[L_{q,i}|busy]) + (1 - \rho_i) \frac{\lambda_i \mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]}. \end{aligned}$$

⁶The last term in [78, formula (3.12)] is missing a factor two and there is a mix up between δ_1 and δ_2 . The expression should read $\dots + \frac{[1-\rho_1-2\rho_2(1-\rho)]\delta_2^2+(1-\rho_1)^2\delta_1^2}{2R(1-\rho+2\rho_1\rho_2)}$.

Combining these last two equations gives

$$\mathbb{E}[L_{q,i}|busy] = \frac{1}{b_i} \left(\mathbb{E}[W_{q,i}] - \frac{(1 - \rho_i)\mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]} \right). \quad (5.32)$$

Putting together equations (5.29)-(5.32) gives

$$\begin{aligned} \mathbb{E}[W_{q,i}] &= (1 - \rho_i) \frac{\mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]} (1 + \lambda_i b_i) + \rho_i \left(\frac{b_i^{(2)}}{2b_i} + \mathbb{E}[W_{q,i}] - \frac{(1 - \rho_i)\mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]} \right) \\ &= \frac{\lambda_i b_i^{(2)}}{2(1 - \rho_i)} + \frac{\mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]} \end{aligned} \quad (5.33)$$

The theorem follows by plugging in the values of $\mathbb{E}[I_{n,i}]$ and $\mathbb{E}[I_{n,i}^2]$. Notice that not once have we assumed the switchover times and the busy periods to be uncorrelated!

An alternative proof is given in Appendix 5.C. ■

5.3.3 Identical queues

Let both queues be identically distributed, i.e., $\lambda := \lambda_i$, $c(j) := c_i(j)$, $b^{(2)} := b_i^{(2)}$, $v := v_i$, $\delta^2 := \delta_i^2$, and $\hat{\rho} := \rho_i$, for $i = 1, 2$. The stability condition is then $\hat{\rho} = \lambda b < 1/2$. In stationary regime,

$$\mathbb{E}[I_{n,1}] = \frac{2v(1 - \hat{\rho})}{1 - 2\hat{\rho}} \quad (\text{Intervisit time})$$

$$\mathbb{E}[L_{n,1}^*] = \frac{2v\lambda(1 - \hat{\rho})}{1 - 2\hat{\rho}} \quad (\text{Queue length when server arrives})$$

$$\mathbb{E}[D_{n,1}] = \frac{2v\hat{\rho}}{1 - 2\hat{\rho}} \quad (\text{Busy period duration})$$

$$\mathbb{E}[C_{n,1}] = \frac{2v}{1 - 2\hat{\rho}} \quad (\text{Cycle time})$$

$$\mathbb{E}[T_{n,1}] = \frac{2v\lambda}{1 - 2\hat{\rho}} \quad (\text{Number served per cycle})$$

The second moments are

$$\begin{aligned}
 \mathbb{E}[I_{n,1}^2] &= 2v \left(\frac{\lambda b^{(2)} + 2v(1 - \hat{\rho})^2}{(1 - 2\hat{\rho})^2} \right) + \frac{2(1 - \hat{\rho})\delta^2}{1 - 2\hat{\rho}} \\
 &\quad + \frac{2(1 - 3\hat{\rho} + 3\hat{\rho}^2)}{\beta(1 - \hat{\rho})^2} \sum_{j=0}^{\infty} \alpha^j c_{12}(j) + \frac{2(1 - \hat{\rho})}{\hat{\rho}} \sum_{j=1}^{\infty} \left[\frac{c(j)}{1 - 2\hat{\rho}} + \frac{1}{\beta} c_{12}(-j) \right] \alpha^j \\
 \mathbb{E}[(L_{n,1}^*)^2] &= \lambda^2 \mathbb{E}[I_{n,1}^2] + \frac{2v\lambda(1 - \hat{\rho})}{1 - 2\hat{\rho}} \\
 \mathbb{E}[D_{n,1}^2] &= \frac{1}{(1 - \hat{\rho})^2} \left(\hat{\rho}^2 \mathbb{E}[I_{n,1}^2] + \frac{2v\lambda b^{(2)}}{1 - 2\hat{\rho}} \right) \\
 \mathbb{E}[C_{n,1}^2] &= \frac{\mathbb{E}[I_{n,1}^2]}{(1 - \hat{\rho})^2} + \frac{2v\lambda b^{(2)}}{(1 - \hat{\rho})^2(1 - 2\hat{\rho})} \\
 \mathbb{E}[T_{n,1}^2] &= \frac{1}{(1 - \hat{\rho})^2} \left(\lambda^2 \mathbb{E}[I_{n,1}^2] + \frac{2v\lambda(1 - \hat{\rho}^2 + \lambda^2 b^{(2)})}{1 - 2\hat{\rho}} \right)
 \end{aligned}$$

with

$$\alpha = \frac{\hat{\rho}^2}{(1 - \hat{\rho})^2}, \quad \beta = \frac{(1 - \hat{\rho})^4 - \hat{\rho}^4}{(1 - \hat{\rho})^2} = \frac{(1 - 2\hat{\rho})(1 - 2\hat{\rho} + 2\hat{\rho}^2)}{(1 - \hat{\rho})^2}.$$

The expected waiting time and average queue length are

$$\begin{aligned}
 \mathbb{E}[W_{q,1}] &= \frac{\lambda b^{(2)} + (1 - \hat{\rho})v}{1 - 2\hat{\rho}} + \frac{\delta^2}{2v} && \text{(Waiting time)} \\
 &\quad + \frac{1 - 3\hat{\rho} + 3\hat{\rho}^2}{v(1 - \hat{\rho})(1 - 2\hat{\rho} + 2\hat{\rho}^2)} \sum_{j=0}^{\infty} \alpha^j c_{12}(j) + \frac{1 - 2\hat{\rho}}{2v\hat{\rho}} \sum_{j=1}^{\infty} \left[\frac{c(j)}{1 - 2\hat{\rho}} + \frac{1}{\beta} c_{12}(-j) \right] \alpha^j \\
 \mathbb{E}[L_{s,1}] &= \lambda \mathbb{E}[W_{q,1}] + \hat{\rho} && \text{(Average queue length)}
 \end{aligned}$$

In the case of uncorrelated switchover times these expressions coincide with the known results for the intervisit time (equation (4.21) in [77]), number of customers at polling instant (equation (4.12b) in [77]), duration of a busy period (4.15b), cycle time (4.24), number of customers served per cycle (4.18b), waiting time (4.33b), and the average queue length (4.34).

5.3.4 Heavily Unbalanced Traffic

What happens to the waiting times if we let the traffic at the second queue approach zero? By letting $\lambda_2 \rightarrow 0$ in (5.28) we see that the expected waiting times,

$$\begin{aligned}\mathbb{E}[W_{q,1}] &= \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{R}{2} + \frac{\Delta^2}{2R}, \\ \mathbb{E}[W_{q,2}] &= \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)^2} + \frac{R}{2(1-\rho_1)} + \left(\frac{\Delta^2}{2} - \rho_1(1-\rho_1)\delta_2^2 \right) \frac{1}{R(1-\rho_1)},\end{aligned}$$

no longer depends on the correlation. Because of this the conclusions of [77, page 83] hold which state that if $\rho_1 \geq 0.5$, or if all switching times of the second queue are constant ($\delta_2 = 0$), we have $\mathbb{E}[W_{q,2}] \geq \mathbb{E}[W_{q,1}]$.

5.3.5 Switchover Times Equal to Zero

Let us examine what happens to the expected waiting time when the switchover times are constant and equal to zero. By looking at the expressions for the expected waiting times (equation (5.28)), we see that the variances and (cross) correlations of the switchover times must go faster to zero than the mean switchover times does since otherwise the expected waiting times could explode (for example, $\lim_{R \rightarrow 0} \delta_2^2/R$ must go to zero). Sending the appropriate variables to zero in the right order produces

$$\mathbb{E}[W_{q,1}] = \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{\lambda_1 \rho_2^2 b_1^{(2)} + \lambda_2 (1-\rho_1)^2 b_2^{(2)}}{2(1-\rho_1)(1-\rho)(1-\rho+2\rho_1\rho_2)}.$$

This is naturally no longer dependent on the correlation and is identical to (4.36b) in [77].

5.4 Exhaustive/Gated Service System

5.4.1 Introduction

Now let the first queue be served exhaustively and the second be served in a gated manner. The time needed to serve N customers in the second queue in the n^{th} cycle is denoted by $\mathcal{S}_{n,2}(N)$. Naturally, $\mathbb{E}[\mathcal{S}_{n,2}(N)] = b_2 \mathbb{E}[N]$. The service time of the second queue, $S_{n,2}$, satisfies the following recursive relationship

$$S_{n+1,2} = \mathcal{S}_{n+1,2}(\mathcal{N}_{n,2}(S_{n,2} + V_{n,2}) + \mathcal{N}_{n+1,2}(D_{n+1,1} + V_{n+1,1})).$$

At the same time, the time the server works per cycle at the first queue, $D_{n,1}$, satisfies

$$D_{n+1,1} = \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(V_{n,1} + S_{n,2} + V_{n,2})).$$

By combining these two expressions we obtain the following theorems.

Theorem 5.4.1 (SRE for the exhaustive/gated system). *The SRE for the service time at the gated queue is given by*

$$S_{n+1,2} = \mathcal{X}_n(S_{n,2}) + \mathcal{Y}_n, \quad (5.34)$$

with

$$\mathcal{X}_n(\cdot) := \mathcal{S}_{n+1,2} \left(\mathcal{N}_{n,2}(\cdot) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(\cdot))) \right), \quad (5.35)$$

$$\mathcal{Y}_n := \mathcal{S}_{n+1,2} \left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})) + V_{n+1,1}) \right). \quad (5.36)$$

Note that $\mathbb{E}[\mathcal{X}_n(S_{n,2})] = \gamma \mathbb{E}[S_{n,2}]$ where $\gamma := \frac{\rho_2}{1-\rho_1}$. \diamond

Theorem 5.4.2 *If $\rho < 1$ then there exists a stationary ergodic regime $S_{n,2}^*$ which satisfies (5.34).* \diamond

The proof of the last theorem is omitted due to the strong similarity with the proof of Theorem 5.3.2.

5.4.2 Performance Measures

Starting with the service time at the second queue, the next couple of sections will present a number of performance measures of the mixed exhaustive/gated service system.

5.4.2.1 Service Time Second Queue

Theorem 5.4.3 (Service Time Second Queue) *Under the stationary regime the expected time per cycle spend on service time at the second queue is given by*

$$\mathbb{E}[S_{n+1,2}] = \frac{\rho_2 R}{1-\rho}. \quad (5.37)$$

The second moment of the same variable is

$$\begin{aligned} & \omega \mathbb{E}[S_{n+1,2}^2] \\ &= \frac{R}{1-\rho} \left(\lambda_1 b_1^{(2)} + \frac{\lambda_2 (1-\rho_1)^2 b_2^{(2)}}{\rho_2^2} \right) + (1-2\rho_1(1-\rho)) \delta_1^2 + \delta_2^2 + \left(\frac{1-\rho_1+\rho_2}{1-\rho} \right) R^2 \\ & \quad + 2\rho \sum_{j=0}^{\infty} c_{12}(j) \gamma^j + 2 \sum_{j=1}^{\infty} \left(c_2(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) \right) \gamma^j. \end{aligned} \quad (5.38)$$

with

$$\gamma := \frac{\rho_2}{1 - \rho_1}, \quad \omega := \frac{(1 - \rho_1)^2 - \rho_2^2}{\rho_2^2}. \quad \diamond$$

The proof forwarded to Appendix 5.D

5.4.2.2 Cycle Time Starting from the Second Queue

The cycle time, $C_{n,2}$, starting from the server arriving at the second queue can be derived by using the relationship

$$S_{n+1,2} = \mathcal{S}_{n+1,2}(\dot{\mathcal{N}}_{n,2}(C_{n,2})),$$

where $\dot{\mathcal{N}}_{n,2}(\cdot)$ expresses⁷ the number of arrivals at the second queue during the n^{th} cycle, with the cycle starting from the arrival of the server at the second queue. The time the server works, per cycle, on the second queue is equal to the time needed to serve the customers that arrived at the second queue during the previous cycle. Taking the expectation ($\mathbb{E}[S_{n+1,2}] = \rho_2 \mathbb{E}[C_{n,2}]$) gives the first moment,

$$\mathbb{E}[C_{n,2}] = \frac{R}{1 - \rho}. \quad (5.39)$$

By taking the expectation of the square one obtains

$$\mathbb{E}[S_{n+1,2}^2] = \mathbb{E}[\mathcal{S}_{n+1,2}^2(\dot{\mathcal{N}}_2(C_{n,2}))] = \rho_2^2 \mathbb{E}[C_{n,2}^2] + \lambda_2 b_2^{(2)} \mathbb{E}[C_{n,2}],$$

where (5.76) was used. Substituting (5.38) into this gives the second moment of the cycle time starting from the polling instant of the second queue:

$$\begin{aligned} \eta \mathbb{E}[C_{n,2}^2] &= \frac{R}{1 - \rho} \left(\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)} \right) + (1 - 2\rho_1(1 - \rho)) \delta_1^2 + \delta_2^2 + \left(\frac{1 - \rho_1 + \rho_2}{1 - \rho} \right) R^2 \\ &\quad + 2\rho \sum_{j=0}^{\infty} c_{12}(j) \gamma^j + 2 \sum_{j=1}^{\infty} \left(c_2(j) + \frac{1 - \rho_1(2 - \rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) \right) \gamma^j. \end{aligned} \quad (5.40)$$

where $\gamma := \frac{\rho_2}{1 - \rho_1}$ and $\eta := (1 - \rho_1)^2 - \rho_2^2$.

⁷This represents a shift from the original arrival process $\mathcal{N}_{n,2}(\cdot)$. By doing so the calculations become significantly less involved. Note that this process is correlated to $\mathcal{N}_{n,2}$ and $\mathcal{N}_{n+1,2}$, but it is independent of $\mathcal{D}_{n,1}(\cdot)$ and $\mathcal{S}_{n,2}(\cdot)$.

5.4.2.3 Intervisit Time First Queue

The time between the server leaving the first queue and arriving back at the first queue is the intervisit time and it is given by

$$I_{n,1} = V_{n,1} + S_{n,2} + V_{n,2}.$$

It's expectation is straightforward,

$$\mathbb{E}[I_{n,1}] = \frac{(1 - \rho_1)R}{1 - \rho}. \quad (5.41)$$

For the second moment a little more work is needed as

$$\mathbb{E}[I_{n,1}^2] = \mathbb{E}[(V_{n,1} + S_{n,2} + V_{n,2})^2] = \delta_1^2 + \delta_2^2 + R^2 + 2\mathbb{E}[S_{n,2}(V_{n,1} + V_{n,2})] + \mathbb{E}[S_{n,2}^2]. \quad (5.42)$$

The second last term remains to be assessed. Similar to the step taking (5.80) to (5.81), Theorem 2 of [3] provides us with the answer. Applying the theorem just as in (5.81) gives

$$\begin{aligned} \mathbb{E}[S_{n,2}(V_{n,1} + V_{n,2})] &= \sum_{j=0}^{\infty} \gamma^j \mathbb{E}[\mathcal{Y}_0(V_{j+1,1} + V_{j+1,2})] \\ &= \sum_{j=0}^{\infty} \gamma^j \mathbb{E}\left[\mathcal{S}_{1,2}\left(\mathcal{N}_{0,2}(V_{0,2}) + \mathcal{N}_{1,2}(\mathcal{D}_{1,1}(\mathcal{N}_{0,1}(V_{0,1} + V_{0,2})) + V_{1,1})\right)(V_{j+1,1} + V_{j+1,2})\right] \\ &= \rho_2 \sum_{j=0}^{\infty} \gamma^j \mathbb{E}\left[\left(V_{0,2} + \frac{\rho_1}{1 - \rho_1}(V_{0,1} + V_{0,2}) + V_{1,1}\right)(V_{j+1,1} + V_{j+1,2})\right]. \end{aligned}$$

Using the covariance functions and collecting terms gives

$$\begin{aligned} \mathbb{E}[S_{n,2}(V_{n,1} + V_{n,2})] &= \rho_2 \sum_{j=0}^{\infty} \gamma^j \left(\frac{R^2 + c_{12}(-j-1) + c_2(j+1)}{1 - \rho_1} \right. \\ &\quad \left. + \frac{\rho_1(c_1(j+1) + c_{12}(j+1))}{1 - \rho_1} + c_1(j) + c_{12}(j) \right). \end{aligned}$$

Here the term R^2 can be taken out of the summation and the covariance terms can be re-indexed (for example, $\gamma \sum_{j=0}^{\infty} \gamma^j c_1(j+1) = \sum_{j=1}^{\infty} \gamma^j c_1(j)$) to give

$$\begin{aligned} \mathbb{E}[S_{n,2}(V_{n,1} + V_{n,2})] &= \frac{\rho_2}{1 - \rho} R^2 + \rho_2(c_1(0) + c_{12}(0)) \\ &\quad + \sum_{j=1}^{\infty} \left[c_{12}(-j) + c_2(j) + \rho_1(c_1(j) + c_{12}(j)) + \rho_2(c_1(j) + c_{12}(j)) \right] \gamma^j \\ &= \frac{\rho_2 R^2}{1 - \rho} + \rho_2(\delta_1^2 + c_{12}(0)) + \sum_{j=1}^{\infty} \left[c_{12}(-j) + c_2(j) + \rho(c_1(j) + c_{12}(j)) \right] \gamma^j. \end{aligned}$$

This expression can be put back into (5.42) to disclose that

$$\begin{aligned}
\mathbb{E}[I_{n,1}^2] &= \delta_1^2 + \delta_2^2 + \left(1 + \frac{2\rho_2}{1-\rho}\right) R^2 + 2\rho_2(\delta_1^2 + c_{12}(0)) \\
&\quad + 2 \sum_{j=1}^{\infty} \left[c_{12}(-j) + c_2(j) + \rho(c_1(j) + c_{12}(j)) \right] \gamma^j \\
&\quad + \frac{1}{\omega} \left[\frac{R}{1-\rho} \left(\lambda_1 b_1^{(2)} + \frac{\lambda_2(1-\rho_1)^2 b_2^{(2)}}{\rho_2^2} \right) + (1-2\rho_1(1-\rho))\delta_1^2 + \delta_2^2 + \left(1 + \frac{2\rho_2}{1-\rho}\right) R^2 \right. \\
&\quad \left. + 2\rho \sum_{j=0}^{\infty} c_{12}(j)\gamma^j + 2 \sum_{j=1}^{\infty} \left(c_2(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) \right) \gamma^j \right],
\end{aligned}$$

with $\gamma := \frac{\rho_2}{1-\rho_1}$ and $\omega := \frac{(1-\rho_1)^2 - \rho_2^2}{\rho_2^2}$. Collecting terms gives the second moment of the intervisit time of the first queue,

$$\begin{aligned}
\mathbb{E}[I_{n,1}^2] &= \frac{\lambda_1 \rho_2^2 b_1^{(2)} + \lambda_2 (1-\rho_1)^2 b_2^{(2)}}{(1-\rho)^2(1-\rho_1+\rho_2)} R + \frac{(1-\rho_1)^2 \Delta^2}{(1-\rho)(1-\rho_1+\rho_2)} + \frac{2\rho_2(1-\rho_1)(1+\rho_2)}{1-\rho_1+\rho_2} \delta_1^2 \\
&\quad + \frac{(1-\rho_1)^2}{(1-\rho)^2} R^2 + \frac{2\rho_2}{1-\rho} \left(\frac{1-\rho_1(2-\rho)}{1-\rho_1+\rho_2} \right) c_{12}(0) \\
&\quad + \frac{2(1-\rho_1)^2}{(1-\rho)(1-\rho_1+\rho_2)} \sum_{j=1}^{\infty} \left[\left(1 + \frac{\rho_2(1-\rho)}{1-\rho_1} \right) (\rho c_1(j) + c_{12}(-j)) + c_2(j) + \rho c_{12}(j) \right] \gamma^j.
\end{aligned} \tag{5.43}$$

5.4.2.4 Duration of the Busy Period at the First Queue

The time the server spends, per cycle, at the first queue is made up of

$$D_{n+1,1} = \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})),$$

the time needed to empty a queue (exhaustively), starting with the number of customers that have arrived at and accumulated at the first queue since the last time the server served the first queue, i.e., since the intervisit time $I_{n,1}$. The expected value is

$$\mathbb{E}[D_{n+1,1}] = \frac{\rho_1 R}{1-\rho_1}. \tag{5.44}$$

Taking once again the expectation over the square gives $\mathbb{E}[D_{n+1,1}^2] = \mathbb{E}[\mathcal{D}_{n+1,1}^2(\mathcal{N}_{n,1}(I_{n,1}))]$, which with (5.9) gives

$$\mathbb{E}[D_{n+1,1}^2] = \frac{\rho_1^2 \mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2} + \frac{\lambda_1 R b_1^{(2)}}{(1-\rho_1)^2(1-\rho)}, \tag{5.45}$$

as the second moment of the duration of the busy period at the first queue.

