Ph.D. Thesis in agreement for international joint supervision UNIANDES - Universidad de los Andes (Colombia) UNICAMP - Universidade Estadual de Campinas (Brasil)

UNIANDES Doctorado en Ingeniería UNICAMP Doutorado em Ciências da Computação



Geometric Aspects and Random Delays in Probabilistic Broadcasting for Wireless Ad Hoc Networks

Felipe Forero R. f.rodriguez22@uniandes.edu.co frodriguez@ic.unicamp.br

Supervisors: Néstor Misael Peña Traslaviña - UNIANDES Nelson Luis Saldanha da Fonseca - UNICAMP

July 2019 Departamento de Ingeniería Eléctrica y Electrónica UNIANDES - Bogotá, Colombia in joint supervision agreement with Instituto de Computação UNICAMP - Campinas, Brasil

The thesis defense took place on July 3rd, 2019.

Members of the Jury:

Nelson Luis Saldanha da Fonseca (President - Supervisor) Full professor Institute of Computing, UNICAMP.

Edmundo Roberto Mauro Madeira (Reviewer) Full professor Institute of Computing, UNICAMP.

Diego Méndez Chávez (Examiner) Associate professor Department of Electronics Engineering, Universidad Javeriana.

Juan Carlos Bohórquez (Examiner) Associate professor Department of Electrical and Electronics Engineering, UNIANDES.

Néstor Misael Peña Traslaviña (Examiner - Supervisor) Full professor Department of Electrical and Electronics Engineering, UNIANDES.

To my parents.

Acknowledgement

This research was funded by COLCIENCIAS under grant 617-2 (2013) according to the Colombian national policy for higher-level scientific and technological education - Chapter of Local Ph.D. graduates in Colombian universities.

This Ph.D. work was conducted in Agreement for International Joint Supervision of Theses between Universidad de Los Andes and Universidade Estadual de Campinas, and was supported by the CNPq, process # 15/24494-8 of the São Paulo Research Foundation (FAPESP).

Resumo

Esta tese aborda o problema de disseminação de pacotes em roteamento ponto-multiponto (do Inglês, broadcast) em redes sem fio do tipo ad hoc com múltiplos saltos. Aborda-se, especificamente, na tese, protocolos ponto-multiponto nos quais os nós transmissores tomam decisões sobre o encaminhamento dos pacotes de forma probabilística. O uso desses protocolos leva à economia de energia e de banda passante, bem como promove a alternância do papel de retransmissores de pacotes entre nós de um mesmo grupo, com o mínimo de *overhead* possível.

Grande parte da literatura especializada aborda a estimativa dos valores das probabilidades de encaminhamento de pacote (p_f) com o objetivo de reduzir o tamanho do conjunto de nós retransmissores de pacotes bem como promover o maior alcance possível do repasse desses pacotes. Esta tese introduz uma análise que inclui novos parâmetros, tais como a geometria do posicionamento dos nós e a distribuição dos atrasos randômicos nos protocolos ponto-multiponto, que influenciam, também, no tamanho desses conjuntos e na alcançabilidade.

Define-se uma métrica que engloba o valor da razão sinal ruído para cada nó receptor. Para tal, definem-se camadas de grupos de nós vizinhos a um nó transmissor. Calcula-se um valor médio da razão sinal ruído para cada nó e para cada camada de nós, assumindo-se um cenário de interferência máxima. Uma vez que existe correlação entre o valor teórico da razão sinal ruído e o sucesso da recepção de pacotes, demonstrada através de experimentos de simulação e de medições, adota-se a razão sinal ruído como um preditor eficiente do sucesso da recepção de pacotes.

O valor da razão sinal ruído é empregada para se avaliar o impacto da geometria da posição dos nós na alcançabilidade das transmissões ponto-multiponto. Resultados derivados na presente tese evidenciam os limites da alcançabilidade das transmissões pontomultiponto, bem como evidenciam a importância da geometria da posição dos nós em comparação com a posição absoluta dos nós. Demonstra-se que redes compostas por nós posicionados randomicamente promovem valores quase ótimos de alcançabilidade, enquanto que transmissões em redes compostas de nós posicionados em malha são sensíveis às colisões de pacotes e à interferência. Esta tese introduz, também, um arcabouço analítico para quantificar o impacto da variabilidade dos valores dos parâmetros de atrasos randômicos na alcançabilidade, bem como em suas distribuições. Propõe-se uma nova estratégia que utiliza a distribuição exponencial truncada ao invés da distribuição uniforme, que é comumente empregada para redução dos atrasos fim-a-fim em transmissões ponto-multiponto.

Propõe-se uma nova forma de computação da distribuição da latência fim-a-fim que considera a distribuição dos atrasos aleatórios e a geometria do posicionamento dos nós. Determina-se, primeiramente, a função densidade dos números de passos necessários de uma transmissão ponto-multiponto. Obtém-se uma formula para computação do número de passos que é utilizada conjuntamente com a função randômica de atrasos aleatórios para se derivar a distribuição da latência fim-a-fim, o que permite a avaliação das latências em transmissões ponto-multiponto sem utilizar a hipótese de atrasos probabilísticos distribuídos de forma uniforme. A computação proposta é facilmente generalizável par outras distribuições.

Abstract

This thesis addresses the problem of disseminating broadcast packets in a multi-hop wireless ad hoc network. Specifically, it focuses on broadcasting protocols in which nodes make probabilistic decisions to forward packets. In connected multi-hop ad hoc networks, this type of protocol produces a broadcasting process that saves energy and bandwidth while offering a natural rotation of the subset of forwarders over time, yet it entails minimum overhead.

Most of the literature addresses the challenge of estimating the values of the forwarding probability of nodes (p_f) that reduce the size of the subset of forwarders (saved rebroadcast) and maintain broadcast packets reaching most nodes (reachability). This thesis provides an analysis of new factors, other than the estimation of p_f , that also influence the reachability and saved rebroadcast of probabilistic protocols, namely the layout of nodes (i.e. the geometry of node position) and the parameters of the random delays used in the broadcast protocols.

A Signal to Interference metric (probabilistic S/I) is defined for each receiving node by grouping subsets of nodes into tiers around the broadcasting source. An average value of S/I is obtained for each node and then for each tier assuming a worst-case scenario of maximum interference. The theoretical results of S/I exhibit excellent correlation with the probability of successful packet reception obtained from simulations and measured-based experiments, which indicates that such analysis of S/I can be used to predict the average percentage of nodes receiving broadcast packets in every broadcasting event.

The S/I analysis is used to evaluate the impact of the geometry of the position of nodes on reachability. The results show the limitations in reachability of transmitted messages in probabilistic broadcasting schemes as well as the influence on the reachability of the node geometry in comparison to the impact of the position of the source node. It is also shown that networks with randomly-placed nodes exhibit near-ideal levels of reachability, whereas the grid layouts are extremely sensitive to the collisions and interference.

Moreover, the thesis introduces an analytical framework to quantify the impact of changing the parameters of random delays on reachability (e.g. a change in the distribution or a change in the interval of random delays). The thesis introduces a new strategy to reduce end-to-end latency in a multi-hop broadcasting scenario by using the truncated exponential distribution for random delays instead of the wide-spread uniform distribution. A novel model to compute the probability distribution function of broadcast end-toend latency is presented. The model considers both the distribution of random delays and the geometry of node layouts. First, probability mass functions are derived to describe the distribution of the number of hops taken by broadcast packets. Since these functions depend on the geometry of the position of nodes, different models apply to grids (regular geometries) and random node layouts (random geometries). The formulas to compute the probability mass function of the number of hops are then combined with the distribution function of random delays (a per-hop random variable) in a single model to find the probability density function of probabilistic broadcasting end-to-end latency. This allows the evaluation of broadcast latency considering random-delay distributions different from the Uniform distribution and it facilitates the creation of latency models for several nodeposition geometries different from the random Uniform.

Key Words

Ad hoc networks, probabilistic broadcast, signal to interference, MAC layer, random delays, spatial point processes, grids, random geometric graphs.

Resumen

Esta tesis aborda el problema de diseminación de paquetes de difusión en una red inalámbrica sin infraestructura (red inalámbrica ad hoc). De manera específica, la tesis estudia los protocolos de difusión que utilizan variables aleatorias al momento de retransmitir los paquetes durante el proceso de difusión (difusión probabilística). El uso de estos protocolos en redes ad hoc con enlaces multi-salto reduce el consumo de energía y la ocupación del canal mientras que se obtiene una rotación natural del subconjunto de nodos responsables por las retransmisiones del proceso de difusión. Estos beneficios tienen un costo muy bajo en la complejidad de implementación de los protocolos.