5.4.2.5 Number of Customers Waiting When Server Arrives

The number of customers, $L_{n+1,i}^*$, waiting in a queue at the moment the queue is polled is equal to either ($i = 1$) the number of customers that arrived during the last intervisit time or ($i = 2$) the number of customers that have arrived since the gate closed. This gives

$$L_{n+1,1}^* = \mathcal{N}_{n,1}(I_{n,1})$$

$$L_{n+1,2}^* = \mathcal{N}_{n,2}(C_{n,2}).$$

Thus the expected number of customers waiting at the polling instances are

$$\mathbb{E}[L_{n+1,1}^*] = \frac{\lambda_1(1 - \rho_1)R}{1 - \rho} \quad (5.46a)$$

$$\mathbb{E}[L_{n+1,2}^*] = \frac{\lambda_2 R}{1 - \rho}. \quad (5.46b)$$

The second moments follows through squaring to give (see (5.6))

$$\mathbb{E}[(L_{n+1,1}^*)^2] = \lambda_1^2 \mathbb{E}[I_{n,1}^2] + \frac{\lambda_1(1 - \rho_1)R}{1 - \rho} \quad (5.47a)$$

$$\mathbb{E}[(L_{n+1,2}^*)^2] = \lambda_2^2 \mathbb{E}[C_{n,2}^2] + \frac{\lambda_2 R}{1 - \rho}. \quad (5.47b)$$

5.4.2.6 Number of Customers Served per Cycle

The derivation of the expected number of customers served, per cycle, at the first queue is completely identical to that of the exhaustive/exhaustive service discipline given in section 5.3.2.5 (with $i = 1$). Equations (5.25) and (5.26) give

$$\mathbb{E}[T_{n,1}] = \frac{\mathbb{E}[L_{n,1}^*]}{1 - \rho_1} = \frac{\lambda_1 R}{1 - \rho}, \quad (5.48)$$

$$\begin{aligned} \mathbb{E}[T_{n,1}^2] &= \frac{1}{(1 - \rho_1)^2} \left(\mathbb{E}[(L_{n,1}^*)^2] + \frac{\lambda_1 R (\rho_1(1 - \rho_1) + \lambda_1^2 b_1^{(2)})}{1 - \rho} \right) \\ &= \frac{1}{(1 - \rho_1)^2} \left(\lambda_1^2 \mathbb{E}[I_{n,1}^2] + \frac{\lambda_1 R (1 - \rho_1^2 + \lambda_1^2 b_1^{(2)})}{1 - \rho} \right). \end{aligned} \quad (5.49)$$

The number of customers served, per cycle, at the second queue is easier since, because of its gated nature, it is equal to the number of customers waiting in the queue at the moment the queue gets polled. Thus $T_{n,2} = L_{n,2}^*$ and, in particular,

$$\mathbb{E}[T_{n,2}] = \mathbb{E}[L_{n,2}^*] = \frac{\lambda_2 R}{1 - \rho}, \quad (5.50)$$

$$\mathbb{E}[T_{n,2}^2] = \mathbb{E}[(L_{n,2}^*)^2] = \lambda_2^2 \mathbb{E}[C_{n,2}^2] + \frac{\lambda_2 R}{1 - \rho}. \quad (5.51)$$

5.4.2.7 Expected Waiting Time and Average Queue Length

Theorem 5.4.4 (Expected Waiting Times for Exhaustive/Gated Served Queues)

If queue 1 is serviced exhaustively and queue 2 has gated service, then the expected time a customer waits in queue i ($i = 1, 2$) until being served is given by

$$\mathbb{E}[W_{q,1}] = \frac{(1-\rho_1)R}{2(1-\rho)} + \frac{\rho_2(1-\rho)(1+\rho_2)\delta_1^2}{R(1-\rho_1+\rho_2)} + \frac{1-\rho_1}{2(1-\rho_1+\rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1-\rho} + \frac{\Delta^2 + 2C_1}{R} \right] \quad (5.52a)$$

$$\mathbb{E}[W_{q,2}] = \frac{(1+\rho_2)R}{2(1-\rho)} - \frac{\rho_1(1-\rho)(1+\rho_2)\delta_1^2}{R(1-\rho_1+\rho_2)} + \frac{1+\rho_2}{2(1-\rho_1+\rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1-\rho} + \frac{\Delta^2 + 2C_2}{R} \right] \quad (5.52b)$$

where

$$C_1 = \sum_{j=1}^{\infty} \left[c_2(j) + \rho c_{12}(j) + \left(1 + \frac{\rho_2(1-\rho)}{1-\rho_1} \right) (c_{12}(-j) + \rho c_1(j)) \right] \gamma^j + \frac{\rho_2(1-\rho_1(2-\rho))}{(1-\rho_1)^2} c_{12}(0)$$

$$C_2 = \sum_{j=1}^{\infty} \left[c_2(j) + \rho c_{12}(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) \right] \gamma^j + 2\rho c_{12}(0)$$

is the increase due to correlated switchover times. Here $\gamma := \frac{\rho_2}{1-\rho_1}$.

The average queue lengths,

$$\mathbb{E}[L_{s,i}] = \lambda_i \mathbb{E}[W_{q,i}] + \rho_i, \quad i = 1, 2. \quad (5.53)$$

follow immediately because of Little. \diamond

If the switchover times for each of the two queues are independent, as well as the switchover times between the two queues, then $C_1 = 0 = C_2$ and we obtain the results given in [66, formulas (25) and (28)] or [78, formula⁸ 4.1].

Proof: For a customer arriving at the first (exhaustive) queue, the derivation of the waiting time is identical to that leading to (5.33) in Theorem 5.3.4. Therefore we have

$$\mathbb{E}[W_{q,1}] = \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{\mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]}$$

Filling in the values of $\mathbb{E}[I_{n,1}]$ and $\mathbb{E}[I_{n,1}^2]$ gives the expected waiting time of a customer in the first queue.

⁸The formula presented in this reference is taken incorrectly from [66]. The first term for the waiting time for customers arriving at the second queue should contain $(1 + \rho_2)$ instead of $(1 + \rho_1)$.

Now imagine a tagged customer arriving at the second queue. The tagged customer has to wait for the cycle to finish (the expected remaining cycle time is $\mathbb{E}[C_{n,2}^2]/2\mathbb{E}[C_{n,2}]$) plus the time needed to serve the customers which arrived in the same cycle, but the before the tagged customer did (the expected number of customers arriving before the tagged customer is $\lambda_2\mathbb{E}[C_{n,2}^2]/2\mathbb{E}[C_{n,2}]$ and they require an expected total time of $\lambda_2 b_2 \mathbb{E}[C_{n,2}^2]/2\mathbb{E}[C_{n,2}]$ to serve). Therefore, the expected waiting time for a customer arriving at the second queue is

$$\mathbb{E}[W_{q,2}] = \frac{\mathbb{E}[C_{n,2}^2]}{2\mathbb{E}[C_{n,2}]}(1 + \rho_2).$$

Filling in the value for the second moment of the cycle time gives $\mathbb{E}[W_{q,2}]$. ■

5.4.3 Identical Queues

In the case of identical queues ($\lambda := \lambda_i$, $\hat{\rho} = \rho_i$, $v := v_i$, $\delta := \delta_i$, $b^{(2)} := b_i^{(2)}$, and $c(j) := c_i(j)$, for $i = 1, 2$) we have as the stability condition $\hat{\rho} = \lambda b < 1/2$ and

$$\mathbb{E}[D_{n,1}] = \frac{2\hat{\rho}v}{1 - \hat{\rho}} \quad (\text{Service Time/Busy Period Q1})$$

$$\mathbb{E}[S_{n,2}] = \frac{2\hat{\rho}v}{1 - 2\hat{\rho}} \quad (\text{Service time/Busy Period Q2})$$

$$\mathbb{E}[I_{n,1}] = \frac{2(1 - \hat{\rho})v}{1 - 2\hat{\rho}} \quad (\text{Intervisit time Q1})$$

$$\mathbb{E}[C_{n,2}] = \frac{2v}{1 - 2\hat{\rho}} \quad (\text{Cycle time Q2})$$

$$\mathbb{E}[L_{n,1}^*] = \frac{2\lambda(1 - \hat{\rho})v}{1 - 2\hat{\rho}} \quad (\text{Number Waiting Q1})$$

$$\mathbb{E}[L_{n,2}^*] = \frac{2\lambda v}{1 - 2\hat{\rho}} \quad (\text{Number Waiting Q2})$$

$$\mathbb{E}[T_{n,i}] = \frac{2\lambda v}{1 - 2\hat{\rho}} \quad (\text{Number Served per Cycle})$$

Writing out the second moments of these quantities does not provide much insight. The expected waiting times are

$$\mathbb{E}[W_{q,1}] = \frac{(1 - \hat{\rho})v}{1 - 2\hat{\rho}} + \frac{\hat{\rho}(1 - 2\hat{\rho})(1 + \hat{\rho})\delta^2}{2v} + \frac{1 - \hat{\rho}}{2} \left[\frac{2\lambda b^{(2)}}{1 - 2\hat{\rho}} + \frac{\delta^2 + C_1}{v} \right] \quad (5.54a)$$

$$\mathbb{E}[W_{q,2}] = \frac{(1 + \hat{\rho})v}{1 - 2\hat{\rho}} - \frac{\hat{\rho}(1 - 2\hat{\rho})(1 + \hat{\rho})\delta^2}{2v} + \frac{1 + \hat{\rho}}{2} \left[\frac{2\lambda b^{(2)}}{1 - 2\hat{\rho}} + \frac{\delta^2 + C_2}{v} \right] \quad (5.54b)$$

with

$$C_1 = \sum_{j=1}^{\infty} \left[2\hat{\rho}c_{12}(j) + \frac{1-2\hat{\rho}^3}{1-\hat{\rho}}c(j) + \frac{1-2\hat{\rho}^2}{1-\hat{\rho}}c_{12}(-j) \right] \gamma^j + \frac{\hat{\rho}(1-2\hat{\rho}(1-\hat{\rho}))}{(1-\hat{\rho})^2}c_{12}(0) \quad (5.55a)$$

$$C_2 = \sum_{j=1}^{\infty} \left[2\hat{\rho}c_{12}(j) + (3-4\hat{\rho}+4\hat{\rho}^2)c(j) + \frac{1-2\hat{\rho}(1-\hat{\rho})}{\hat{\rho}}c_{12}(-j) \right] \gamma^j + 4\hat{\rho}c_{12}(0) \quad (5.55b)$$

being the increase due to correlated switchover times. In this case $\gamma := \frac{\hat{\rho}}{1-\hat{\rho}}$.

The average queue lengths are given by

$$\mathbb{E}[L_{s,i}] = \lambda \mathbb{E}[W_{q,i}] + \hat{\rho}, \quad i = 1, 2. \quad (5.56)$$

5.4.4 Comparison of Waiting Times in Exhaustive/Gated System

In the subsequent sections the effect of the service discipline at the two queues will be studied. We start by examining the expected waiting times in the exhaustive/gated system. Although the discussions are about the expected waiting times, the same arguments hold for the average queue length.

5.4.4.1 Single Server Queue with Correlated Vacations

By turning off one of the queues one obtains an M/G/1 queue with multiple correlated vacations. Let us start by turning off the second queue (by setting $\lambda_2 = 0$, $\rho_2 = 0$, $v_2 = 0$, and $v_2^{(2)} = 0$, which leads to $c_2(j) = 0$, $c_{12}(j) = 0$, and $\gamma = 0$) to end up with an exhaustively served M/G/1 queue where the expected waiting time

$$\mathbb{E}[W_{q,1}] = \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{v_1^{(2)}}{2v_1} \quad (\text{Exhaustive M/G/1})$$

is independent of the correlation between the vacations! This result was previously pointed out in [3, paragraph 3.6] which causes it to correspond to the expression for the expected waiting time but with i.i.d. vacation times [79, page 123].

On the other hand, by turning off the first queue in the exhaustive/gated system, we are left with an M/G/1 queue with a gated service discipline. After setting the appropriate parameters to zero we obtain

$$\mathbb{E}[W_{q,2}] = \frac{\lambda_2 b_2^{(2)}}{2(1-\rho_2)} + \frac{v_2^{(2)}}{2v_2} + \frac{\rho_2 v_2}{1-\rho_2} + \frac{1}{v_2} \sum_{j=1}^{\infty} c_2(j) \rho_2^j \quad (\text{Gated M/G/1})$$

as the waiting time of a customer arriving at a gated M/G/1 queue with correlated vacations. If there is no correlation then this expression is in agreement with the result previously obtained in [3, Theorem 5] and [79, equation (5.24a)].

It is interesting to compare the difference between these two waiting times due to the server behaving differently. Assuming queues with identical parameters (by dropping the indices of ones and twos and setting $\hat{\rho} := \rho_1 = \rho_2$) we see that

$$\mathbb{E}[W]_{gated\ M/G/1} - \mathbb{E}[W]_{exhaustive\ M/G/1} = \frac{\hat{\rho}v}{1 - \hat{\rho}} + \frac{1}{v} \sum_{j=1}^{\infty} c(j)\hat{\rho}^j, \quad (5.57)$$

where the first term on the right hand side is the mean length of a service period (which is the same for the exhaustive and the gated service systems). If there is no correlation, then it is well known that the expected waiting time in an exhaustively served queue is less than that in a gated serviced queue. In the presence of correlated vacation times this difference is larger but remains a surprisingly simple expression.

5.4.4.2 Nonidentical Queues

The difference between the waiting times at the two queues is found by subtracting (5.52) from (5.52a) to give $\mathbb{E}[W_{q,2}] - \mathbb{E}[W_{q,1}] =$

$$\begin{aligned} & \frac{\rho R}{2(1-\rho)} - \frac{\rho(1-\rho)(1+\rho_2)\delta_1^2}{R(1-\rho_1+\rho_2)} + \frac{\rho}{2(1-\rho_1+\rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1-\rho} + \frac{\Delta^2 + 2((1+\rho_2)C_2 - (1-\rho_1)C_1)}{\rho R} \right] \\ &= \frac{\rho R}{2(1-\rho)} - \frac{\rho(1-2(\rho_1+\rho_2\rho))\delta_1^2}{2R(1-\rho_1+\rho_2)} + \frac{\rho}{2(1-\rho_1+\rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1-\rho} + \frac{\delta_2^2 + 2((1+\rho_2)C_2 - (1-\rho_1)C_1)}{\rho R} \right], \end{aligned}$$

with C_1 and C_2 defined in Theorem 5.4.4. From this expression we see that if $1-2(\rho_1+\rho_2\rho) > 0$, and $C_1 = 0 = C_2$, and if δ_1^2 is sufficiently large, then it may very well be possible that the expected waiting time at the gated queue is smaller than the expected waiting time at the exhaustive queue! However, the range of parameter settings for which this is the case is fairly small. In particular, in the next sections it will be shown that this does not happen when there are no switchover times or if the system is heavily loaded.

5.4.4.3 Identical Queues

Assume that, except for the service discipline, the two queues in the exhaustive/gated system have equal parameter settings. The difference between the waiting time of a customer arriving at the gated or at the exhaustive queue is

$$\mathbb{E}[W_{q,2}] - \mathbb{E}[W_{q,1}] = \hat{\rho} \left(\frac{2\lambda b^{(2)} + 2v}{1 - 2\hat{\rho}} + \frac{\hat{\rho}(1 + 2\hat{\rho})\delta^2 + C_2 - C_1}{v} \right),$$

with C_1 and C_2 defined in (5.55b). Since it is not necessarily true that $C_2 \geq C_1$, we can deduct that, under equal parameter settings and correlated switchover times, *the average*

waiting time at the gated queue can be smaller than the average waiting time at the exhaustive queue.

5.4.4.4 Switchover Times Equal to Zero

Let us now examine the two service disciplines in the interesting situation when the switchover times are equal to zero. By looking at the expressions for the expected waiting times (equation (5.52b)), we see that the variances and (cross) correlations of the switchover times must go faster to zero than the mean switchover times does since otherwise the expected waiting times could explode (for example, $\lim_{R \rightarrow 0} \delta_1^2/R$ must go to zero).

By first sending the second moments and the (cross) correlations of the switchover times to zero, and then the mean switchover times to zero, we see that the expected waiting times for the exhaustive/gated service discipline are

$$\begin{aligned}\mathbb{E}[W_{q,1}] &= \frac{1 - \rho_1}{2(1 - \rho_1 + \rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1 - \rho} \right], \\ \mathbb{E}[W_{q,2}] &= \frac{1 + \rho_2}{2(1 - \rho_1 + \rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1 - \rho} \right].\end{aligned}$$

From this it is clear that a customer arriving at the second queue is expected to wait

$$\mathbb{E}[W_{q,2}] - \mathbb{E}[W_{q,1}] = \frac{\rho}{2(1 - \rho_1 + \rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1 - \rho} \right]$$

longer than a customer that arrives at the first queue.

5.4.4.5 Heavily Loaded System

Let us now consider what happens if the system is heavily loaded (ρ close to 1). In this case (with the exception of the case where $\rho_1 \rightarrow 1$ and $\rho_2 \rightarrow 0$, in which case we refer to section 5.4.4.1) the expected waiting times (taken from 5.52b),

$$\begin{aligned}\mathbb{E}[W_{q,1}] &\rightarrow \frac{(1 - \rho_1)R}{2(1 - \rho)} + \frac{1 - \rho_1}{2(1 - \rho_1 + \rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1 - \rho} \right] \\ \mathbb{E}[W_{q,2}] &\rightarrow \frac{(1 + \rho_2)R}{2(1 - \rho)} + \frac{1 + \rho_2}{2(1 - \rho_1 + \rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1 - \rho} \right]\end{aligned}$$

explode due to the factor $1 - \rho$ in the denominators. As we can see, the correlation no longer plays a role. The difference between the expected waiting times tends towards

$$\mathbb{E}[W_{q,2}] - \mathbb{E}[W_{q,1}] \rightarrow \frac{\rho R}{2(1 - \rho)} + \frac{\rho}{2(1 - \rho_1 + \rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1 - \rho} \right]$$

which is always positive.

More discussions on the differences arising due to the server behaviour can be found in the examples of the next section.

5.5 Examples

In the following sections we will consider a number of examples in which the sequences of switchover times are correlated. The covariance functions will be calculated explicitly after which the effect of the correlation on the waiting times will be studied. In all of the examples the expected waiting time of a customer arriving at the first queue of the exhaustive/exhaustive system is given by (5.28), whereas in the exhaustive/gated system the expected waiting times are given by (5.52b). The difference between each of the examples is that C , C_1 , and C_2 take on different values.

5.5.1 Example 1: Correlated Switchover Times

Consider a sequence of switchover times where there is no correlation between the switchover times of the two queues (this gives $c_{12}(j) = 0$, for $j \in \mathbb{Z}$). Let the individual sequence of switchover times per queue satisfy

$$V_{n+1,i} = x_i V_{n,i} + (1 - x_i) \varepsilon_{n,i}, \quad i = 1, 2, \quad (5.58)$$

where $x_i \in [0, 1)$ is a constant and $\varepsilon_{n,i}$ are positive i.i.d. variables with finite expectation $\mathbb{E}[\varepsilon_{n,i}] =: \bar{\varepsilon}_i$ and second moment $\mathbb{E}[\varepsilon_{n,i}^2] =: \varepsilon_i^{(2)}$. The parameter x_i determines the amount of correlation in the sequence; with $x_i = 0$ the sequence is i.i.d., whereas when x_i tends to one the correlation is maximal. Notice that there exists a stationary ergodic sequence of switchover times which satisfies (5.58). By taking the expectation it follows that $\mathbb{E}[V_{n+1,i}] = x_i \mathbb{E}[V_{n,i}] + (1 - x_i) \bar{\varepsilon}_i$. Due to the stationarity of the process $\mathbb{E}[V_{0,i}] = \mathbb{E}[V_{n,i}] = v_i$ is independent of x_i , and therefore $v_i = \bar{\varepsilon}_i$. A similar relationship can be derived for the second moments by taking the expectation over the square of (5.58) to give

$$\begin{aligned} \mathbb{E}[V_{n+1,i}^2] &= x_i^2 \mathbb{E}[V_{n,i}^2] + (1 - x_i)^2 \mathbb{E}[\varepsilon_{n,i}^2] \\ &\quad + 2x_i(1 - x_i) \mathbb{E}[\varepsilon_{n,i}] \mathbb{E}[V_{n,i}]. \end{aligned}$$

Due to the stationarity ($\mathbb{E}[V_{n+1,i}^2] = \mathbb{E}[V_{n,i}^2] = v_i^{(2)}$) this implies that

$$v_i^{(2)} = \frac{(1 - x_i) \varepsilon_i^{(2)} + 2x_i \bar{\varepsilon}_i v_i}{1 + x_i},$$

which gives a second relationship (since $\bar{\varepsilon}_i = v_i$),

$$\delta_i^2 = \frac{1 - x_i}{1 + x_i} \text{Var}(\varepsilon_{n,i}).$$

Thus we see that for $x_i \in [0, 1)$ there exists a $\varepsilon_{n,i}$ such that any desired values of v_i and δ_i^2 can be obtained. Now we will derive the covariance functions and the expected waiting time.

By iterating (5.58) a number of times it is quickly seen that

$$V_{n,i} = x_i^n V_{0,i} + (1 - x_i) \sum_{k=0}^{n-1} \varepsilon_{n-1-k,i} x_i^k. \quad (5.59)$$

From this we obtain

$$\mathbb{E}[V_{0,i} V_{j,i}] = x_i^j \mathbb{E}[V_{0,i}^2] + (1 - x_i^j) \bar{\varepsilon}_i \mathbb{E}[V_{0,i}] = x_i^j v_i^{(2)} + (1 - x_i^j) \bar{\varepsilon}_i v_i.$$

This means that the covariance functions, $c_i(j) = \mathbb{E}[V_{0,i} V_{j,i}] - \mathbb{E}[V_{0,i}] \mathbb{E}[V_{j,i}]$, are given by

$$c_i(j) = x_i^j v_i^{(2)} + (1 - x_i^j) \bar{\varepsilon}_i v_i - v_i (x_i^j v_i + (1 - x_i^j) \bar{\varepsilon}_i) = x_i^j (v_i^{(2)} - v_i^2) = x_i^j \delta_i^2. \quad (5.60)$$

Since

$$\sum_{j=1}^{\infty} c_i(j) \alpha^j = \delta_i^2 \sum_{j=1}^{\infty} (\alpha x_i)^j = \frac{\alpha x_i \delta_i^2}{1 - \alpha x_i}, \quad (5.61)$$

we have from Theorem 5.3.4 that the expected waiting time in the exhaustive/exhaustive system is given by (5.28) with

$$C := \frac{\alpha x_1 \delta_1^2}{1 - \alpha x_1} + \frac{\alpha x_2 \delta_2^2}{1 - \alpha x_2} \left(1 + \frac{(1 - \rho)^2}{\rho_1 (1 - \rho_1)} \right)$$

and $\alpha = \frac{\rho_1 \rho_2}{(1 - \rho_1)(1 - \rho_2)}$.

Equivalently, in the exhaustive/gated system we have from equation (5.61) and from Theorem 5.4.4 that the expected waiting times are given by (5.52b) where $\gamma = \frac{\rho_2}{1 - \rho_1}$ and

$$\begin{aligned} C_1 &= \frac{\gamma x_1 \delta_1^2}{1 - \gamma x_1} \left(1 + \frac{\rho_2 (1 - \rho)}{1 - \rho_1} \right) \rho + \frac{\gamma x_2 \delta_2^2}{1 - \gamma x_2} \\ C_2 &= \frac{\gamma x_1 \delta_1^2}{1 - \gamma x_1} \left(1 + \frac{\rho_1 (1 - \rho)^2}{\rho_2} \right) + \frac{\gamma x_2 \delta_2^2}{1 - \gamma x_2}. \end{aligned}$$

Numerical examples of the influence of the correlation on the expected waiting times can be found in Figure 5.2. Shown in each of the figures is the expected waiting time divided by the expected waiting time for uncorrelated sequences of switchover times. There are Poisson arrivals with $\lambda_i = 0.4$. The first two moments of the switchover times are always kept fixed (first moment for each of the switchover time distributions is fixed at $v_i = 3$) and the service times are taken to be exponential with $b_i = 0.4$ or $b_i = 1.2$.

Based on the figures and equations above the following important conclusions can be made:

- If $x_1 = x_2 = 0$ then there is no correlation between the sequences of switchover times and $C = 0$, $C_1 = 0$, and $C_2 = 0$;
- The increase in expected waiting times due to correlated switchover times can be up to several times (3.5 times in the example) the expected waiting times if there would be uncorrelated switchover times.
- The increase in the expected waiting times due to correlation grows linearly with the variance δ_i^2 of the switchover times;
- Under light traffic (α and γ are small and so) the increase in the expected waiting time is (approximately) linear in x_i .
- Under heavy traffic α and γ are close to one and, due to the factor $1 - \alpha x_i$ or $1 - \gamma x_i$ in the denominators, the increase in waiting time due to correlated switchover times can be significant. Hence the presence of correlation has the biggest impact on the waiting time if the system has a heavy load (and the switching times have a high variance). This can be seen clearly in Figure 5.2.
- It can be shown that, under identical parameter setting, in the exhaustive/gated system the expected waiting time at the exhaustive queue is always smaller than at the gated queue. In addition to this, we see from Figure 5.2 that in lightly loaded systems the gated queue (Q2) suffers most from correlated switchover times whereas in heavily loaded traffic both queues are effected (relatively) equally by the correlated switchover times.

5.5.2 Example 2: Stochastic Recursive Sequence of Switchover Times

Consider a sequence of switchover times which satisfy the following stochastic recursive relationship

$$V_{n+1,i} = \mathcal{F}_{n,i}(V_{n,i}) + \mathcal{E}_{n,i}, \quad (5.62)$$

where $\mathcal{F}_{n,i}(\cdot)$ are independent, infinitely divisible stochastic processes with $\mathbb{E}[\mathcal{F}_{n,i}(T)] = x_i \mathbb{E}[T]$ and $\mathbb{E}[\mathcal{F}_{n,i}^2(T)] = x_i^{(2)} \mathbb{E}[T^2] + y_i \mathbb{E}[T]$. Here $x_i \in [0, 1)$, $y_i \geq 0$ and $x_i^{(2)} \geq x_i^2$. The sequence $\mathcal{E}_{n,i}$ is a sequence of independent variables with $\mathbb{E}[\mathcal{E}_{n,i}] = \bar{\varepsilon}_i$ and $\mathbb{E}[\mathcal{E}_{n,i}^2] = \varepsilon_i^{(2)}$. Iterating gives

$$V_{n,i} = \left(\prod_{k=0}^{n-1} \mathcal{F}_{k,i} \right) V_{0,i} + \sum_{k=0}^{n-1} \left(\prod_{l=k+1}^{n-1} \mathcal{F}_{l,i} \right) \mathcal{E}_{k,i},$$

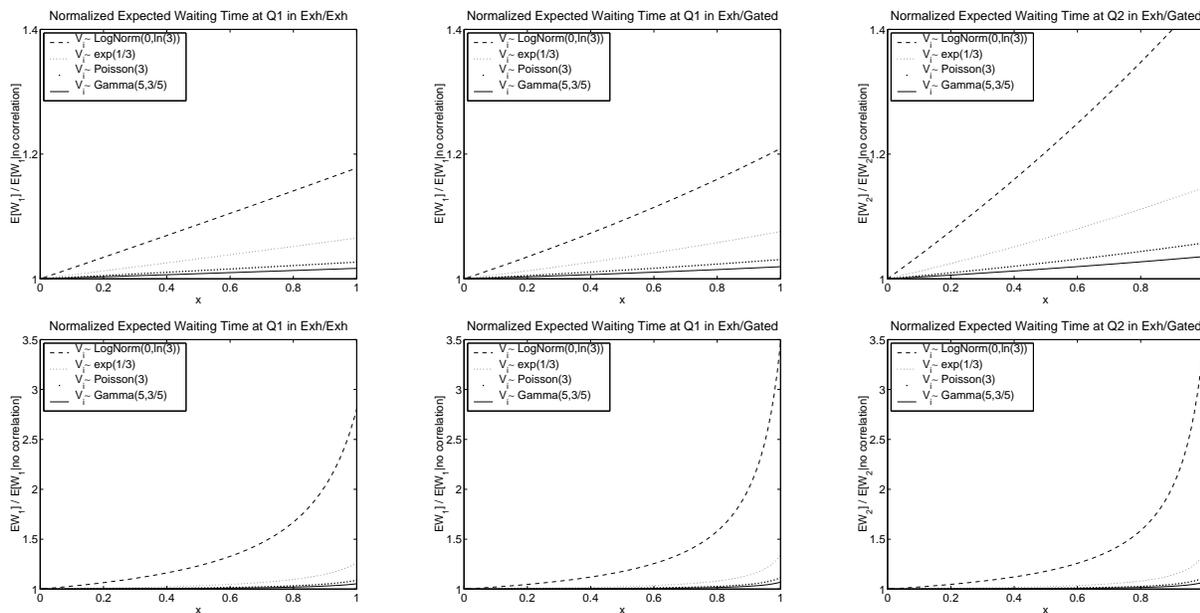


Figure 5.2: *Example: Correlated Switchover Times.* The expected waiting time divided by the expected waiting time with uncorrelated switchover times. The different lines correspond to different switchover time distributions, all with mean $v_i = 3$. Here $x := x_1 = x_2$ determines the level of correlation, there is no cross correlation, the service times are exponential, and $\lambda_i = 0.4$. The top figures are with mean service times $b_i = 0.4$ ($\rho = 0.32$) whereas the second row of figures are under heavy traffic with $b_i = 1.2$ ($\rho = 0.96$). The first column of figures correspond to the exhaustive/exhaustive system, whereas the second and third column of figures show the normalized waiting times for, respectively, the exhaustive queue (Q1) and the gated queue (Q2) in the exhaustive/gated system.

and so

$$\mathbb{E}[V_{n,i}] = x_i^n v_i + \bar{\varepsilon}_i \sum_{k=0}^{n-1} x_i^k = x_i^n v_i + \frac{1 - x_i^n}{1 - x_i} \bar{\varepsilon}_i.$$

This leads to the condition $v_i = \frac{\bar{\varepsilon}_i}{1 - x_i}$. The second moment of the switchover times is found by taking the expectation over the square of (5.62). Doing this produces

$$\mathbb{E}[V_{n+1,i}^2] = \mathbb{E}[\mathcal{F}_{n,i}^2(V_{n,i})] + \mathbb{E}[\mathcal{E}_{n,i}^2] + 2\mathbb{E}[\mathcal{F}_{n,i}(V_{n,i}) \cdot \mathcal{E}_{n,i}] = x_i^{(2)} v_i^{(2)} + y_i v_i + \varepsilon_i^{(2)} + 2x_i v_i \bar{\varepsilon}_i.$$

Since, by definition, $\mathbb{E}[V_{n+1,i}^2]$ also equals $v_i^{(2)}$, we have

$$v_i^{(2)} = \frac{\varepsilon_i^{(2)} + (y_i + 2x_i \bar{\varepsilon}_i) v_i}{1 - x_i^{(2)}}.$$

Thus we see that there exists a $\mathcal{E}_{n,i}$ which generates a sequence $V_{n,i}$ with arbitrary correlation and any desired values of v_i and $v_i^{(2)}$. Similarly, it is easy to show that

$$\mathbb{E}[V_{0,i} V_{n,i}] = x_i^n v_i^{(2)} + \frac{1 - x_i^n}{1 - x_i} \bar{\varepsilon}_i v_i.$$

This gives the covariance functions

$$c_i(j) = \mathbb{E}[V_{0,i} V_{j,i}] - \mathbb{E}[V_{0,i}] \mathbb{E}[V_{j,i}] = x_i^j v_i^{(2)} + \frac{1 - x_i^j}{1 - x_i} \bar{\varepsilon}_i v_i - v_i \left(x_i^j v_i + \frac{1 - x_i^j}{1 - x_i} \bar{\varepsilon}_i \right) = x_i^j \delta_i^2$$

which are completely identical to (5.60)! This means that if the first two moments of the switchover times are kept fixed while varying $x \in [0, 1)$, that the expected waiting times are once again given by (5.28) and (5.52b) with C , C_1 , and C_2 as given in the first example. Furthermore, the conclusions of the first example also hold here.

As a special case of (5.62) we can take $V_{n+1,i} = x_i V_n + \varepsilon_{n,i}$ where $x_i \in [0, 1)$ is a constant and $\varepsilon_{n,i}$ is a positive sequence of i.i.d. variables.

5.5.3 Example 3: Identical Switchover Times

Set $V_{n,2} = V_{n,1}$. This introduces cross-correlation between the two sequences of switchover times and it gives $v_2 = v_1$ and $\delta_2^2 = \delta_1^2$. In addition to this let $V_{n+1,1} = x V_{n,1} + (1 - x) \varepsilon_{n,1}$ just as in the first example. From (5.60) we have $c_1(j) = x^j \delta_1^2$ after which

$$\begin{aligned} c_2(j) &= \mathbb{E}[V_{0,2} V_{j,2}] - \mathbb{E}[V_{0,2}] \mathbb{E}[V_{j,2}] = \mathbb{E}[V_{0,1} V_{j,1}] - \mathbb{E}[V_{0,1}] \mathbb{E}[V_{j,1}] = c_1(j) = x^j \delta_1^2 \\ c_{12}(j) &= \mathbb{E}[V_{0,1} V_{j,2}] - \mathbb{E}[V_{0,1}] \mathbb{E}[V_{j,2}] = \mathbb{E}[V_{0,1} V_{j,1}] - \mathbb{E}[V_{0,1}] \mathbb{E}[V_{j,1}] = c_1(j) = x^j \delta_1^2 \\ c_{12}(-j) &= \mathbb{E}[V_{0,2} V_{j,1}] - \mathbb{E}[V_{0,2}] \mathbb{E}[V_{j,1}] = \mathbb{E}[V_{0,1} V_{j,1}] - \mathbb{E}[V_{0,1}] \mathbb{E}[V_{j,1}] = c_1(j) = x^j \delta_1^2 \end{aligned}$$

immediately follow. This means that ($i = 1, 2$)

$$\sum_{j=1}^{\infty} c_i(j)\alpha^j = \sum_{j=1}^{\infty} c_{12}(j)\alpha^j = \sum_{j=1}^{\infty} c_{12}(-j)\alpha^j = \frac{\alpha x \delta_1^2}{1 - \alpha x}$$

can all be plugged into Theorem 5.3.4 so that the expected waiting time in the exhaustive/exhaustive system is given by (5.28) with

$$C = \frac{\alpha x \delta_1^2}{1 - \alpha x} \left(\frac{1 - \rho_2(1 - \alpha)}{\alpha x} + 2 + \frac{(1 - \rho)^2}{\rho_1(1 - \rho_1)} + \frac{1 - \rho_2}{\rho_1} \right),$$

$$\psi := \frac{1 - \rho_1}{2v_1(1 - \rho + 2\rho_1\rho_2)}, \text{ and } \alpha = \frac{\rho_1\rho_2}{(1 - \rho_1)(1 - \rho_2)}.$$

Equivalently, the expected waiting times in the exhaustive/gated system are given by (5.52b) with

$$C_1 = \frac{\gamma x \delta_1^2}{1 - \gamma x} \left(2 + \frac{\rho_2(1 - \rho)}{1 - \rho_1} \right) (1 + \rho) + \frac{\rho_2(1 - \rho_1(2 - \rho))}{(1 - \rho_1)^2} \delta_1^2,$$

$$C_2 = \frac{\gamma x \delta_1^2}{1 - \gamma x} \left(1 + \frac{1 - \rho_1(2 - \rho)}{\rho_2} \right) (1 + \rho) + 2\rho \delta_1^2.$$

To get a feeling of the impact of the cross correlation, the expected waiting times are plotted in Figure 5.3 for various switchover time distributions and traffic loads. Shown in each of the figures is the expected waiting time divided by the expected waiting time for uncorrelated sequences of switchover times. There are Poisson arrivals with $\lambda_i = 0.4$. The first two moments of the switchover times are always kept fixed (first moment for each of the switchover time distributions is fixed at $v_i = 3$) and the service times are taken to be exponential with $b_i = 0.4$ or $b_i = 1.2$.

Striking is the impact of the cross correlation on the waiting times. For example, if there is no correlation within each sequence of switchover times ($x = 0$), then there is still an increase in the expected waiting time due to the cross-correlation. For the exhaustive/exhaustive system this increase is $(1 - \rho_2(1 - \alpha))\psi\delta_1^2$ and for the exhaustive/gated system this increase is given by $\frac{\rho_2(1 - \rho_1(2 - \rho))}{(1 - \rho_1)^2}\delta_1^2$ and $2\rho\delta_1^2$ for, respectively, the exhaustive and the gated queue. For exponentially distributed switchover times this can mean an increase of tens of percents in the expected waiting time. Besides this, all of the conclusions made in the first example also hold here, with the exception that the increase in expected waiting time can up to a factor 5.

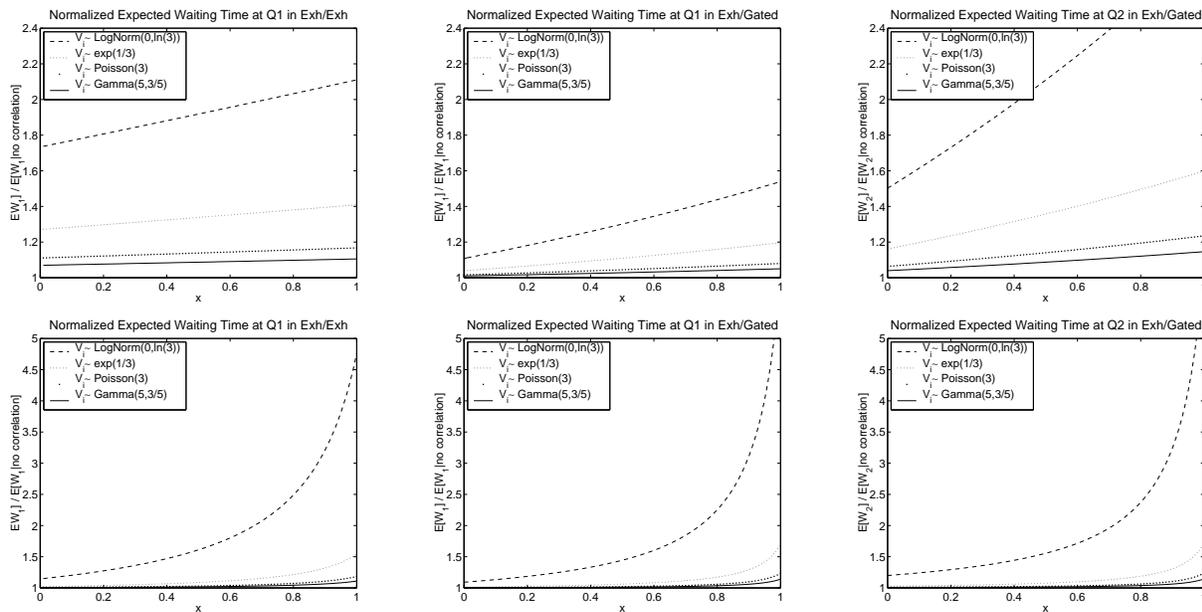


Figure 5.3: *Example: Identical Switchover Times.* The expected waiting time divided by the expected waiting time if there would be no correlation between the switchover times. The different lines correspond to different switchover time distributions. Here x determines the level of correlation, cross correlation is introduced by setting $V_{n,2} = V_{n,1}$, the service times are exponential, and $\lambda_i = 0.4$. The figures in the first row are with mean service time $b_i = 0.4$ ($\rho = 0.32$) and the figures on the bottom row are under heavy traffic with $b_i = 1.2$ ($\rho = 0.96$). The first column of figures correspond to the exhaustive/exhaustive system, whereas the second and third column of figures show the normalized waiting times for, respectively, the exhaustive queue (Q1) and the gated queue (Q2) in the exhaustive/gated system.

5.5.4 Example 4: Switchover Times Coming from the Same Sequence

As a last example consider the case where the two sequences of switchover times come from a single sequence:

$$Y_{n+1} = xY_n + (1-x)\varepsilon_n, \quad \begin{aligned} V_{n,1} &:= Y_{2n} \\ V_{n,2} &:= Y_{2n+1}. \end{aligned} \quad (5.63)$$

Here Y_n is a stationary ergodic sequence, $\mathbb{E}[\varepsilon_n] = \varepsilon$, and $\mathbb{E}[\varepsilon_n^2] = \varepsilon^{(2)}$. In this case there is a large correlation between the two sequences, but it is not as strong as in example 3. Once again we will derive an explicit expression for the expected waiting times. Equation (5.63) can be rewritten as

$$\begin{aligned} V_{n,1} &= xV_{n-1,2} + (1-x)\varepsilon_{2n} \\ V_{n,2} &= xV_{n,1} + (1-x)\varepsilon_{2n+1}. \end{aligned}$$

Iterating twice gives

$$\begin{aligned} V_{n,1} &= x^2V_{n-1,1} + (1-x)(x\varepsilon_{2n-1} + \varepsilon_{2n}) \\ V_{n,2} &= x^2V_{n-1,2} + (1-x)(x\varepsilon_{2n} + \varepsilon_{2n+1}), \end{aligned}$$

and iterating a few more times shows that

$$\begin{aligned} V_{n,1} &= x^{2n}V_{0,1} + (1-x) \sum_{k=0}^{n-1} (x\varepsilon_{2(n-k)-1} + \varepsilon_{2(n-k)}) x^{2k} \\ V_{n,2} &= x^{2n}V_{0,2} + (1-x) \sum_{k=0}^{n-1} (x\varepsilon_{2(n-k)} + \varepsilon_{2(n-k)+1}) x^{2k}. \end{aligned}$$

Taking the expectation over this gives

$$\mathbb{E}[V_{n,i}] = x^{2n}v_i + (1-x^{2n})\varepsilon, \quad i = 1, 2,$$

which with the stationarity gives $v_i = \varepsilon$. This last result can be generalized to saying that all of the moments of $V_{n,1}$ and $V_{n,2}$ are equal to each other, in particular, $\delta_1^2 = \delta_2^2$. This follows from the stationarity of Y_n , i.e., $\mathbb{E}[V_{n,2}^k] = \mathbb{E}[Y_{2n+1}^k] = \mathbb{E}[Y_{2n}^k] = \mathbb{E}[V_{n,1}^k]$.