La mayoría de los estudios se enfocan en estimar el valor de la probabilidad de retransmisión de los paquetes (valor de p_f), por parte de los nodos, de modo tal que se reduzca el tamaño del subconjunto de *retransmisores* (ahorro en retransmisiones) y se mantenga el número promedio de nodos que reciben los mensajes de difusión (alcance de los paquetes de difusión). Esta tesis aporta un análisis de nuevos factores, diferentes del valor estimado de p_f , que tienen una influencia en el alcance y el ahorro en retransmisiones provisto por los protocolos de difusión probabilística. De forma específica, se analizan factores como la geometría de la posición de los nodos y la configuración de los retrasos aleatorios utilizados en el proceso de difusión.

Se define una métrica basada en la relación Señal a Interferencia (S/I probabilística) para todos los nodos de la red al momento de recibir los paquetes de difusión. Dicha métrica se calcula agrupando los nodos en *anillos* al rededor del nodo que emitió el paquete original (nodo fuente). El valor promedio de S/I probabilística se obtiene en cada nodo para el caso de máxima interferencia (peor caso). Los resultados teóricos de S/I muestran una correlación alta con respecto a la probabilidad de recibir un paquete sin errores; probabilidades tomadas de estudios basados en mediciones de transmisiones inalámbricas punto a punto. Los resultados indican que el análisis teórico de S/I puede ser utilizado para predecir el alcance promedio de los paquetes de difusión.

El análisis de S/I se utiliza para evaluar el impacto de la geometría de la posición de los nodos en el alcance de los paquetes de difusión. Los resulados muestran las limitaciones de los esquemas de difusión probabilística en términos del alcance. También se muestra la influencia que tiene la geometría de la posición de los nodos en comparación con el impacto de la posición del nodo fuente. El análisis de los resultados permite establecer

que una distribución aleatoria de las posiciones de los nodos (geometría aleatoria) produce un alcance promedio casi ideal, mientras que las geometrías regulares (nodos ubicados de acuerdo a alguna retícula) hacen que la red sea muy susceptible a los efectos de la interferencia y las colisiones entre paquetes.

Adicionalmente, la tesis introduce un marco analítico para cuantificar el impacto de los retrasos aleatorios sobre el alcance de los paquetes de difusión; impacto que se produce al modificar los parámetros de dichos retrasos, específicamente cuando se cambia la función de distribución de probabilidad o el intervalo donde se define el retraso. La tesis introduce una estrategia nueva que reduce la latencia de extremo a extremo, en redes multi-salto, mediante el uso de la distribución exponencial truncada como alternativa ante el uso generalizado de la distribución uniforme en los retrasos aleatorios.

Se propone además un modelo inovador para calcular la función de distribución de probabilidad de la latencia extremo a extremo. El modelo considera tanto la distribucón de los retrasos aleatorios como la geometría de la posición de los nodos. Primero se deducen funciones de masa de probabilidad para describir el comportamiento del número de saltos de los paquetes de difusión. Como las funciones de masa de probabilidad están asociadas a la geometría de la red, se aplican diferentes modelos a las retículas (nodos con posiciones en geometría regular) y a las redes con geometría aleatoria. Las funciones de masa de probabilidad obtenidas se combinan con las funciones de distribución de los retrasos aleatorios (funciones de densidad de probabilidad) en un solo modelo de donde se obtiene finalmente la función de densidad de probabilidad de la latencia extremo a extremo de los paquetes de difusión. Este modelo permite evaluar la latencia de los protocolos de difusión basados en probabilidad considerando diferentes tipos de distribución para los retrasos aleatorios (diferentes de la distribución uniforme) y a su vez facilita la creación de modelos de latencia que consideran varias geometrías para la posición de los nodos (diferentes retículas y posiciones aleatorias).