The covariance functions can be derived in the same way as in the previous examples,

$$c_1(j) = \mathbb{E}[V_{0,1}V_{j,1}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{j,1}] = x^{2j}(\mathbb{E}[V_{0,1}^2] - \mathbb{E}[V_{0,1}]^2) = x^{2j}\delta_1^2. \quad (5.64)$$

In a similar manner it can be derived that $c_2(j) = x^{2j}\delta_2^2 = x^{2j}\delta_1^2$. The cross covariance function for $j \in \mathbb{N}$ is

$$\begin{aligned} c_{12}(j) &= \mathbb{E}[V_{0,1}V_{j,2}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{j,2}] \\ &= \mathbb{E}[V_{0,1}(xV_{j,1} + (1-x)\varepsilon_{2j+1})] - \mathbb{E}[V_{0,1}]\mathbb{E}[xV_{j,1} + (1-x)\varepsilon_{2j+1}] \\ &= x\mathbb{E}[V_{0,1}V_{j,1}] + x\mathbb{E}[V_{0,1}]\mathbb{E}[V_{j,1}] = xc_1(j) = x^{2j+1}\delta_1^2. \end{aligned} \quad (5.65)$$

Similarly,

$$\begin{aligned}
 c_{12}(-j) &= \mathbb{E}[V_{-j,2}V_{0,1}] - \mathbb{E}[V_{-j,2}]\mathbb{E}[V_{0,1}] = \mathbb{E}[V_{0,2}V_{j,1}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{j,1}] \\
 &= \mathbb{E}[V_{0,2} \cdot (xV_{j-1,2} + (1-x)\varepsilon_{2n})] - \mathbb{E}[V_{0,2}]\mathbb{E}[xV_{j-1,2} + (1-x)\varepsilon_{2n}] \\
 &= xc_2(j-1) = x^{2j-1}\delta_1^2
 \end{aligned} \tag{5.66}$$

Putting these expressions for the covariance functions into (5.19) tells us that the expected waiting time for a customer arriving at the first queue of an exhaustive/exhaustive system is given by (5.28) where $\psi := \frac{1-\rho_1}{R(1-\rho+2\rho_1\rho_2)}$, $\alpha = \frac{\rho_1\rho_2}{(1-\rho_1)(1-\rho_2)}$, and

$$C = \left(\frac{1-\rho_2(1-\alpha)}{\alpha x} + 2 + \frac{(1-\rho)^2}{\rho_1(1-\rho_1)} + \frac{1-\rho_2}{x\rho_1} \right) \frac{\alpha x^2 \delta_1^2}{1-\alpha x^2}. \tag{5.67}$$

Similarly, in the exhaustive/gated system the expected waiting times follow by putting (5.64)-(5.66) into Theorem 5.4.4. This leads to the expected waiting time being given by (5.52b) with $\gamma = \frac{\rho_2}{1-\rho_1}$ and

$$C_1 = \frac{\gamma x^2 \delta_1^2}{1-\gamma x^2} \left(\rho + \frac{1}{x} \right) \left(x + 1 + \frac{\rho_2(1-\rho)}{1-\rho_1} \right) + \frac{\rho_2(1-\rho_1(2-\rho))}{(1-\rho_1)^2} x \delta_1^2 \tag{5.68a}$$

$$C_2 = \frac{\gamma x^2 \delta_1^2}{1-\gamma x^2} \left(p + \frac{1}{x} \right) \left(x + \frac{1-\rho_1(2-\rho)}{\rho_2} \right) + 2\rho x \delta_1^2 \tag{5.68b}$$

The waiting times as a function of x and for different switchover time distributions can be seen in Figure 5.4. Due to the strong similarities the same conclusions can be made as for the first examples, with the exception that in this case the increase is not entirely linear in x for systems with a low traffic load.

5.6 Concluding Remarks

The performance of alternating-priority queues with very weak assumptions on the switchover time sequences was studied; all that is assumed is that these sequences are stationary ergodic. In spite of this generality explicit expressions were derived for the expected waiting times and number of customers in each queue. The expressions obtained involve the weighted sum of all correlations where the weights decrease exponentially fast to zero. With the help of these explicit expressions, through numerical studies and examples, it was shown that correlation can add up to 400% to the expected waiting times. This has important implications for (ad-hoc) networks where a common communication channel is shared amongst a number of users and the number of users between consecutive data transfers are correlated.

The analysis of the gated/gated queueing systems is slightly more evolved because the stochastic recursive equations for that system can not be written in a one-dimensional version. It remains a subject for future research.

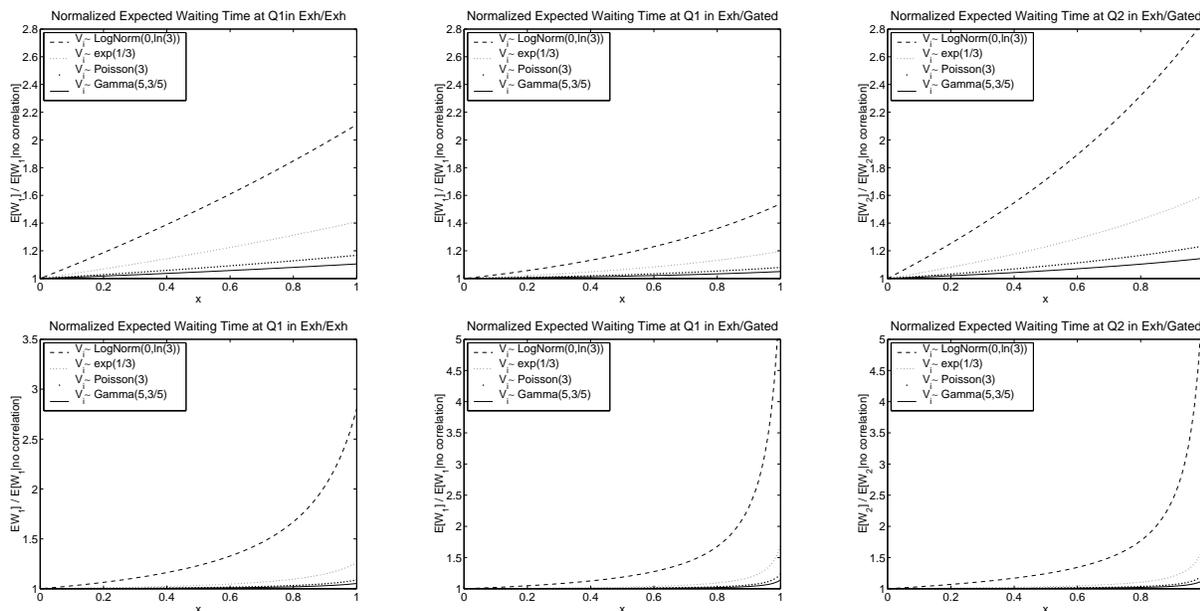


Figure 5.4: *Switchover Times from the Same Sequence.* The expected waiting time divided by the expected waiting time if there would be no correlation between the switchover times. The different lines correspond to different switchover time distributions. Here x determines the level of correlation, cross correlation is introduced by having the switchover times come from the same sequence, the service times are exponential, and $\lambda_i = 0.4$. The figures in the first row are with mean service time $b_i = 0.4$ ($\rho = 0.32$) and the figures on the bottom row are under heavy traffic with $b_i = 1.2$ ($\rho = 0.96$). The first column of figures correspond to the exhaustive/exhaustive system, whereas the second and third column of figures show the normalized waiting times for, respectively, the exhaustive queue (Q1) and the gated queue (Q2) in the exhaustive/gated system.

The extension of the models discussed here to more than two queues is another open problem that remains to be addressed. With independent switchover times the expected waiting times can be found by solving a set of equations. It will be interesting to see if a similar solution can be given for correlated switchover times and what the influence is of correlation on a larger number of queues.

Acknowledgements

I would like to thank Uri Yechiali for proofreading the original manuscript of this chapter as well as Daniele Miorandi for his discussions on polling systems.

5.A Proof of Theorem 5.3.2.

The proof of Theorem 5.3.2 will make use the following lemma [19, Propositions 6.6 and 6.31] [27, p.292 and p.295]

Lemma 5.A.1 *Let X_1, X_2, \dots be a stationary and ergodic process, and $\phi(\mathbf{x})$ be Borel measurable, then the process B_1, B_2, \dots defined by $B_n = \phi(X_n, X_{n+1}, \dots)$ is stationary ergodic.*

First of all note that $\mathcal{A}_n(y)$ and \mathcal{B}_n are nonnegative (component-wise) for all n and y . Moreover, $\mathcal{A}_n(y)$ are monotone increasing in y for all n and for each sample path. Theorem 1 in [3] then tells us that

- (i) if the sequence $\{(\mathcal{A}_n(\cdot), \mathcal{B}_n), -\infty < n < \infty\}$ is stationary ergodic⁹ and
- (ii) if $P(\overline{\lim}_{n \rightarrow \infty} I_{n,1} \text{ is finite}) > 0$ for $I_{0,1} = 0$,

then there exists a stationary ergodic regime $I_{n,1}^*$ defined on the probability space and $I_{n,1}^*$ satisfies (5.13).

We start by verifying that \mathcal{B}_n is stationary ergodic by analyzing it term by term. First of all the processes $V_{n,1}$ and $V_{n,2}$ are by assumption stationary ergodic. Furthermore it can be shown that $\mathcal{N}_{n,2}(V_{n,i})$, $i = 1, 2$, is also stationary ergodic. To do this introduce the sequence of i.i.d. uniform variables $U_{n,i}$ on $(0, 1)$ and let $\mathcal{N}_{n,2}(\omega)$, $\omega \in \mathbb{R}$, have cumulative distribution function $F_{\mathcal{N}_{n,2}(\omega)}(x) := P(\mathcal{N}_{n,2}(\omega) \leq x)$. With the help of the inverse-transformation method for drawing random numbers it follows that $\mathcal{N}_{n,2}(\omega) = F_{\mathcal{N}_{n,2}(\omega)}^{-1}(U_{n,i})$ which leads to $\mathcal{N}_{n,2}(V_{n,i}) = F_{\mathcal{N}_{n,2}(V_{n,i})}^{-1}(U_{n,i})$. In other words, $\mathcal{N}_{n,2}(V_{n,i})$ is a function of two parameters, $U_{n,i}$ and $V_{n,i}$. Call this function $f(U, V) := F_{\mathcal{N}_{n,2}(V)}^{-1}(U)$. This function is Borel measurable, and

⁹We mean by that notation that the sequence $\{((\mathcal{A}_n(y))_{y \in \mathcal{Y}}, \mathcal{B}_n), -\infty < n < \infty\}$ is stationary ergodic rather than $\{(\mathcal{A}_n(I_{n,1}), \mathcal{B}_n), -\infty < n < \infty\}$. Here \mathcal{Y} is a subset of \mathbb{R} .

since $X_n := (U_{n,i}, V_{n,i})$ is stationary ergodic it follows with Lemma 5.A.1 that $\mathcal{N}_{n,2}(V_{n,i}) = f(U_{n,i}, V_{n,i})$ is stationary ergodic. An identical argumentation with the sequence of i.i.d. uniform variables W_n on $(0, 1)$ leads to the function $g(W_n, N) := F_{\mathcal{D}_{n,2}(N)}^{-1}(W_n) = \mathcal{D}_{n,2}(N)$ with which it can be shown that $\mathcal{D}_{n,2}(N_n) = g(W_n, N_n)$ is also stationary ergodic. Getting back to \mathcal{B}_n , with the vector $X_n := (V_{n,1}, V_{n,2}, U_{n,1}, U_{n,2}, W_n)$ and the projection function $P_i(X_n)$ which gives the i -th coordinate of X_n , we can rewrite \mathcal{B}_n as

$$\begin{aligned}\mathcal{B}_n &= V_{n+1,1} + V_{n+1,2} + g(W_{n+1}, f(U_{n,2}, V_{n,2}) + f(U_{n+1,1}, V_{n+1,1})) \\ &= P_1(X_{n+1}) + P_2(X_{n+1}) + g(P_4(X_{n+1}), f(P_3(X_n), P_2(X_n)) + f(P_3(X_{n+1}), P_1(X_{n+1})))\end{aligned}$$

Since projections, additions and compositions of measurable functions are once again measurable it follows with Lemma 5.A.1 that \mathcal{B}_n is stationary ergodic.

Next we show that the sequence $\mathcal{A}_n(\cdot)$ is stationary ergodic by starting with the innermost term of (5.14) and working outwards. Since the number of arrivals in cycle n is independent of the number of arrivals in cycle k , for $k \neq n$, it follows that $\mathcal{N}_{n,1}(\cdot)$ is an i.i.d. sequence in n . With a similar argumentation it follows that $\mathcal{D}_{n,1}(\cdot)$ is an i.i.d. sequence in n . Furthermore, a composition of independent processes results once again in an i.i.d. sequence. To prove this let $\mathcal{D}_k(\cdot)$ and $\mathcal{N}_m(\cdot)$ be i.i.d. sequences and independent of each other. Conditioning on the inner process gives ($k \neq l$ and $m \neq n$)

$$\begin{aligned}P(\mathcal{D}_k(N_m) \leq x; \mathcal{D}_l(N_n) \leq y) &= \\ &= \int_0^\infty \int_0^\infty P(\mathcal{D}_k(r) \leq x; \mathcal{D}_l(s) \leq y) dP(N_m \leq r) dP(N_n \leq s) \\ &= \int_0^\infty \int_0^\infty P(\mathcal{D}_k(r) \leq x) P(\mathcal{D}_l(s) \leq y) dP(N_m \leq r) dP(N_n \leq s) \\ &= P(\mathcal{D}_k(N_m) \leq x) P(\mathcal{D}_l(N_n) \leq y).\end{aligned}$$

In particular, this means that the sequence $\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(\cdot))$ is an i.i.d. sequence in n . Performing two more compositions like this proves that $\mathcal{A}_n(\cdot)$ is an i.i.d. sequence (and therefore also stationary ergodic). This proves (i).

To show that (ii) holds it is sufficient to show that certain conditions of Lemma 1 in [3] hold as that lemma leads to (ii). The first condition is that $\mathcal{A}_n(\cdot)$ forms an i.i.d. sequence independent of the stationary ergodic sequence \mathcal{B}_n . The second condition is that $\mathbb{E}[|\mathcal{A}_0(y)|] \leq \alpha|y|$ for some $\alpha < 1$ and all $y \geq D$ where $D > 0$ is some constant.

It has already been shown that the sequence $\mathcal{A}_n(\cdot)$ forms an i.i.d. sequence and that \mathcal{B}_n is a stationary ergodic sequence. The sequence $\mathcal{A}_n(\cdot)$ is also independent of \mathcal{B}_n since (compare (5.15) to (5.14)) $V_{n+1,1}$ and $V_{n+1,2}$ are independent of $\mathcal{A}_n(\cdot)$, $\mathcal{N}_{n,2}$ is independent of $\mathcal{N}_{n+1,2}$, and $V_{n+1,1}$ is independent of $\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(\cdot))$.

Finally, since $\mathbb{E}[|\mathcal{A}_0(y)|] = \mathbb{E}[\mathcal{A}_n(y)] = \alpha y$, with $\alpha = \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$, it follows that the second condition holds since $\alpha < 1$ whenever $\rho < 1$, which was assumed throughout. Thus (ii) holds if $\rho < 1$.

Since (i) and (ii) hold there exists (Theorem 1 in [3]) a stationary regime $I_{n,1}^*$ which satisfies (5.13). \blacksquare

5.B Proof of Theorem 5.3.3.

Proof: Taking the expectation on both sides of equation (5.13) gives

$$\mathbb{E}[I_{n+1,1}] = R + \frac{\rho_2}{1 - \rho_2} \left(R + \frac{\rho_1 \mathbb{E}[I_{n,1}]}{1 - \rho_1} \right).$$

Under the stationary regime $\mathbb{E}[I_{n+1,1}] = \mathbb{E}[I_{n,1}]$ which immediately leads to the first moment

$$E[I_{n,1}] = \frac{R(1 - \rho_1)}{1 - \rho}. \quad (5.69)$$

To obtain the second moment we need to derive the terms on the right hand side of (5.16). First of all,

$$\begin{aligned} \mathbb{E}[\mathcal{A}_n^2(I_{n,1})] &= \mathbb{E} \left[\mathcal{D}_{n+1,2}^2 \left(\mathcal{N}_{n+1,2} \left(\mathcal{D}_{n+1,1} \left(\mathcal{N}_{n,1}(I_{n,1}) \right) \right) \right) \right] \\ &= \lambda_2^2 d_2^2 \mathbb{E}[\mathcal{D}_{n+1,1}^2(\mathcal{N}_{n,1}(I_{n,1}))] + \lambda_2 d_2^{(2)} \mathbb{E}[\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1}))] \\ &= \lambda_2^2 d_2^2 \left(\lambda_1^2 d_1^2 \mathbb{E}[I_{n,1}^2] + \lambda_1 d_1^{(2)} \mathbb{E}[I_{n,1}] \right) + \lambda_1 d_1 \lambda_2 d_2^{(2)} \mathbb{E}[I_{n,1}]. \end{aligned}$$

Plugging equations (5.7) and (5.17) into this results in

$$\mathbb{E}[\mathcal{A}_n^2(I_{n,1})] = \frac{\rho_1^2 \rho_2^2 \mathbb{E}[I_{n,1}^2]}{(1 - \rho_1)^2 (1 - \rho_2)^2} + \frac{R}{(1 - \rho_2)^2 (1 - \rho)} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1 - \rho_1)^2} + \frac{\lambda_2 \rho_1 b_2^{(2)}}{1 - \rho_2} \right). \quad (5.70)$$

Next we proceed with the second unknown of expression (5.16), $\mathbb{E}[\mathcal{B}_n^2]$, where we recall that \mathcal{B}_n is defined in equation (5.15). Making use of (5.3) gives

$$\begin{aligned} \mathbb{E}[\mathcal{B}_n^2] &= \mathbb{E} \left[V_{n+1,1} + V_{n+1,2} + \mathcal{D}_{n+1,2} \left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1}) \right) \right]^2 \\ &= v_1^{(2)} + v_2^{(2)} + 2v_1 v_2 + 2c_{12}(0) + 2\lambda_2 d_2 \left(v_1^{(2)} + v_2^2 + 2v_1 v_2 + c_2(1) + c_{12}(-1) + c_{12}(0) \right) \\ &\quad + \lambda_2^2 d_2^2 \left(v_1^{(2)} + v_2^{(2)} + 2v_1 v_2 + 2c_{12}(-1) \right) + \lambda_2 d_2^{(2)} R \\ &= \frac{\Delta^2 + R^2}{(1 - \rho_2)^2} - \frac{2\rho_2 \delta_2^2}{1 - \rho_2} + \frac{R \lambda_2 b_2^{(2)}}{(1 - \rho_2)^3} + \frac{2\rho_2 c_2(1) + 2c_{12}(0)}{1 - \rho_2} + \frac{2\rho_2 c_{12}(-1)}{(1 - \rho_2)^2}. \end{aligned} \quad (5.71)$$

To solve the last part first notice that the processes $\mathcal{N}_{n,1}(\cdot)$, $\mathcal{N}_{n,2}(\cdot)$, $\mathcal{N}_{n+1,1}(\cdot)$, $\mathcal{N}_{n+1,2}(\cdot)$, $\mathcal{D}_{n+1,1}(\cdot)$, and $\mathcal{D}_{n+1,2}(\cdot)$ are all independent of each other, and each of them is independent

of $I_{n,1}$, $V_{n,2}$, $V_{n+1,1}$, and $V_{n+1,2}$. This means that

$$\mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n] = \alpha \mathbb{E}[I_{n,1}\mathcal{B}_n], \quad (5.72)$$

with $\alpha := \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$. The last piece of the puzzle can be derived with the help of Theorem 2 in [3] which states that

$$I_{n,1} = \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} \mathcal{A}_i^{(n-j)} \right) (\mathcal{B}_{n-j-1}), \quad n \in \mathbb{Z},$$

where for each integer i , $\{\mathcal{A}_i^{(-j)}\}_j$ are independent of each other and have the same distribution as $\mathcal{A}_i(\cdot)$. To apply the theorem it is sufficient to have $\alpha < 1$, which turns out to be equivalent to $\rho < 1$, and that $\mathbb{E}[\mathcal{B}_n] < \infty$ (see Lemma 1 in [3]). The latter indeed holds as $\mathbb{E}[\mathcal{B}_n] = \frac{R}{1-\rho_2}$.

Applying the theorem gives

$$\begin{aligned} \mathbb{E}[I_{n,1}\mathcal{B}_n] &= \sum_{j=0}^{\infty} \mathbb{E} \left[\prod_{i=n-j}^{n-1} \left(\mathcal{D}_{i+1,2} \left(\mathcal{N}_{i+1,2} \left(\mathcal{D}_{i+1,1} \left(\mathcal{N}_{i,1}(\mathcal{B}_{n-j-1}) \right) \right) \right) \right) \cdot \mathcal{B}_n \right] \\ &= \sum_{j=0}^{\infty} \alpha^j \mathbb{E}[\mathcal{B}_{n-j-1}\mathcal{B}_n] = \sum_{j=0}^{\infty} \alpha^j \mathbb{E}[\mathcal{B}_0\mathcal{B}_{j+1}]. \end{aligned} \quad (5.73)$$

because of the independence of the processes $\mathcal{D}_{i,1}(\cdot)$, $\mathcal{D}_{n,2}(\cdot)$, $\mathcal{N}_{i,1}(\cdot)$, and $\mathcal{N}_{i,2}(\cdot)$, for all $i \in \mathbb{Z}$. Writing out the last term yields

$$\begin{aligned} \mathbb{E}[\mathcal{B}_0\mathcal{B}_{j+1}] &= \mathbb{E} \left[\left(V_{1,1} + V_{1,2} + \mathcal{D}_{1,2} \left(\mathcal{N}_{0,2}(V_{0,2}) + \mathcal{N}_{1,2}(V_{1,1}) \right) \right) \right. \\ &\quad \cdot \left. \left(V_{j+2,1} + V_{j+2,2} + \mathcal{D}_{j+2,2} \left(\mathcal{N}_{j+1,2}(V_{j+1,2}) + \mathcal{N}_{j+2,2}(V_{j+2,1}) \right) \right) \right] \\ &= v_1^2 + c_1(j+1) + v_1v_2 + c_{12}(j+1) + \lambda_2 d_2 \left(v_1v_2 + c_{12}(j) + v_1^2 + c_1(j+1) \right) \\ &\quad + v_1v_2 + c_{12}(-j-1) + v_2^2 + c_2(j+1) + \lambda_2 d_2 \left(v_2^2 + c_2(j) + v_1v_2 + c_{12}(-j-1) \right) \\ &\quad + \lambda_2 d_2 \left(v_1v_2 + c_{12}(-j-2) + v_2^2 + c_2(j+2) + \lambda_2 d_2 \left(v_2^2 + c_2(j+1) + v_1v_2 + c_{12}(-j-2) \right) \right) \\ &\quad + \lambda_2 d_2 \left(v_1^2 + c_1(j+1) + v_1v_2 + c_{12}(j+1) + \lambda_2 d_2 \left(v_1v_2 + c_{12}(j) + v_1^2 + c_1(j+1) \right) \right) \\ &= \frac{R^2}{(1-\rho_2)^2} + \frac{c_1(j+1) + c_2(j+1)}{(1-\rho_2)^2} + \frac{\rho_2(c_2(j) - 2c_2(j+1) + c_2(j+2))}{1-\rho_2} \\ &\quad + \frac{c_{12}(-j-1) + c_{12}(j+1)}{1-\rho_2} + \frac{\rho_2(c_{12}(-j-2) + c_{12}(j))}{(1-\rho_2)^2}. \end{aligned} \quad (5.74)$$

Putting (5.72)-(5.74) together and re-indexing the summation (for example, $\sum_{j=0}^{\infty} \alpha^{j+1} c_2(j) = \sum_{j=1}^{\infty} \alpha^j c_2(j-1) = \alpha c_2(0) + \sum_{j=1}^{\infty} \alpha^j \alpha c_2(j)$) produces

$$\begin{aligned} \mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n] &= \sum_{j=1}^{\infty} \left[\frac{R^2 + c_1(j) + c_2(j)}{(1-\rho_2)^2} + \frac{\rho_2 \left(\alpha - 2 + \frac{1}{\alpha} \right) c_2(j)}{1-\rho_2} \right. \\ &\quad \left. + \frac{c_{12}(-j)}{1-\rho_2} \left(1 + \frac{\rho_2}{\alpha(1-\rho_2)} \right) + \frac{c_{12}(j)}{1-\rho_2} \left(1 + \frac{\alpha\rho_2}{1-\rho_2} \right) \right] \alpha^j \\ &\quad + \frac{\rho_2(\alpha c_2(0) - c_2(1))}{1-\rho_2} + \frac{\rho_2(-c_{12}(-1) + \alpha c_{12}(0))}{(1-\rho_2)^2}. \end{aligned}$$

All of the terms with v_1 and v_2 can be pulled out of the summation and under stationary regime $c_2(0) = \delta_2^2$. This gives

$$\begin{aligned} \mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n] &= \frac{\rho_1 \rho_2 R^2}{(1-\rho_2)^2(1-\rho)} + \frac{\rho_1 \rho_2^2 \delta_2^2}{(1-\rho_1)(1-\rho_2)^2} - \frac{\rho_2 c_2(1)}{1-\rho_2} + \frac{\rho_2(-c_{12}(-1) + \alpha c_{12}(0))}{(1-\rho_2)^2} \\ &\quad + \frac{1}{(1-\rho_2)^2} \sum_{j=1}^{\infty} \left[c_1(j) + c_2(j) + \frac{(1-\rho)^2 c_2(j)}{\rho_1(1-\rho_1)} + \frac{(1-\rho_2)c_{12}(-j)}{\rho_1} \right] \alpha^j \\ &\quad + \frac{1-\rho_2(1-\alpha)}{(1-\rho_2)^2} \sum_{j=1}^{\infty} c_{12}(j) \alpha^j. \end{aligned} \tag{5.75}$$

Putting equations (5.70),(5.71), and (5.75) into (5.16) and collecting terms gives

$$\begin{aligned} \mathbb{E}[I_{n+1,1}^2] &= \frac{\rho_1^2 \rho_2^2 \mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2(1-\rho_2)^2} + \frac{R}{(1-\rho_2)^2(1-\rho)} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1-\rho_1)^2} + \lambda_2 b_2^{(2)} \right) \\ &\quad + \frac{1}{(1-\rho_2)^2} \left(\Delta^2 - \frac{2\rho_2(1-\rho)\delta_2^2}{1-\rho_1} + \left(\frac{1-\rho+2\rho_1\rho_2}{1-\rho} \right) R^2 \right) \\ &\quad + \frac{2(1-\rho_2(1-\alpha))}{(1-\rho_2)^2} \sum_{j=0}^{\infty} c_{12}(j) \alpha^j \\ &\quad + \frac{2}{(1-\rho_2)^2} \sum_{j=1}^{\infty} \left[c_1(j) + c_2(j) + \frac{(1-\rho)^2 c_2(j)}{\rho_1(1-\rho_1)} + \frac{(1-\rho_2)c_{12}(-j)}{\rho_1} \right] \alpha^j. \end{aligned}$$

Under stationary regime $\mathbb{E}[I_{n+1,1}^2] = \mathbb{E}[I_{n,1}^2]$. The theorem follows by putting these terms on the same side and by making use of the identity

$$1 - \left(\frac{\rho_1}{1-\rho_1} \right)^2 \left(\frac{\rho_2}{1-\rho_2} \right)^2 = \frac{(1-\rho)(1-\rho+2\rho_1\rho_2)}{(1-\rho_1)^2(1-\rho_2)^2}. \quad \blacksquare$$

5.C Alternative proof of Theorem 5.3.4.

Proof: Under a number of assumptions [33] presents a decomposition theory which states that the expected waiting time in an M/G/1 queue with vacations can be decomposed into two parts, namely

$$\mathbb{E}[W_{q,i}] = \frac{\lambda_i b_i^{(2)}}{2(1 - \rho_i)} + \mathbb{E}[\tilde{V}_i].$$

The first part is the Pollaczek-Khinchin formula for the expected waiting time in an ordinary M/G/1 queue without vacations. The second part is the forward-recurrence time defined as

$$\mathbb{E}[\tilde{V}_i] = \frac{\mathbb{E}[I_{n,i}^2]}{2\mathbb{E}[I_{n,i}]}.$$

See [25] for an application of this theory. Since a customer in a polling system arriving at queue i sees the system as an M/G/1 with vacations, and the theorem does not assume independence between vacations and service times (or busy periods), we can apply the theorem to our polling system. \blacksquare

5.D Proof of Theorem 5.4.3.

Proof: First of all note that the expected service time of N customers at the second queue is the sum of their individual service times and hence $\mathbb{E}[S_{n,2}] = b_2 \mathbb{E}[N]$. Taking the expectation over (5.34) then gives

$$\begin{aligned} \mathbb{E}[S_{n+1,2}] &= \rho_2 \left(\mathbb{E}[S_{n,2}] + R + \mathbb{E}[\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2} + V_{n,1} + V_{n,2}))] \right) \\ &= \frac{\rho_2}{1 - \rho_1} \left(\mathbb{E}[S_{n,2}] + R \right), \end{aligned}$$

where the last line follows since $\mathbb{E}[D_{n,1}] = \rho_1 \mathbb{E}[I_{n,1}] / (1 - \rho_1)$ (equation (5.12a)). Because of the stationarity ($\mathbb{E}[S_{n+1,2}] = \mathbb{E}[S_{n,2}]$), this gives the first moment, $\mathbb{E}[S_{n+1,2}] = \rho_2 R / (1 - \rho)$.

Before we derive the second moment, we first need to point out (with the help of (5.4) and (5.6)) that

$$\begin{aligned} \mathbb{E}[S_{n,2}^2(\mathcal{N}_{n,2}(T))] &= b_2^2 \mathbb{E}[\mathcal{N}_{n,2}^2(T)] + (b_2^{(2)} - b_2^2) \mathbb{E}[\mathcal{N}_{n,2}(T)] \\ &= \rho_2^2 \mathbb{E}[T^2] + \lambda_2 b_2^{(2)} \mathbb{E}[T]. \end{aligned} \tag{5.76}$$

Squaring (5.34) gives the second moment of the service time of the second queue as

$$\mathbb{E}[S_{n+1,2}^2] = \mathbb{E}[\mathcal{X}_n^2(S_{n,2})] + \mathbb{E}[\mathcal{Y}_n^2] + 2\mathbb{E}[\mathcal{X}_n(S_{n,2})\mathcal{Y}_n]. \tag{5.77}$$

The right hand side will be solved piece by piece. First of all, using (5.76), and then (5.6) along with the independence of $\mathcal{N}_{n,2}(\cdot)$ and $\mathcal{N}_{n+1,2}(\cdot)$ gives

$$\begin{aligned}
 \mathbb{E}[\mathcal{X}_n^2(S_{n,2})] &= \mathbb{E} \left[\mathcal{S}_{n+1,2}^2 \left(\mathcal{N}_{n,2}(S_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))) \right) \right] \\
 &= b_2^2 \mathbb{E} \left[\left(\mathcal{N}_{n,2}(S_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))) \right)^2 \right] \\
 &\quad + (b_2^{(2)} - b_2^2) \mathbb{E} \left[\mathcal{N}_{n,2}(S_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))) \right] \\
 &= \lambda_2 b_2^2 \left(\lambda_2 \mathbb{E}[S_{n,2}^2] + \mathbb{E}[S_{n,2}] + 2\lambda_2 \mathbb{E} [S_{n,2} \cdot \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))] \right) \\
 &\quad + \lambda_2 \mathbb{E} [\mathcal{D}_{n+1,1}^2(\mathcal{N}_{n,1}(S_{n,2}))] + \mathbb{E} [\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))] \\
 &\quad + \lambda_2 (b_2^{(2)} - b_2^2) \mathbb{E} [S_{n,2} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))].
 \end{aligned}$$

A couple of these terms can be crossed out to give

$$\begin{aligned}
 \mathbb{E}[\mathcal{X}_n^2(S_{n,2})] &= \rho_2^2 \left(\mathbb{E}[S_{n,2}^2] + 2\mathbb{E} [S_{n,2} \cdot \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))] + \mathbb{E} [\mathcal{D}_{n+1,1}^2(\mathcal{N}_{n,1}(S_{n,2}))] \right) \\
 &\quad + \lambda_2 b_2^{(2)} \mathbb{E} [S_{n,2} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))].
 \end{aligned}$$

Applying formula (5.8) and then substituting $d_1 = b_1/(1 - \rho_1)$ and $d_1^{(2)} = b_1^{(2)}/(1 - \rho_1)^3$ (see (5.7)) leads to

$$\begin{aligned}
 \mathbb{E}[\mathcal{X}_n^2(S_{n,2})] &= \rho_2^2 \left(\mathbb{E}[S_{n,2}^2] \left(1 + \frac{2\rho_1}{1 - \rho_1} + \lambda_1^2 d_1^2 \right) + \lambda_1 d_1^{(2)} \mathbb{E}[S_{n,2}] \right) + \frac{\lambda_2 b_2^{(2)}}{1 - \rho_1} \mathbb{E}[S_{n,2}] \\
 &= \frac{\rho_2^2 \mathbb{E}[S_{n,2}^2]}{(1 - \rho_1)^2} + \frac{\mathbb{E}[S_{n,2}]}{1 - \rho_1} \left(\frac{\lambda_1 b_1^{(2)} \rho_2^2}{(1 - \rho_1)^2} + \lambda_2 b_2^{(2)} \right) \\
 &= \frac{\rho_2^2 \mathbb{E}[S_{n,2}^2]}{(1 - \rho_1)^2} + \frac{\rho_2 R}{(1 - \rho_1)(1 - \rho)} \left(\frac{\lambda_1 b_1^{(2)} \rho_2^2}{(1 - \rho_1)^2} + \lambda_2 b_2^{(2)} \right) \tag{5.78}
 \end{aligned}$$

The second part of (5.77) can be derived along similar lines. Taking the expectation of the

square of (5.36),

$$\begin{aligned}
\mathbb{E}[\mathcal{Y}_n^2] &= \mathbb{E}\left[\mathcal{S}_{n+1,2}^2\left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}\left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)\right)\right] \\
&= b_2^2 \mathbb{E}\left[\left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}\left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)\right)^2\right] \\
&\quad + \left(b_2^{(2)} - b_2^2\right) \mathbb{E}\left[\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}\left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)\right] \\
&= \lambda_2 b_2^2 \left(\lambda_2 \mathbb{E}[V_{n,2}^2] + \mathbb{E}[V_{n,2}] + 2\lambda_2 \mathbb{E}\left[V_{n,2} \cdot \left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)\right]\right) \\
&\quad + \lambda_2 \mathbb{E}\left[\left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)^2\right] \\
&\quad + \mathbb{E}\left[\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right] \\
&\quad + \lambda_2 \left(b_2^{(2)} - b_2^2\right) \mathbb{E}\left[V_{n,2} + \mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right].
\end{aligned}$$

Also here terms can be crossed out¹⁰. Doing this gives

$$\begin{aligned}
\mathbb{E}[\mathcal{Y}_n^2] &= \rho_2^2 \left(\mathbb{E}[V_{n,2}^2] + 2\mathbb{E}\left[V_{n,2} \cdot \left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)\right]\right) \\
&\quad + \mathbb{E}\left[\left(\mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right)^2\right] \\
&\quad + \lambda_2 b_2^{(2)} \mathbb{E}\left[V_{n,2} + \mathcal{D}_{n+1,1}\left(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})\right) + V_{n+1,1}\right],
\end{aligned}$$

after which the expectations can be moved farther back in the equation to give

$$\begin{aligned}
\mathbb{E}[\mathcal{Y}_n^2] &= \rho_2^2 \left(\mathbb{E}[V_{n,2}^2] + \frac{2\rho_1}{1 - \rho_1} \mathbb{E}[V_{n,1}V_{n,2} + V_{n,2}^2] + 2\mathbb{E}[V_{n+1,1}V_{n,2}]\right) \\
&\quad + \lambda_1^2 d_1^2 \mathbb{E}[(V_{n,1} + V_{n,2})^2] + \lambda_1 d_1^{(2)} \mathbb{E}[V_{n,1} + V_{n,2}] + \mathbb{E}[V_{n+1,1}^2] \\
&\quad + \frac{2\rho_1}{1 - \rho_1} \mathbb{E}[V_{n,1}V_{n+1,1} + V_{n+1,1}V_{n,2}] + \frac{\lambda_2 b_2^{(2)} R}{1 - \rho_1}.
\end{aligned}$$

¹⁰The slightly more elaborate calculations and then the crossing out of terms is due to $N_{n,2}(\cdot)$ and $N_{n+1,2}(\cdot)$ being treated separately. Although it is tempting to combine them into one term, it is better not to do so since the independence and correlations between the different processes $\mathcal{N}_{n,1}(T)$, $\mathcal{N}_{n,2}(T)$, $\mathcal{D}_{n+1,2}(T)$, $\mathcal{S}_{n,1}(T)$, and $\mathcal{S}_{n+1,1}(T)$ are not always obvious and have to be treated with care when squaring them.

In terms of the covariance functions $c_1(j)$, $c_2(j)$, and $c_{12}(\pm j)$, $j \in \mathbb{N}$ (see equation (5.3)),

$$\begin{aligned}
 \mathbb{E}[\mathcal{Y}_n^2] &= \rho_2^2 \left(v_2^{(2)} + \frac{2\rho_1(v_1v_2 + c_{12}(0) + v_2^{(2)})}{1 - \rho_1} + 2v_1v_2 + 2c_{12}(-1) \right. \\
 &\quad \left. + \frac{\rho_1^2(v_1^{(2)} + v_2^{(2)} + 2v_1v_2 + 2c_{12}(0))}{(1 - \rho_1)^2} + \frac{\lambda_1 b_1^{(2)} R}{(1 - \rho_1)^3} + v_1^{(2)} \right. \\
 &\quad \left. + \frac{2\rho_1(v_1^2 + c_1(1) + v_1v_2 + c_{12}(-1))}{1 - \rho_1} \right) + \frac{\lambda_2 b_2^{(2)} R}{1 - \rho_1} \\
 &= \frac{\rho_2^2}{1 - \rho_1} \left(\frac{\delta_1^2 + \delta_2^2 + R^2}{1 - \rho_1} - 2\rho_1 \delta_1^2 + \frac{\lambda_1 b_1^{(2)} R}{(1 - \rho_1)^2} \right) + \frac{\lambda_2 b_2^{(2)} R}{1 - \rho_1} \\
 &\quad + \frac{2\rho_2^2}{1 - \rho_1} \left(\rho_1 c_1(1) + c_{12}(-1) + \frac{\rho_1 c_{12}(0)}{1 - \rho_1} \right). \tag{5.79}
 \end{aligned}$$

The third part of (5.77) is

$$\begin{aligned}
 \mathbb{E}[\mathcal{X}_n(S_{n,2})\mathcal{Y}_n] &= \mathbb{E} \left[\mathcal{S}_{n+1,2} \left(\mathcal{N}_{n,2}(S_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))) \right) \cdot \mathcal{Y}_n \right] \\
 &= b_2 \mathbb{E} \left[\left(\mathcal{N}_{n,2}(S_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2}))) \right) \cdot \mathcal{Y}_n \right] \\
 &= \rho_2 \mathbb{E} \left[\left(S_{n,2} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(S_{n,2})) \right) \cdot \mathcal{Y}_n \right] = \gamma \mathbb{E}[S_{n,2}\mathcal{Y}_n] \tag{5.80}
 \end{aligned}$$

where in each step the independence and non-overlapping of $S_{n,2}$ and $V_{n,i}$ ($i = 1, 2$) was used. Theorem 2 in [3] provides us with a mean to work out (5.80). The theorem states that

$$S_{n,2} = \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} \mathcal{X}_i^{(n-j)} \right) (\mathcal{Y}_{n-j-1}), \quad n \in \mathbb{Z},$$

where for each integer i , $\{\mathcal{X}_i^{(-j)}\}_j$ are independent of each other and have the same distribution as $\mathcal{X}_i(\cdot)$. To apply the theorem it is sufficient to have $\gamma < 1$, which turns out to be equivalent to $\rho < 1$, and that $\mathbb{E}[\mathcal{Y}_n] < \infty$ (see Lemma 1 in [3]). The latter indeed holds as $\mathbb{E}[\mathcal{Y}_n] = \frac{\rho_2 R}{1 - \rho_1}$.