Contents

1	Intr	oducti	ion	17			
2	Concept Framework						
	2.1	Basic	Description	21			
	2.2	Forma	l Description	27			
	2.3	Techni	ical Description	33			
	2.4	Frame	work Summary	37			
3	Literature Review						
	3.1	Neight	bor-based probabilistic broadcast	39			
		3.1.1	One-hop neighborhood	40			
		3.1.2	Two-hop neighborhood	44			
		3.1.3	Timeline of neighbor-based schemes	45			
	3.2	Area-b	pased probabilistic broadcast	46			
		3.2.1	Area Estimation by Number of Nodes	48			
		3.2.2	Area Estimation by Distance Estimation	49			
		3.2.3	Area Estimation by counting copies	51			
		3.2.4	Timeline of Area-based schemes	51			
	3.3	Altern	ative probabilistic broadcast schemes	53			
		3.3.1	Timeline of Alternative schemes	55			
	3.4	Recent	t Studies	56			
		3.4.1	Congestion Awareness in Studies	56			
		3.4.2	Routing-oriented Studies	57			
		3.4.3	Studies on Energy Conservation	58			
		3.4.4	Schemes applied to VANETs	59			
		3.4.5	Studies with Analytical Models	60			
		3.4.6	Other Studies	61			
	3.5	Tradeo	offs and Issues	62			
	3.6	Locati	ng the Contributions of this Thesis	63			
4	Geometric Aspects of Probabilistic Broadcasting in Ad Hoc Networks						
	4.1	Introd	uction	65			
	4.2	Table	of Chapter Variables	67			
	4.3	Relate	d Work	67			

	4.4	Impact of Geometrical Node Distribution on Probabilistic Broadcast 6	59 75
	4.5	Simulation Experiments	о 77
		4.5.1 Random and Regular Placement of Nodes	С 77
	1.0	4.5.2 Other Grid Geometries and Node Distributions	ິ () ຄ
	4.0	Chapter Summary and Future work	52
	4.7	Table with Related Work	54
5	Lat	ency Reduction in Probabilistic Broadcast Protocols for Ad Hoc Net-	
	wor	KS 8	
	5.1	Introduction	57
	5.2	Table of Chapter Variables	58 20
	5.3	Related Work	<u> 59</u>
	5.4	Impact of the employment of exponentially distributed random delays on	0
			59 50
		5.4.1 Probability of two overlapping backon periods	1U 1
		5.4.2 Probability of two transmissions beginning in the same slot 9)1 \0
	0.0 5 C	Simulation Results	13
	5.0	Table with Dalated Week)4)5
	5.7	Table with Related Work	10
6	Dis	tribution of Latency in Ad Hoc Probabilistic Broadcast 9	7
	6.1)7
	6.2	Table of Chapter Variables)9
	6.3	Related Work	99
	6.4	Path Length According to Node Layout)()
		6.4.1 Regular Geometries (Grids))1
		6.4.1.1 Path-length Model)]
		$6.4.1.2 \text{Square Grid} \dots \dots \dots \dots \dots \dots \dots \dots \dots $)2
		6.4.1.3 Hexagonal Grid)6
		$6.4.1.4$ Triangular Grid \ldots)7
		$6.4.2 \text{Random Geometries} \dots \dots \dots \dots \dots \dots \dots \dots \dots $.2
		$6.4.2.1 \text{Estimating the Constant} \dots \dots \dots \dots \dots \dots \dots \dots \dots $.2
	с г	$6.4.2.2$ Finding the pmi \ldots	.4
	0.0	par of Random Sums	.ð .0
	6.6	0.5.1 Latency pdf for uniform random delays	.ð .0
	0.0	Validation of Latency Models	.9
		6.6.1.1 Validation for the Severe Crid	20
		6.6.1.2 Validation for the Heneranal Crid	2U 01
		6.6.1.2 Validation for the Triangular Crid	51 51
		$0.0.1.5$ valuation for the margurar Grid \ldots \ldots 12	52 52
		6.6.2 Validation of Latoney Model for Pandem Networks	50 50
	67	Chapter Summary and Future Work	:0)E
	0.1 6 9	Table with Related Work	61 AC
	0.0	TADIE WITH RELATED WOLK	0
	~		

7 Conclusions and Recommendations

Chapter ONE

Introduction

This thesis addresses the problem of disseminating broadcast packets in a multi-hop wireless ad hoc network (e.g. MANET, VANET or WSN). Specifically, the thesis focuses on the study of broadcasting protocols in which nodes make probabilistic decisions to forward packets. Unlike broadcasting by flooding (i.e. allowing every node to forward broadcast packets), probabilistic protocols enable each node to draw a random variable to decide whether or not a packet should be forwarded. In connected multi-hop ad hoc networks, this type of protocol produces a broadcasting process that saves energy and bandwidth while offering a natural rotation of the subset of forwarders over time, yet it entails minimum overhead.