Using the theorem and because of the independence of the processes $\mathcal{D}_{n,1}(\cdot)$, $\mathcal{S}_{n,2}(\cdot)$,

$\mathcal{N}_{n,1}(\cdot)$, and $\mathcal{N}_{n,2}(\cdot)$, for all $n \in \mathbb{N}$, we obtain

$$\begin{aligned}
\mathbb{E}[S_{n,2}\mathcal{Y}_n] &= \sum_{j=0}^{\infty} \mathbb{E} \left[\prod_{i=n-j}^{n-1} \left(\mathcal{S}_{i+1,2} \left(\mathcal{N}_{i,2}(\mathcal{Y}_{n-j-1}) + \mathcal{N}_{i+1,2}(\mathcal{D}_{i+1,1}(\mathcal{N}_{i,1}(\mathcal{Y}_{n-j-1}))) \right) \right) \cdot \mathcal{Y}_n \right] \\
&= \sum_{j=0}^{\infty} b_2^j \mathbb{E} \left[\prod_{i=n-j}^{n-1} \left(\mathcal{N}_{i,2}(\mathcal{Y}_{n-j-1}) + \mathcal{N}_{i+1,2}(\mathcal{D}_{i+1,1}(\mathcal{N}_{i,1}(\mathcal{Y}_{n-j-1}))) \right) \cdot \mathcal{Y}_n \right] \\
&= \sum_{j=0}^{\infty} \rho_2^j \mathbb{E} \left[\prod_{i=n-j}^{n-1} \left(\mathcal{Y}_{n-j-1} + \mathcal{D}_{i+1,1}(\mathcal{N}_{i,1}(\mathcal{Y}_{n-j-1})) \right) \cdot \mathcal{Y}_n \right] \\
&= \sum_{j=0}^{\infty} \left(\frac{\rho_2}{1-\rho_1} \right)^j \mathbb{E}[\mathcal{Y}_{n-j-1}\mathcal{Y}_n] = \sum_{j=0}^{\infty} \gamma^j \mathbb{E}[\mathcal{Y}_0\mathcal{Y}_{j+1}]. \tag{5.81}
\end{aligned}$$

with $\gamma := \frac{\rho_2}{1-\rho_1}$. Finally, writing out the last term of this expression yields

$$\begin{aligned}
\mathbb{E}[\mathcal{Y}_0\mathcal{Y}_{j+1}] &= \mathbb{E} \left[\mathcal{S}_{1,2} \left(\mathcal{N}_{0,2}(V_{0,2}) + \mathcal{N}_{1,2}(\mathcal{D}_{1,1}(\mathcal{N}_{0,1}(V_{0,1} + V_{0,2})) + V_{1,1}) \right) \right. \\
&\quad \left. \cdot \mathcal{S}_{j+2,2} \left(\mathcal{N}_{j+1,2}(V_{j+1,2}) + \mathcal{N}_{j+2,2}(\mathcal{D}_{j+2,1}(\mathcal{N}_{j+1,1}(V_{j+1,1} + V_{j+1,2})) + V_{j+2,1}) \right) \right] \\
&= \rho_2^2 \mathbb{E} \left[\left(V_{0,2} + \mathcal{D}_{1,1}(\mathcal{N}_{0,1}(V_{0,1} + V_{0,2})) + V_{1,1} \right) \right. \\
&\quad \left. \cdot \left(V_{j+1,2} + \mathcal{D}_{j+2,1}(\mathcal{N}_{j+1,1}(V_{j+1,1} + V_{j+1,2})) + V_{j+2,1} \right) \right] \\
&= \rho_2^2 \mathbb{E} \left[V_{1,1}V_{j+2,1} + V_{0,2}V_{j+1,2} + V_{1,1}V_{j+1,2} + V_{j+2,1}V_{0,2} \right. \\
&\quad \left. + \frac{\rho_1}{1-\rho_1} \left(V_{1,1}V_{j+1,1} + V_{0,1}V_{j+2,1} + 2V_{0,2}V_{j+1,2} \right. \right. \\
&\quad \left. \left. + V_{j+1,1}V_{0,2} + V_{0,1}V_{j+1,2} + V_{j+2,1}V_{0,2} + V_{1,1}V_{j+1,2} \right) \right. \\
&\quad \left. + \frac{\rho_1^2}{(1-\rho_1)^2} \left(V_{0,1}V_{j+1,1} + V_{0,2}V_{j+1,2} + V_{0,1}V_{j+1,2} + V_{j+1,1}V_{0,2} \right) \right].
\end{aligned}$$

Using the covariance functions (5.3)

$$\begin{aligned}
 \mathbb{E}[\mathcal{Y}_0\mathcal{Y}_{j+1}] &= \rho_2^2 \left[v_1^2 + c_1(j+1) + v_2^2 + c_2(j+1) + 2v_1v_2 + c_{12}(j) + c_{12}(-j-2) \right. \\
 &\quad + \frac{\rho_1}{1-\rho_1} \left(2v_1^2 + c_1(j) + c_1(j+2) + 2v_2^2 + 2c_2(j+1) \right. \\
 &\quad \quad \left. \left. + 4v_1v_2 + c_{12}(-j-1) + c_{12}(j+1) + c_{12}(-j-2) + c_{12}(j) \right) \right. \\
 &\quad \left. + \frac{\rho_1^2}{(1-\rho_1)^2} \left(v_1^2 + c_1(j+1) + v_2^2 + c_2(j+1) + 2v_1v_2 + c_{12}(j+1) + c_{12}(-j-1) \right) \right] \\
 &= \frac{\rho_2^2}{1-\rho_1} \left[\frac{R^2}{1-\rho_1} + \frac{c_1(j+1) + c_2(j+1)}{1-\rho_1} + \rho_1(c_1(j) - 2c_1(j+1) + c_1(j+2)) \right. \\
 &\quad \left. + c_{12}(-j-2) + c_{12}(j) + \frac{\rho_1(c_{12}(-j-1) + c_{12}(j+1))}{1-\rho_1} \right]. \tag{5.82}
 \end{aligned}$$

Putting equations (5.81) and (5.82) into (5.80) and re-indexing the summation gives

$$\begin{aligned}
 \mathbb{E}[\mathcal{X}_n(S_{n,2})\mathcal{Y}_n] &= \frac{\rho_2^2}{1-\rho_1} \sum_{j=1}^{\infty} \left(\frac{R^2 + c_1(j) + c_2(j)}{1-\rho_1} + \rho_1(c_1(j-1) - 2c_1(j) + c_1(j+1)) \right. \\
 &\quad \left. + c_{12}(-j-1) + c_{12}(j-1) + \frac{\rho_1(c_{12}(-j) + c_{12}(j))}{1-\rho_1} \right) \gamma^j.
 \end{aligned}$$

In this expression the term R^2 can be pulled out of the summation and the various covariance functions can be collected together (for example, $\sum_{j=1}^{\infty} \gamma^j c_2(j-1) = \gamma c_1(0) + \sum_{j=1}^{\infty} \gamma^j \gamma c_2(j)$) to reveal

$$\begin{aligned}
 \mathbb{E}[\mathcal{X}_n(S_{n,2})\mathcal{Y}_n] &= \frac{\rho_2^2}{1-\rho_1} \left[\frac{\rho_2 R^2}{(1-\rho_1)(1-\rho)} + \rho_1(\gamma c_1(0) - c_1(1)) - c_{12}(-1) + \gamma c_{12}(0) \right. \\
 &\quad \left. + \sum_{j=1}^{\infty} \gamma^j \left(\frac{c_1(j) + c_2(j)}{1-\rho_1} + \rho_1 c_1(j) \left(\gamma - 2 + \frac{1}{\gamma} \right) + c_{12}(-j) \left(\frac{1}{\gamma} + \frac{\rho_1}{1-\rho_1} \right) + c_{12}(j) \left(\gamma + \frac{\rho_1}{1-\rho_1} \right) \right) \right] \\
 &= \frac{\rho_2^2}{1-\rho_1} \left[\frac{\rho_2 R^2}{(1-\rho_1)(1-\rho)} + \rho_1(\gamma c_1(0) - c_1(1)) - c_{12}(-1) + \gamma c_{12}(0) \right] \\
 &\quad + \frac{\rho_2^2}{(1-\rho_1)^2} \sum_{j=1}^{\infty} \left(c_1(j) + c_2(j) + \frac{\rho_1}{\rho_2} (1-\rho)^2 c_1(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} c_{12}(-j) + \rho c_{12}(j) \right) \gamma^j. \tag{5.83}
 \end{aligned}$$

Putting equations (5.78), (5.79), and (5.83) into (5.77), using $c_1(0) = v_1^{(2)} - v_1^2$, and collect-

ing terms gives

$$\begin{aligned}
\mathbb{E}[S_{n+1,2}^2] &= \frac{\rho_2^2 \mathbb{E}[S_{n,2}^2]}{(1-\rho_1)^2} + \frac{\rho_2 R}{(1-\rho_1)(1-\rho)} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1-\rho_1)^2} + \lambda_2 b_2^{(2)} \right) \\
&+ \frac{\rho_2^2}{1-\rho_1} \left(\frac{\delta_1^2 + \delta_2^2 + R^2}{1-\rho_1} - 2\rho_1 \delta_1^2 + \frac{\lambda_1 b_1^{(2)} R}{(1-\rho_1)^2} \right) + \frac{\lambda_2 b_2^{(2)} R}{1-\rho_1} \\
&+ \frac{2\rho_2^2}{1-\rho_1} \left(\rho_1 c_1(1) + c_{12}(-1) + \frac{\rho_1 c_{12}(0)}{1-\rho_1} \right) \\
&+ \frac{2\rho_2^2}{1-\rho_1} \left[\frac{\rho_2 R^2}{(1-\rho_1)(1-\rho)} + \rho_1 (\gamma \delta_1^2 - c_1(1)) - c_{12}(-1) + \gamma c_{12}(0) \right] \\
&+ \frac{2\rho_2^2}{(1-\rho_1)^2} \sum_{j=1}^{\infty} \gamma^j \left(c_2(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) + \rho c_{12}(j) \right) \\
&= \frac{\rho_2^2 \mathbb{E}[S_{n,2}^2]}{(1-\rho_1)^2} + \frac{R}{1-\rho} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1-\rho_1)^2} + \lambda_2 b_2^{(2)} \right) \\
&+ \frac{\rho_2^2}{(1-\rho_1)^2} \left((1-2\rho_1(1-\rho)) \delta_1^2 + \delta_2^2 + \left(\frac{1-\rho_1+\rho_2}{1-\rho} \right) R^2 + 2\rho c_{12}(0) \right) \\
&+ \frac{2\rho_2^2}{(1-\rho_1)^2} \sum_{j=1}^{\infty} \left(c_2(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) + \rho c_{12}(j) \right) \gamma^j.
\end{aligned}$$

The final expression is obtained by assuming stationary, putting the terms $\mathbb{E}[S_{n+1,2}^2] = \mathbb{E}[S_{n,2}^2]$ on the same side, and multiplying both sides by $(1-\rho_1)^2/\rho_2^2$. ■

5.E List of Notation

$\mathcal{A}_n(\cdot)$ = A nested combination of stochastic processes defined in (5.14).

\mathcal{B}_n = A nested combination of stochastic processes defined in (5.15).

$\mathcal{D}_{n,i}(N)$ = Total busy period generated by N customers in queue i with arrival rate λ_i and first and second moment of the service time b_i and $b_i^{(2)}$, respectively.

$\mathcal{D}_{n,i,k}$ = Single busy period generated by the k -th customer in queue i with arrival rate λ_i and first and second moment of the service time b_i and $b_i^{(2)}$, respectively.

$\mathcal{S}_{n,i}(N)$ = Service time of N customers at queue i with first and second moment of the service time b_i and $b_i^{(2)}$, respectively.

$\mathcal{N}_{n,i}(T)$ = Number of arrivals at queue i in time T in the n^{th} cycle (cycle starting from the polling instant of the first queue).

$\dot{\mathcal{N}}_{n,2}(\cdot)$ = The number of arrivals at the second queue during the n^{th} cycle, with the cycle starting from the arrival of the server at the second queue.

$\mathcal{T}_{n,i}(N)$ = The number of customers served at queue i during the n^{th} cycle if there are N customers in the queue at the moment of polling.

$\mathcal{X}_n(\cdot)$ = A nested combination of stochastic processes defined in (5.35).

\mathcal{Y}_n = A nested combination of stochastic processes defined in (5.36).

$\alpha = \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$ = Central quantity in the exhaustive/exhaustive queueing system.

Comes forth from $\mathbb{E}[\mathcal{A}_n(I)] = \alpha \mathbb{E}[I]$, with $\mathcal{A}_n(\cdot)$ defined in (5.14).

$b_i = \mathbb{E}[B_i]$ = Expected service time at queue i .

$b_i^{(2)} = \mathbb{E}[B_i]^2$ = Second moment of the service time at queue i .

$c_i(n) = \mathbb{E}[V_{0,i}V_{n,i}] - \mathbb{E}[V_{0,i}]\mathbb{E}[V_{n,i}]$ = Covariance function for the vacation sequences at queue i .

$c_{12}(n) = \mathbb{E}[V_{0,1}V_{n,2}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,2}]$ = Covariance function for the vacation sequences between the two queues.

$d_i = \rho_i/(1-\rho_i)$ = The expected duration of a single busy period, where the arrival rate is λ_i and the average service time is b_i .

$d_i^{(2)} = b_i^{(2)}/(1-\rho_i)^3$ = Second moment of the duration of a single busy period, where the arrival rate is λ_i , the average service time is b_i , and the second moment of the service time is $b_i^{(2)}$.

$\delta_i^2 = v_i^{(2)} - v_i^2$ = Variance of the switching time from queue i to the other queue.

$\Delta^2 = \delta_1^2 + \delta_2^2$ = Sum of the variances of the switching times.

$\gamma = \frac{\rho_2}{1-\rho_1}$ = Central quantity in the exhaustive/gated queueing system. Comes

forth from $\mathbb{E}[\mathcal{X}_n(I)] = \gamma \mathbb{E}[I]$, with $\mathcal{X}_n(\cdot)$ defined in (5.36).

λ_i = (Poisson) arrival rate at queue i .

$\rho_i = \lambda_i d_i$ = load at queue i .

$\rho = \rho_1 + \rho_2$ = load of the system. The assumption is made throughout that $\rho < 1/2$.

$\hat{\rho} = \rho_1 = \rho_2$ if the parameter settings for the two queues are equal.

$R = v_1 + v_2$.

$v_i = \mathbb{E}[V_{n,i}]$ = Expected switching time from queue i to the other queue.

$v_i^{(2)} = \mathbb{E}[V_{n,i}^2]$ = Second moment of the switching time away from queue i .

- $C_{n,i}$ = Duration of the n^{th} cycle starting from the polling instant of the i^{th} queue.
- $D_{n,i}$ = Duration of the busy period at queue i in the n^{th} cycle.
- $I_{n,i}$ = Inter-visit of the i^{th} queue in the n^{th} cycle. This is the time between the server switching away from queue i until the time that the server comes back to queue i . ($I_{n,1} = V_{n,1} + D_{n,2} + V_{n,2}$ for exhaustive/exhaustive queues and $I_{n,1} = V_{n,1} + S_{n,2} + V_{n,2}$ for exhaustive/gated queues).
- K = A constant used in the expressions for the inter-visit time and the waiting times. Sometimes K_1 or K_2 is used if there is a difference between the two queues.
- $L_{n,i}^*$ = Number of customers in queue i in the n^{th} cycle at the moment the queue is polled.
- $L_{q,i}$ = Average number of customers in queue i . This is also the number of customers that arrived at queue i during a vacation, and are still in the queue, before a tag customer arrived in that same vacation.
- $L_{s,i}$ = Average number of customers at queue i (including the customer in service). This is also the number of customers that arrived at queue i during a vacation (and are in the queue or in service) before a tag customer arrived in that same vacation.
- $S_{n,i}$ = Service time at queue i in the n^{th} cycle (=similar to the duration $D_{n,i}$ of the busy period but then with no arrivals).
- $T_{n,i}$ = The number of customers served, per cycle, at queue i .
- $V_{n,i}$ = Switching time from queue i to the other queue in the n^{th} cycle.
- $W_{q,i}$ = Random variable for the waiting time of a customer in queue i (not including service).
- $W_{s,i}$ = Random variable for the total sojourn time of a customer at queue i (waiting time plus service time).
- Γ_i = The number of customers served during a busy period, where the arrival rate is λ_i , average service time is b_i , and the second moment of the service time is $b_i^{(2)}$.

Appendix

Appendix **A**

The Value Function of a Tandem Queue

Contents

A.1 Introduction	162
A.2 A tandem queue with no arrivals	162
A.3 A tandem queue with arrivals	168
A.4 Appendix: Proof of Proposition A.2.1	170
A.5 Appendix: Proof of Proposition A.2.2	172

We study the tandem queue with independent servers with exponentially distributed service times. The system is subject to holding costs. An explicit expression is derived for the value function of the average costs in the absence of customer arrivals. The derivation is accomplished through the inversion of its generating function. The resulting expression is a weighted combination of Catalan and ballot numbers which give an intuitive explanation to the problem at hand. The techniques and tricks used—along with the insights gained—can provide valuable information for the derivation of the value function in the presence of arrivals. Once derived, the value function can be used for important optimization and control of many queueing networks.

Note: The work in this chapter was sponsored by a van Gogh project. It was done jointly with Sandjai Bhulai at the Vrije Universiteit in Amsterdam and is a continuation of [15, section 4.3].

A.1 Introduction

The optimal control of servers and the routing in queueing networks (such as tandem queues, networks of queues, re-entry queues, or packets being routed through the Internet) is important from both a financial and a congestion perspective. Obtaining the optimal control or service policy can be achieved through linear programming, value and policy iteration [67], or the power series algorithm [50]. However, the optimal policy can often not be determined explicitly due to the curse of dimensionality. It has been shown [16] [80] that the first step of policy iteration produces policies which are very close to the optimal policy, while in some cases not having problems with the curse of dimensionality. To perform this one-step policy iteration the value function needs to be known.

The value function, also known as the bias vector, of a queueing system provides us with information about the cost of operating the system. The costs can be either averaged over time [42]—the situation which will be considered here—or they can be discounted [49] [59] [60]. For further information on the value function the reader is referred to [15, Chapter 4] or [58].

The structure of this chapter is as follows. In Section A.2 the value function of the tandem queue is discussed, in the absence of customer arrivals. Its generating function is given in Section A.2.1 after which the value function is obtained in Section A.2.2. In Section A.3 the tandem queue with arrivals is discussed.

A.2 A tandem queue with no arrivals

Consider two single server queues in tandem as shown in Figure A.1. Arrivals at the first are determined by a Poisson process with rate λ . Once a customer has been served at the first queue (s)he proceeds to the second queue. When the second server has finished serving the customer, the customer leaves the system. The duration of the service in the first (respectively second) queue is exponentially distributed with parameter μ_1 (resp. μ_2).

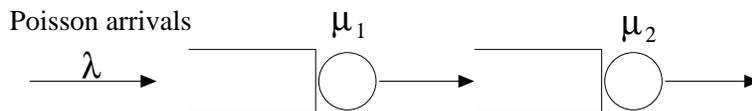


Figure A.1: A tandem queue with arrival rate λ and service rates μ_1 and μ_2 .

The system is exposed to holding costs h_i for each customer in queue i for each one unit of time. Without restrictions we will take $h_1 = 0$ and $h_2 = 1$. Holding costs could also be included in the first queue, but because this queue behaves as an ordinary M/M/1 queue—for which the value function and the average holding costs are known [15, Sections 3.4 and 4.3]—we take $h_1 = 0$.

The second queue also behaves as an M/M/1 queue with arrival rate λ and service rate

μ_2 , but *only* from a long-run average point of view. Since the value function captures the transient behaviour of the system the results from the M/M/1 queue cannot be used.

To understand the transient behaviour of how the system is emptied we start by looking at the system with no arrivals, i.e., setting $\lambda = 0$. If there are no arrivals, then the long-run average (holding) cost is equal to zero.

The dynamic programming optimality equation (or Poisson equation) for the system is given by¹

$$(\mu_1 + \mu_2)V(x, y) = y + \mu_1 V(x - 1, y + 1) \cdot 1_{(x>0)} + \mu_1 V(x, y) \cdot 1_{(x=0)} + \mu_2 V(x, [y - 1]^+), \quad (\text{A.1})$$

The function $V(x, y)$ is the (relative) value function and it is our goal to find an expression for this quantity. The Poisson equation is derived as follows. If the system is in state (x, y) , for $x > 0$ and $y > 0$ (this corresponds to x , respectively, y customers in the first, respectively, the second queue), then the expected time until the system changes states is $1/(\mu_1 + \mu_2)$. The holding costs occurred during this amount of time is y times this quantity, which gives the first term in the right hand side of (A.1), after multiplication by $\mu_1 + \mu_2$. From here the system can go to either state $(x - 1, y + 1)$ with probability $\mu_1/(\mu_1 + \mu_2)$, or to state $(x, y - 1)$ with probability $\mu_2/(\mu_1 + \mu_2)$. The cases where $x = 0$ and $y = 0$ are treated in a similar way.

The function $V(x, y)$ has the interpretation of the asymptotic difference in total costs that results from starting the process with x , respectively, y customers in the first, respectively, second queue as compared to starting from some reference state. Without loss of generality we will take the reference state to be the empty tandem queue ($x = 0$ and $y = 0$), giving $V(0, 0) = 0$.

A.2.1 The generating function of the value function

The expression for the value function will be found with the help of its generating function. For this purpose define $G(z_1, z_2) := \sum_{x \geq 0, y \geq 0} V(x, y) z_1^x z_2^y$. With equation (A.1) this gives

$$\left[\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2 \right] G(z_1, z_2) = \frac{z_2}{(1 - z_1)(1 - z_2)^2} + \mu_1 G(0, z_2) + \left[\mu_2 - \mu_1 \frac{z_1}{z_2} \right] G(z_1, 0). \quad (\text{A.2})$$

We proceed by first obtaining an expression for $G(z_1, z_2)$, through the derivation of $G(z_1, 0)$ and $G(0, z_2)$, after which $G(z_1, z_2)$ will be inverted to give us the value function.

By taking a simple M/M/1 queue, or $z_1 = 0$ in equation (A.2), we have from (A.1)

¹The notation $[\cdot]^+$ is used to indicate $[y]^+ = \max(0, y)$.

$$V(0, y) = \frac{y(y+1)}{2\mu_2}, \quad (\text{A.3})$$

and thus

$$G(0, z_2) = \frac{z_2}{\mu_2(1-z_2)^3}. \quad (\text{A.4})$$

To obtain $G(z_1, 0)$, we can choose z_2 such that the left hand side of equation (A.2) vanishes (i.e., choose z_2 such that $\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2 = 0$), giving us an expression for $G(z_1, 0)$ in terms of $G(0, z_2)$. Some calculus shows that z_2 defined by

$$z_2 = \frac{\mu_1 + \mu_2 - \sqrt{(\mu_1 + \mu_2)^2 - 4\mu_1\mu_2}}{2\mu_2} = \mu_1 \left(\frac{\mu_1 + \mu_2}{\mu_1\mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma}}{2} \right)$$

is the correct root, with $\gamma := \mu_1\mu_2/(\mu_1 + \mu_2)^2$.

Denote $S(z_1) := \sqrt{(\mu_1 + \mu_2)^2 - 4\mu_1\mu_2 z_1}$. By inserting the above value of z_2 into (A.2) and performing some calculus we find

$$\begin{aligned} G(z_1, 0) &= \frac{2(\mu_1 + \mu_2 - S(z_1))^2(\mu_1 + \mu_2 + S(z_1) - 2\mu_1 z_1)}{(1-z_1)(\mu_2 - \mu_1 + S(z_1))^3(-\mu_1 - \mu_2 + S(z_1) + 2\mu_1 z_1)} \\ &= \frac{z_1(\mu_1 - \mu_2 + S(z_1))}{2\mu_1\mu_2(1-z_1)^3} \\ &= \frac{z_1}{\mu_2(1-z_1)^3} - \frac{1}{\mu_1 + \mu_2} \left(\frac{z_1^2}{(1-z_1)^3} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2\gamma z_1} \right). \end{aligned} \quad (\text{A.5})$$

Inverting this expression gives

$$V(x, 0) = V(0, x) - \frac{1}{\mu_1 + \mu_2} \sum_{k=0}^{x-2} \binom{x-k}{2} \underbrace{\frac{1}{k+1} \binom{2k}{k}}_{\text{Catalan numbers}} \gamma^k. \quad (\text{A.6})$$

The Catalan numbers which show up in the last summation have numerous interpretations [74, pages 221-247], of which one is very closely related to the tandem queue.

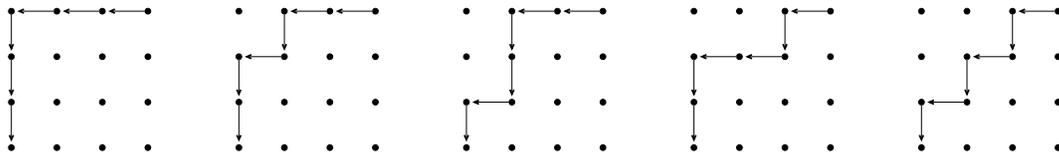


Figure A.2: The paths from $(3, 3)$ to $(0, 0)$, without going below the diagonal. The number of paths is given by the Catalan number $C_N = \frac{1}{N+1} \binom{2N}{N}$ with $N = 3$ ($C_3 = 5$).

Interpretation of Catalan numbers. Consider an N -by- N lattice. The Catalan numbers $C_N := \frac{1}{N+1} \binom{2N}{N}$ describe the number of lattice paths (also referred to as Dyck paths) from the point (N, N) to $(0, 0)$, without going below the diagonal line (x, x) . An example for $N = 3$ is given in Figure A.2.

Next take a look at the possible state transitions shown in Figure A.3 for the tandem queue with $x = 3$ customers in the first queue and $y = 0$ customers in the second queue. It should be clear that with a simple shift Figures A.2 and A.3 describe the same situation.

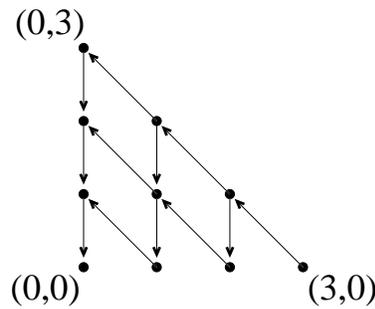


Figure A.3: The possible state transitions from state $(3, 0)$ to state $(0, 0)$.

With the above interpretation it is not surprising that (a weighted sum of) the Catalan numbers show up in the expression for $V(x, 0)$, c.f., equation (A.6).

Plugging (A.4) and (A.5) into (A.2) gives the first part of the following proposition.

Proposition A.2.1 (Generating function of the value function for $\lambda = 0$)

The generating function of the bias vector for the average cost of a tandem queue, with holding costs in the second queue and with no arrivals ($\lambda = 0$), is given by

$$\begin{aligned} \left[\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2 \right] G(z_1, z_2) &= \frac{z_2}{(1 - z_1)(1 - z_2)^2} + \frac{\mu_1}{\mu_2} \left(\frac{z_2}{(1 - z_2)^3} \right) \quad (\text{A.7a}) \\ &+ \left(\frac{z_1}{(1 - z_1)^3} \right) \left(\mu_2 - \mu_1 \frac{z_1}{z_2} \right) \left(\frac{1}{\mu_2} - \left(\frac{\mu_1 + \mu_2}{\mu_1 \mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right) \right), \end{aligned}$$

where $\gamma := \mu_1 \mu_2 / (\mu_1 + \mu_2)^2$.

Alternatively, $G(z_1, z_2)$ can be rewritten as

$$\begin{aligned}
G(z_1, z_2) = & \frac{1}{\mu_2} \left(\frac{z_2}{(1-z_1)(1-z_2)^3} + \frac{z_1 z_2}{(1-z_1)^2(1-z_2)^2} + \frac{z_1}{(1-z_1)^3(1-z_2)} \right) \\
& - \left(\frac{z_1^2}{(1-z_1)^3(1-z_2)} \right) \left(\frac{1}{\mu_1 + \mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2\gamma z_1} \right) \\
& - \frac{1}{\mu_2} \left(\frac{z_1}{(1-z_1)^2(1-z_2)} \right) \left(\frac{z_2}{S_1(z_1) - z_2} \right), \tag{A.7b}
\end{aligned}$$

where

$$S_1(z_1) := \left(\frac{\mu_1 + \mu_2}{\mu_2} \right) \left(\frac{1 + \sqrt{1 - 4\gamma z_1}}{2} \right). \quad \blacksquare$$

The proof of (A.7b) is forwarded to Appendix A.4 for the sake of readability.

By writing $G(z_1, z_2)$ in the second form it can be inverted to give $V(x, y)$, even though, to the best of the author's knowledge, it not known how to invert the kernel $(\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2)^{-1}$. Inversion of equation (A.7b) provides us with the main result of this chapter.

A.2.2 The value function of a tandem queue

Proposition A.2.2 (Value function of the tandem queue for $\lambda = 0$)

The value function for the average cost of a tandem queue with holding costs in the second queue and no arrivals ($\lambda = 0$) is given by

$$\begin{aligned}
V(x, y) = & \frac{(x+y)(x+y+1)}{2\mu_2} - \frac{1}{\mu_1 + \mu_2} \sum_{k=0}^{x-2} \binom{x-k}{2} \overbrace{\frac{1}{k+1} \binom{2k}{k}}^{\text{Catalan numbers}} \gamma^k \\
& - \frac{1}{\mu_2} \sum_{0 \leq k \leq x-1} \sum_{1 \leq m \leq y} (x-k) \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^m \underbrace{\frac{m}{k+m} \binom{2k+m-1}{k+m-1}}_{\text{Ballot numbers}} \gamma^k, \tag{A.8a}
\end{aligned}$$

where $\gamma := \mu_1 \mu_2 / (\mu_1 + \mu_2)^2$.

The value function can also be written as

$$\begin{aligned}
V(x, y) = & V(x, 0) + V(0, y) + \frac{xy}{\mu_2} \\
& - \frac{1}{\mu_2} \sum_{0 \leq k \leq x-1} \sum_{1 \leq m \leq y} (x-k) \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^m \frac{m}{k+m} \binom{2k+m-1}{k+m-1} \gamma^k, \tag{A.8b}
\end{aligned}$$

where

$$V(x, 0) = \frac{x(x+1)}{2\mu_2} - \frac{1}{\mu_1 + \mu_2} \sum_{k=0}^{x-2} \binom{x-k}{2} \frac{1}{k+1} \binom{2k}{k} \gamma^k,$$

$$V(0, y) = \frac{y(y+1)}{2\mu_2}$$

are the boundary value functions. ◇

The proof is given in Appendix A.5.

The ballot numbers are related to the classical *ballot problem*² and can be used to describe the number of paths in a N -by- $(N+m)$ lattice going to from $(0, 0)$ to $(N, N+m)$, while never going below the diagonal (x, x) . For $m = 1$ the ballot numbers turn into the Catalan numbers.

Corollary A.2.1 (Special cases)

Some special cases of the value function are given by the following situations.

- The servers at both queues work at the same speed ($\mu_1 = \mu_2 = \mu$):

$$\begin{aligned} V(x, y) &= \frac{(x+y)(x+y+1)}{2\mu} - \frac{1}{2\mu} \sum_{k=0}^{x-2} \binom{x-k}{2} \frac{1}{k+1} \binom{2k}{k} \left(\frac{1}{4}\right)^k \\ &\quad - \frac{1}{\mu} \sum_{0 \leq k \leq x-1} \sum_{1 \leq m \leq y} (x-k) \left(\frac{1}{2}\right)^m \frac{m}{k+m} \binom{2k+m-1}{k+m-1} \left(\frac{1}{4}\right)^k \\ &= \frac{2xy + y(y+1)}{2\mu} + \left(\frac{1}{2}\right)^y \sum_{0 \leq k \leq x-1} (x-k) \binom{2k+y}{k+y} \left(\frac{1}{4}\right)^k. \end{aligned}$$

- The server at the second queue works at half the speed of the server at the first queue ($\mu_1 = \frac{1}{2}\mu_2$):

$$V(x, y) = V(x, 0) + 2V(0, y) + \frac{2}{3\mu_2} \sum_{k=0}^{x-2} \binom{x-k}{2} \frac{1}{k+1} \binom{2k}{k} \left(\frac{2}{9}\right)^k.$$

²Suppose that, in an election, candidate 0 receives $N+k$ votes and candidate 1 receives N votes. What is the probability that candidate 0 is always in the lead during the counting process? A related question is in how many ways can the ballots be counted such that candidate 0 is always in the lead during the counting process. For detailed discussions on the ballot problem the reader is referred to [24] and [31].

- *The server at the second queue works at twice the speed of the server at the first queue ($\mu_1 = 2\mu_2$):*

$$\begin{aligned}
V(x, y) = & \frac{(x+y)(x+y+1)}{2\mu_2} - \frac{1}{3\mu_2} \sum_{k=0}^{x-2} \binom{x-k}{2} \frac{1}{k+1} \binom{2k}{k} \left(\frac{2}{9}\right)^k \\
& - \frac{1}{\mu_2} \left(\frac{y}{2^y} + (x-2) \left(1 - (1/2)^y\right) \right). \quad \diamond
\end{aligned}$$

A.3 A tandem queue with arrivals

In the previous sections the value function was derived for a tandem queue with no arrivals ($\lambda = 0$). Unfortunately, the extension of these results to value function in the presence of arrivals ($\lambda > 0$) remains an open problem. The problem lies in finding an expression for $G(0, z_2)$ or for $G(z_1, 0)$. Once one of these two expressions is known, the other can be derived from this and we obtain an expression for $G(z_1, z_2)$. The final step would then be to invert $G(z_1, z_2)$.

For completeness we give the Poisson equations for $\lambda \geq 0$ and show the relationship between $G(z_1, z_2)$, $G(0, z_2)$, and $G(z_1, 0)$.

Let $\lambda \geq 0$ and let g denote the average costs. Define $\rho_i = \lambda/\mu_i$ for $i = 1, 2$ and assume that the stability conditions $\rho_1 < 1$ and $\rho_2 < 1$ hold. The dynamic programming equation is given by

$$\begin{aligned}
g + (\lambda + \mu_1 + \mu_2)V(x, y) = & y + \mu_1 V(x-1, y+1) \cdot 1_{(x>0)} + \mu_1 V(x, y) \cdot 1_{(x=0)} \\
& + \mu_2 V(x, [y-1]^+) + \lambda V(x+1, y).
\end{aligned}$$

It is known that the average cost $g = \lambda/(\mu_2 - \lambda)$. From this we then have $V(1, 0) = g/\lambda = 1/(\mu_2 - \lambda)$.

The Poisson equations can also be written as

$$\begin{aligned}
g + \lambda V(0, 0) &= \lambda V(1, 0), \\
g + (\lambda + \mu_1)V(x, 0) &= \lambda V(x+1, 0) + \mu_1 V(x-1, 1), & x > 0, \\
g + (\lambda + \mu_2)V(0, y) &= y + \lambda V(1, y) + \mu_2 V(0, y-1), & y > 0, \\
g + (\lambda + \mu_1 + \mu_2)V(x, y) &= y + \lambda V(x+1, y) + \mu_1 V(x-1, y+1) \\
&+ \mu_2 V(x, y-1), & x > 0, y > 0.
\end{aligned}$$

By taking the z -transform we obtain

$$g + \left[\lambda + \mu_1 + \mu_2 - \frac{\lambda}{z_1} - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2 \right] G(z_1, z_2) = \frac{z_2}{(1 - z_1)(1 - z_2)^2} + \left[\mu_1 - \frac{\lambda}{z_1} \right] G(0, z_2) + \left[\mu_2 - \mu_1 \frac{z_1}{z_2} \right] G(z_1, 0).$$

From here on the problem lies in finding an expression for $G(0, z_2)$ or for $G(z_1, 0)$. Once one of these two expressions is known, the other can be derived from this and an expression for $G(z_1, z_2)$ is obtained.

A.4 Appendix: Proof of Proposition A.2.1

Proof: The starting point is the generating function given in equation (A.7a). This generating function has an—let us call it—interesting kernel which is difficult to invert. To invert the generating function we shall deploy two simple tricks: (1) subtract and add terms (which sum up to zero) so as to obtain three separate terms for which it is known how to invert them, and (2) during the derivation of the third term an expression will appear in both the numerator and in the denominator—therefore it can be crossed out—which renders the terms benign (for the purpose of inversion).

Bringing the kernel to the other side gives

$$\begin{aligned}
 G(z_1, z_2) = & \left(\frac{z_2}{(1-z_1)(1-z_2)^2} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 & + \frac{\mu_1}{\mu_2} \left(\frac{z_2}{(1-z_2)^3} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 & + \left(\frac{z_1}{(1-z_1)^3} \right) \left(\frac{\mu_2 - \mu_1 \frac{z_1}{z_2}}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \left(\frac{1}{\mu_2} - \left(\frac{\mu_1 + \mu_2}{\mu_1 \mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right) \right).
 \end{aligned} \tag{A.9}$$

By using properties similar to $\frac{\mu_2 - \mu_1 \frac{z_1}{z_2}}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} = 1 - \frac{\mu_1 - \mu_2 z_2}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2}$, and by rearranging terms we obtain

$$\begin{aligned}
 G(z_1, z_2) = & \left(\frac{z_2}{(1-z_1)(1-z_2)^2} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 & - \left(\frac{z_2}{(1-z_1)(1-z_2)^2} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 & + \frac{1}{\mu_2} \left(\frac{z_2}{(1-z_1)(1-z_2)^3} + \frac{z_1 z_2}{(1-z_1)^2 (1-z_2)^2} \right) \\
 & + \frac{1}{\mu_2} \left(\frac{z_1}{(1-z_1)^2 (1-z_2)} \right) \left(\frac{\mu_1 - \mu_2 z_2}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 & + \frac{1}{\mu_2} \left(\frac{z_1}{(1-z_1)^3 (1-z_2)} \right) \left(1 - \frac{\mu_1 (1-z_1)}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 & - \left(\frac{z_1}{(1-z_1)^3 (1-z_2)} \right) \left(1 - \frac{\mu_1 (1-z_1)}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \left(\frac{\mu_1 + \mu_2}{\mu_1 \mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right).
 \end{aligned} \tag{A.10}$$

Here we see that the first and second term of (A.9) will cancel each other out. If in addition

we subtract a z_2 term from the last term in the last line of (A.9), then we obtain

$$\begin{aligned}
 G(z_1, z_2) &= \frac{1}{\mu_2} \left(\frac{z_2}{(1-z_1)(1-z_2)^3} + \frac{z_1 z_2}{(1-z_1)^2(1-z_2)^2} + \frac{z_1}{(1-z_1)^3(1-z_2)} \right) \\
 &\quad - \left(\frac{z_1 z_2}{(1-z_1)^2(1-z_2)} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \\
 &\quad - \left(\frac{z_1}{(1-z_1)^3(1-z_2)} \right) \left(\frac{\mu_1 + \mu_2}{\mu_1 \mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right) \\
 &\quad + \left(\frac{z_1}{(1-z_1)^2(1-z_2)} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right) \left[\left(\frac{\mu_1 + \mu_2}{\mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right) - z_2 \right] \\
 &\quad + \left(\frac{z_1 z_2}{(1-z_1)^2(1-z_2)} \right) \left(\frac{1}{\mu_1 + \mu_2 - \mu_1 \frac{z_1}{z_2} - \mu_2 z_2} \right).
 \end{aligned}$$

Notice that the second term and the fifth term cancel each other out. Because of the recurrence of a couple of terms it is worthwhile to introduce

$$\begin{aligned}
 S_1(z_1) &:= \left(\frac{\mu_1 + \mu_2}{\mu_2} \right) \left(\frac{1 + \sqrt{1 - 4\gamma z_1}}{2} \right) \\
 S_2(z_1) &:= \left(\frac{\mu_1 + \mu_2}{\mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right)
 \end{aligned}$$

This gives us

$$\begin{aligned}
 G(z_1, z_2) &= \frac{1}{\mu_2} \left(\frac{z_2}{(1-z_1)(1-z_2)^3} + \frac{z_1 z_2}{(1-z_1)^2(1-z_2)^2} + \frac{z_1}{(1-z_1)^3(1-z_2)} \right) \\
 &\quad - \left(\frac{z_1}{(1-z_1)^3(1-z_2)} \right) \left(\frac{\mu_1 + \mu_2}{\mu_1 \mu_2} \right) \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2} \right) \\
 &\quad - \frac{1}{\mu_2} \left(\frac{z_1 z_2}{(1-z_1)^2(1-z_2)} \right) \left(\frac{S_2(z_1) - z_2}{(S_2(z_1) - z_2)(S_1(z_1) - z_2)} \right)
 \end{aligned}$$

Now we are finally in a position to see why the rearranging of the terms pays off: in the last the term $S_2(z_1) - z_2$ shows up in both the numerator and in the denominator. Since $S_2(z_1) - z_2 \neq 0$ these terms vanish and we obtain equation (A.7b). \blacksquare

A.5 Appendix: Proof of Proposition A.2.2

Proof: To invert the generating function $G(z_1, z_2)$ given in equation (A.7b) we will make use of the following two lemmas.

Lemma A.5.1 *We have*

$$\frac{z_2}{S_1(z_1) - z_2} = \sum_{x \geq 0} \sum_{y \geq 1} \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^y \frac{y}{x + y} \binom{2x + y - 1}{x + y - 1} \gamma^x z_1^x z_2^y \quad (\text{A.11})$$

◇

Proof: We can rewrite $\frac{z_2}{S_1(z_1) - z_2}$ as

$$\begin{aligned} \frac{z_2}{S_1(z_1) - z_2} &= \sum_{y \geq 1} \left(\frac{1}{S_1(z_1)} \right)^y z_2^y \\ &= \sum_{y \geq 1} \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^y \left(\frac{1 - \sqrt{1 - 4\gamma z_1}}{2\gamma z_1} \right)^y z_2^y \end{aligned}$$

This is done by making use of the geometric serie $\sum_{n \geq 1} x^n = x/(1 - x)$ and then by filling in the value of $S_1(z_1)$. The geometric expansion is allowed as long as $|x| < 1$, which in our case is equivalent to saying that z_1 and z_2 need to so such that $\frac{z_2}{S_1(z_1) - z_2} < 1$. As $S_1(z_1) > \frac{1}{2}$ this is not a problem.

Equation (A.11) is derived from this by making use of the following formula which is obtained from the book Analysis of Algorithms [70, page 128]

$$\left(\frac{1 - \sqrt{1 - 4z}}{2} \right)^y = \sum_{x \geq y} \frac{y}{x} \binom{2x - y - 1}{x - 1} z^x.$$

Dividing this last expression by z^y and taking $z := \gamma z_1$ gives the formulated result. This series is known as the *Ballot* numbers of which the Catalan numbers are a special case. ■

Lemma A.5.2 *The convolution of two generating functions A and B defined by $A(z_1, z_2) = \sum_{x \geq 0} \sum_{y \geq 0} a(x, y) z_1^x z_2^y$ and $B(z_1, z_2) = \sum_{x \geq 0} \sum_{y \geq 1} b(x, y) z_1^x z_2^y$ is given by*

$$A(z_1, z_2) * B(z_1, z_2) = \sum_{x \geq 0} \sum_{y \geq 0} \sum_{0 \leq k \leq x} \sum_{1 \leq m \leq y} a(x - k, y - m) b(k, m) z_1^x z_2^y \quad \diamond$$

Proof: The lemma follows by writing out the convolution:

$$\begin{aligned}
 A(z_1, z_2) * B(z_1, z_2) &= \left(\sum_{x \geq 0} \underbrace{\sum_{y \geq 1} a(x, y) z_2^y z_1^x}_{a'(x)} \right) \left(\sum_{x \geq 0} \underbrace{\sum_{y \geq 0} b(x, y) z_2^y z_1^x}_{b'(x)} \right) \\
 &= \sum_{x \geq 0} \left(\sum_{0 \leq k \leq x} \overbrace{\sum_{y \geq 0} a(x-k, y) z_2^y}^{a'(x-k)} \overbrace{\sum_{y \geq 1} b(k, y) z_2^y}^{b'(k)} \right) z_1^x \\
 &= \sum_{x \geq 0} \left(\sum_{0 \leq k \leq x} \sum_{y \geq 0} \left(\sum_{1 \leq m \leq y} a(x-k, y-m) b(k, m) \right) z_2^y \right) z_1^x \\
 &= \sum_{x \geq 0} \sum_{y \geq 0} \sum_{0 \leq k \leq x} \sum_{1 \leq m \leq y} a(x-k, y-m) b(k, m) z_1^x z_2^y \quad \blacksquare
 \end{aligned}$$

With these two lemmas the inversion of $G(z_1, z_2)$ does not require much more work.

The first line of equation (A.7b) is a combination of known generating functions which are easy to invert. The inversion of the remaining two lines follows from the two lemmas. ■

Bibliography

- [1] <http://www.webster.com>.
- [2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance in complex networks. *Nature*, 406:378–382, 2000.
- [3] Eitan Altman. Stochastic recursive equations with applications to queues with dependent vacations. *Annals of Operations Research*, 112(1):43–61, 2002.
- [4] Eitan Altman. On stochastic recursive equations and infinite server queues. In *Proc. of IEEE Infocom 2005*, Miami, FL, March 13-17 2005.
- [5] Eitan Altman, Chadi Barakat, and Konstantin Avratchenkov. A stochastic model of TCP/IP with stationary ergodic random losses. In *ACM-Sigcomm*, Aug. 28 - Sept. 1 2000. See also INRIA Research Report RR-3824.
- [6] Eitan Altman, Tania Jimenez, and Daniel Kofman. DPS queues with stationary ergodic service times and the performance of TCP in overload. In *Proceedings of IEEE Infocom*, Hong-Kong, March 2004.
- [7] Eitan Altman, Panagiotis Konstantopoulos, and Zhen Liu. Stability, monotonicity and invariant quantities in general polling systems. *Queueing Systems*, 11:35–57, 1992. Special issue on 'Polling Systems'.
- [8] N. Bansal and Z. Liu. Capacity, delay and mobility in wireless ad-hoc networks. In *Proc. of IEEE Infocom Conf.*, San Francisco, CA, April 1-3 2003.
- [9] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [10] Albert-László Barabási, Réka Albert, Hawoong Jeong, and G. Bianconi. Power-law distribution of the world wide web. *Science*, 287(2115a), 2000.
- [11] Christian Bettstetter. Mobility modeling in wireless networks: Categorization, smooth movement, and border effects. *ACM Mobile Computing and Communications Review*, 5(3):55–67, July 2001.

- [12] Christian Bettstetter, Hannes Hartenstein, and Xavier Pérez-Costa. Stochastic properties of the random waypoint mobility model. *ACM/Kluwer Wireless Networks: Special Issue on Modelling and Analysis of Wireless Networks*, 10(5):555–567, 2004.
- [13] Christian Bettstetter, Giovanni Resta, and Paolo Santi. The node distribution of the random waypoint mobility model for wireless ad hoc networks. *IEEE Transactions on Mobile Computing*, 2(3), July-September 2003.
- [14] Christian Bettstetter and C. Wagner. The spatial node distribution of the random waypoint mobility model. *Proceedings First German Workshop Mobile Ad-Hoc Networks (WMAN)*, pages 41–58, 2002.
- [15] Sandjai Bhulai. *Markov Decision Processes: the Control of High-dimensional Systems*. PhD thesis, Vrije Universiteit, Amsterdam, the Netherlands, 2002.
- [16] Sandjai Bhulai and Ger Koole. On the structure of value functions for threshold policies in queueing models. *Journal of Applied Probability*, 40:613–622, 2003.
- [17] Thomas Bonald, M. Jonckheere, and A. Proutière. Insensitive load balancing. In *ACM Sigmetrics Performance Evaluation Review*, volume 32, pages 367–377, June 2004.
- [18] Thomas Bonald, A. Proutière, G. Régnié, and J.W. Roberts. Insensitivity results in statistical bandwidth sharing. In *ITC*, number 17, Salvador, Brazil, 2001.
- [19] Leo Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 1992.
- [20] B. Brewin. Delta has success in rfid baggage tag test. *Computerworld*, December 18 2003.
- [21] William L. Briggs and Van Emden Henson. *The DFT. An Owner's Manual for the Discrete Fourier Transform*. SIAM. Society for Industrial & Applied Mathematics, March 1995.
- [22] Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proceedings of ACM International Conference on Mobile Computing and Networks (ACM MOBICOM)*, pages 85–97, Dallas, TX, October 25-30 1998.
- [23] Florence Clévenot, Philippe Nain, and Keith Ross. Stochastic fluid model for cache clusters. *Performance Evaluation*, 59(1):1–18, January 2005.
- [24] Louis Comtet. *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. D. Reidel, Dordrecht, the Netherland, 1974.
- [25] Robert B. Cooper, Shun-Chen Niu, and Mandyam M Srinivasan. A decomposition theorem for polling models: The switchover times are effectively additive. *Operations Researchs*, 44(4):629–633, Jul-Aug 1996.

-
- [26] D. Cox and H. Miller. *The Theory of Stochastic Processes*. Chapman & Hall, London, UK, 1965.
- [27] Richard Durrett. *Probability: Theory and Examples*. Wadsworth & Brooks/Cole Statistics/Probability Series. Brooks/Cole, 1991.
- [28] Martin Eisenberg. Queues with periodic service and changeover time. *Operations Research*, 20(2):440–451, March-April 1972.
- [29] B.J. Feder. Delta to invest in radio tags for luggage at airports. *The New York Times*, July 1 2004.
- [30] B.J. Feder. Psst. This is your sensor. Your grapes are thirsty. *The New York Times*, July 26 2004.
- [31] William Feller. *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, Inc., 1950.
- [32] Richard Frenkiel, B.R. Badrinath, Joan Borrás, and Roy Yates. The infostations challenge: Balancing cost and ubiquity in delivering wireless data. *IEEE Personal Communications*, 7(2):66–71, April 2000.
- [33] S.W. Fuhrmann and Robert B. Cooper. Stochastic decompositions in the M/G/1 queue with generalized vacations. *Operations Research*, 33(5):1117–1129, 1985.
- [34] Abbas El Gamal, James Mammen, Balaji Prabhakar, and Devavrat Shah. Throughput-delay trade-off in wireless networks. In *Proc. of IEEE Infocom Conf.*, Hong Kong, March 2004.
- [35] C.W. Gardiner. *Handbook of Stochastic Processes*. Springer-Verlag, New York, 1985.
- [36] Zihui Ge, Daniel Figueiredo, Sharad Jaiswal, James F. Kurose, and Don Towsley. Modeling peer-peer file sharing systems. In *Proceedings of IEEE Infocom*, pages 1312–1321, San Francisco, April 1-3 2003.
- [37] I.S. Gradshteyn and I.M. Ryzhik. *Tables of Integrals, Series, and Products*. Academic Press, Inc. (London) Ltd., fourth edition, 1983.
- [38] H. Green. Sensor revolution: Bugging the world; soon, sensor networks will track everything from weather to inventory. *Business Week*, August 25 2003.
- [39] Robin Groenevelt and Eitan Altman. Analysis of alternating-priority queueing models with (cross) correlated switchover times. Technical Report RR-5368, INRIA, Sophia-Antipolis, November 2004.
- [40] Robin Groenevelt and Eitan Altman. Analysis of alternating-priority queueing models with (cross) correlated switchover times. In *Proc. of IEEE Infocom 2005*, Miami, FL, March 13-17 2005.

- [41] Robin Groenevelt, Eitan Altman, and Philippe Nain. Relaying in mobile ad hoc networks. In *Proc. of Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'04)*, UK, March 2004.
- [42] Robin Groenevelt, Ger Koole, and Philippe Nain. On the bias vector of a two-class preemptive priority queue. *Mathematical Methods of Operations Research*, 55(1):107–120, 2002.
- [43] Robin Groenevelt, Philippe Nain, and Ger Koole. Message delay in mobile ad hoc networks. Technical Report RR-5372, INRIA Sophia Antipolis, November 2004.
- [44] Robin Groenevelt, Philippe Nain, and Ger Koole. Message delay in mobile ad hoc networks. In *To appear as an extended abstract in the proceedings of ACM Sigmetrics*, June 6-10 2005. Longer version is available at [43].
- [45] Matthias Grossglauser and David Tse. Mobility increases the capacity of ad-hoc wireless networks. *ACM/IEEE Transactions in Networking*, 10(4):477–486, August 2002. This paper has also been published in the Proceedings of IEEE INFOCOM 2001 where it received the best paper award.
- [46] R.A. Guérin. Channel occupancy time distribution in a cellular radio system. *IEEE Transactions on Vehicular Technology*, 36:89–99, August 1987.
- [47] P. Gupta and P.R. Kumar. The capacity of wireless networks. *IEEE Trans. Information Theory*, 46(2):388–404, March 2000.
- [48] Y. Han, R.J. La, A.M. Makowski, and S. Lee. Distribution of path durations in mobile ad-hoc networks – palm’s theorem at work. Submitted to Special Issue of Computer Networks Journal on Network Modeling and Simulation, 2004.
- [49] J.M. Harrison. A priority queue with discounted linear costs. *Operations Research*, 23:260–269, 1975.
- [50] Gerard Hooghiemstra and Ger Koole. On the convergence of the power series algorithm. *Performance Evaluation*, 42:21–39, 2000.
- [51] Esa Hyytiä, Pasi Lassila, Laura Nieminen, and Jorma Virtamo. Spatial node distribution in the random waypoint mobility model. Submitted for publication, April 2004.
- [52] Esa Hyytiä and Jorma Virtamo. Random waypoint model in n -dimensional space. 2005.
- [53] K. Itô. On stochastic differential equations. *Memoirs, American Mathematical Society*, 4:1–51, 1951.
- [54] Alain Jean-Marie and Philippe Robert. On the transient behaviour of the processor sharing queue. *QUESTA*, 17(1-2):129–136, 1994.

- [55] David B Johnson and David A Maltz. Dynamic source routing in ad hoc wireless networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers, 1996.
- [56] D.B. Johnson and D.A. Maltz. Dynamic source routing in ad hoc wireless networks. *Mobile Computing*, 353:151–181, 1996. chapter 5.
- [57] Leonard Kleinrock. *Queueing Systems, Volume I: Theory*, volume I. John Wiley and Sons, 1976.
- [58] Ger Koole. Convexity in tandem queues. *Probability in the Engineering and Informational Sciences*, 18:13–31, 2004.
- [59] Ger Koole and Philippe Nain. On the value function of a priority queue with an application to a controlled polling model. *Queueing Systems*, 34:199–214, 2000.
- [60] Ger Koole and Philippe Nain. An explicit solution for the value function of a priority queue. *Queueing Systems*, 47:251–282, 2004.
- [61] Q. Li and D. Rus. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *Proc. 6th ACM MobiCom*, Boston, MA, August 6-11 2000.
- [62] J.D.C. Little. A proof for the queueing formula: $L = \lambda W$. *Operations Research*, 9:383–387, May-June 1961.
- [63] D. Nain, N. Petigara, and H. Bakakrishnan. Integrated routing and storage for messaging applications in mobile ad hoc networks. *Mobile Networks and Applications (MONET)*, 9(6):595–604, December 2004. An earlier version of this paper appeared in the Proc. of Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt’03).
- [64] P. Nain, D. Towsley, B. Liu, and Z. Liu. Properties of random direction models. In *Proc. of IEEE Infocom 2005*, Miami, FL, March 13-17 2005.
- [65] William Navidi and Tracy Camp. Stationary distributions for the random waypoint mobility model. *IEEE Transactions of Mobile Computing*, 3(1):99–108, January-March 2004.
- [66] T. Ozawa. An analysis for multi-queueing systems with cyclic-service discipline - models with exhaustive and gated services. Technical Report 4, The Institute of Electronics, Information and Communication Engineers (IEICE) (in Japanese), 1987.
- [67] M.L. Puterman. *Markov Decision Processes*. John Wiley & Sons, 1994.
- [68] Sidney Redner. *A Guide to First-Passage Processes*. Cambridge University Press, 2001.
- [69] Philippe Nain Robin Groenevelt, Ger Koole. Message delay in mobile ad hoc networks. In *Proceedings of Performance*, Juan-les-Pins, October 5-7 2005.

- [70] Robert Sedgewick and Philippe Flajolet. *An Introduction to the Analysis of Algorithms*. Addison-Wiley, 1996.
- [71] G. Sharma and Ravi R. Mazumdar. Delay and capacity trade-offs for wireless ad hoc networks with random mobility. Submitted for publication, October 2004.
- [72] G. Sharma and Ravi R. Mazumdar. Scaling laws for capacity and delay in wireless ad hoc networks with random mobility. In *Proc. of International Conference on Communications (ICC 2004)*, Paris, France, June 20-24 2004.
- [73] T. Small and Z. Haas. The shared wireless infostation model - a new ad hoc networking paradigm (or where there is a whale, there is a way). In *Proc. of ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc)*, Annapolis, Maryland, USA, June 2003.
- [74] Richard P. Stanley. *Enumerative Combinatorics*, volume II. Cambridge University Press, second edition, June 2001. Page 221-247 are available at <http://www-math.mit.edu/~rstan/ec/>.
- [75] Jack S. Sykes. Simplified analysis of an alternating-priority queueing model with setup times. *Operations Research*, 18(6):1182–1192, November-December 1970.
- [76] V. Syrotiuk and C.J. Colbourn. Routing in mobile aerial networks. In *Proc. of Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'03)*, Sophia Antipolis, France, March 2003.
- [77] Hideaki Takagi. *Analysis of Polling Systems*. MIT Press, Cambridge, Massachusetts, 1986.
- [78] Hideaki Takagi. Queueing analysis of polling models: An update. *Stochastic Analysis of Computer and Communication Systems*, pages 267–318, 1990.
- [79] Hideaki Takagi. *Queueing Analysis, Vacations and Priority Systems, Part 1*, volume 1. Elsevier Science Publishers B.V., The Netherlands, 1991.
- [80] T.J.Ott and K.R.Krishnan. Seperable routing: A scheme for state-dependent routing of circuit switched telephone traffic. *Annals of Operations Research*, 35:43–68, 1992.
- [81] Author unknown. In dust we trust. *The Economist*, June 12 2004.
- [82] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, the Netherlands, 1997. revised edition.
- [83] Duncan J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 1999.
- [84] Duncan J. Watts and Steven Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, June 4 1998.

-
- [85] G.H. Weiss. *Aspects and Applications of the Random Walk*. Random Materials and Processes. Elsevier Science B.V., North Holland, Amsterdam, The Netherlands, 1994.
- [86] Jungkeun Yoon, Mingyan Liu, and Brian Noble. Random waypoint considered harmful. In *Proceedings of IEEE Infocom*, pages 1312–1321, San Francisco, April 1-3 2003.
- [87] R.M. Young. Euler’s constant. *Math.Gazette*, 472:187–190, 1975.
- [88] Wing Ho Yuen, Roy D. Yates, and Siun-Chuon Mau. Exploiting data diversity and multiuser diversity in noncooperative mobile infostation networks. In *Proc. of IEEE Infocom Conf.*, San Francisco, CA, April 1-3 2003.
- [89] W. Zhao and M. Ammar. Message ferrying: Proactive routing in highly-partioned wireless ad hoc networks. In *Proc. of the 9th IEEE Workshop on Future Trends in Distributed Computed Systems (FTDCS’03)*, San Juan, Puerto Rico, May, 28-30 2003.
- [90] Gil Zussman, Uri Yechiali, and Adrian Segall. Exact probabilistic analysis of the limited scheduling algorithm for symmetrical bluetooth piconets. In *Personal Wireless Communications (PWC)*, Venice, Italy, Sept. 23-25 2003.

Résumé

Dans la première partie de cette thèse, nous étudions la mobilité dans les réseaux ad hoc mobiles. Plus précisément, nous étudions le temps de transfert d'un message, défini comme étant le temps requis pour transmettre un message d'un nœud à un autre nœud, en utilisant éventuellement d'autres nœuds comme relais. Dans un premier temps, pour plusieurs modèles de mobilité, nous trouvons le temps requis afin que deux nœuds puissent communiquer. Dans un deuxième temps, dans les cas unidimensionnel et bidimensionnel, nous caractérisons le temps de transfert d'un message.

Dans le cas bidimensionnel le calcul devient générique et moins sensible au modèle de mobilité considéré, ce qui est très intéressant comme propriété. Ceci permet d'obtenir la transformée de Laplace du temps de transfert d'un message selon deux protocoles de routage différents. En particulier, nous obtenons une approximation simple pour le temps moyen de transfert. Nous montrons également, par des simulations, que cette approximation est robuste et valide pour plusieurs modèles de mobilité et pour une grande gamme de rayons de communication.

La deuxième partie de cette thèse traite d'un système à "Polling" qui consiste en deux files d'attente. Nous obtenons l'expression de plusieurs mesures de performance, notamment le temps moyen d'attente et la taille moyenne de la file d'attente. Grâce à ces expressions et à des exemples, nous montrons que ces quantités augmentent significativement quand il y a corrélation entre les temps de commutation de chaque file d'attente. Cela signifie que la corrélation ne peut pas être ignorée et qu'elle a des implications importantes sur les systèmes de communication dans lesquels un canal de communication commun est partagé entre plusieurs utilisateurs (par exemple dans les réseaux ad hoc).

Dans l'appendice nous étudions deux files d'attente en série avec des "holding costs" pour chaque client dans le système. La fonction de valeur est calculée pour le coût moyen lorsqu'il n'y a pas d'entrée de clients. Cette fonction peut être utilisée pour l'optimisation des systèmes en série ou pour le calcul complet de la fonction de valeur.

Abstract

In the first part of this thesis we focus on the mobility in mobile ad hoc networks. This is achieved by studying the message delay, defined as the time required to send a message from a source node to a destination node, while possibly making use of intermediary relay nodes. For a number of mobility patterns, we first obtain the distribution of the time until two nodes come within one another's communication range. Building on these results, explicit expressions for the message delay are obtained, both in one and in two dimensions, when nodes relay messages for each other and when the network is not fully connected.

We discover that in two dimensions something special happens: calculations become generic in nature (i.e., not so much dependent on the underlying mobility pattern). This enables us to derive closed-form expressions for the Laplace-Stieltjes transform of the message delay for two different relay protocols. In particular, a simple approximation is obtained for the mean message delay. Through simulations it is confirmed that this expression is robust as it holds for all of the mobility patterns considered, any number of nodes, and a wide range of communication radii.

The second part is devoted to the study of polling systems composed of two queues. Explicit expressions are derived for various performance measures, most notably the mean waiting times and the queue lengths. Through examples it is shown that these quantities can significantly increase in the presence of correlation in and between the sequences of switchover times from each queue. This has important implications for communication systems in which a common communication channel is shared amongst various users and where the time between consecutive data transfers is correlated (as is the case in ad hoc networks).

In the appendix we consider a tandem queue with holding costs for each customer. An explicit expression is obtained for the value function of the average costs when there is no inflow of customers. The expression obtained provides an intuitive explanation and can be used for optimisation purposes and for the full derivation of the value function when there is an inflow of customers.