Most of the literature addresses the challenge of estimating the values of the forwarding probability of nodes (p_f) that reduce the size of the subset of forwarders (Saved Rebroadcast) and maintain broadcast packets reaching most nodes (Reachability). This thesis provides an analysis of new factors, other than the estimation of p_f , that also influence the Reachability and Saved Rebroadcast of probabilistic protocols, namely the layout of nodes (i.e. the geometry of node position) and the parameters of the random delays used in protocols.

This thesis begins by describing the concepts of packet dissemination in multi-hop, ad-hoc networks at a basic, formal, and technical level (Chapter 2). Chapter 3 offers a comparative survey of the mainstream of proposals. Studies are classified according to one fundamental characterisitic in the operation of protocols, namely the use of Hello packets. This characteristic is strongly connected to a well-known division of probabilistic broadcasting schemes into neighbor-based schemes and area-based schemes. These two groups of probabilistic broadcasting protocols are compared by analyzing their results in scenarios with static nodes placed according to the random uniform distribution in a square region or mobile nodes following the Random Way Point and Random Direction models. Subsequently, families of probability functions, used by authors to estimate the forwarding probability (p_f) of nodes are also classified. Finally, the resulting values of p_f as a function of node density are compared with the protocol parameters suggested by authors.

The evidence gathered in Chapter 3 shows a convergence of the values of p_f obtained with different proposals as a function of node density. Figure 1.1 shows a map of the context of probabilistic broadcasting studies that summarizes the collection of results analyzed in Chapter 3.



Figure 1.1: A map of the main variables analyzed in the literature about probabilistic broadcast for wireless ad hoc networks. Shaded areas summarize the expected performance of probabilistic broadcast results within a research time span of 15 years. The color code to the right indicates the success of the dissemination process in terms of reachability (i.e. the average percentage of nodes reached by broadcast packets).

In Figure 1.1, the two most sensible variables in probabilistic broadcast studies, namely reachability (y-axis) and saved rebroadcast (SRB in the top x-axis) are related to influential aspects of network topology and protocol parameters, such as node degree (bottom x-axis), the geometry of node position, and the role of random delays. The gray strips represent the regions of performance of probabilistic broadcasting protocols, *flooding* (the default ad hoc broadcast scheme), and the analytical models found in the literature, in which an ideal MAC layer is commonly assumed. For example, the performance region associated to analytical models indicates that reachability above 95% is obtained for any value of node degree, continuously reducing the rebroadcast rate as node density increases.

Regarding the other performance regions of Figure 1.1, the graceful drop in the shaded region of flooding means that savings in message retransmissions (i.e. an increase in SRB) only occur when the broadcast process stops at an early stage due to an excess of collision events near the broadcast source, leading to failure of the dissemination process (ending in the red region of reachability). The shaded region of probabilistic broadcast indicates that a poor calibration of broadcast protocols can lead to very limited reachability (there is an example with reachability ending up in the yellow region). However, with proper protocol calibration, probabilistic broadcasting schemes continue to reduce unnecessary resource waste (higher SRB) and reach over 90% of the nodes as node density increases). Finally, the lower-left corner of the diagram indicates that, in random networks, average node degrees lower than 6 produce disconnected networks, which becomes the subject of study for Delay Tolerant Network (DTN) protocols.

The ideas in Figure 1.1 are derived for the two most frequent network scenarios, namely static networks with nodes placed according to a two-dimensional random uniform distribution, and mobile nodes following the Random Way Point (RWP) model or Random Direction Model.

The content of Chapters 4–6 focuses on modelling and evaluating the impact of node layout geometries and random delays on probabilistic broadcast performance. As indicated by the small arrows in Figure 1.1, the geometrical conditions of node layouts can reduce the expected performance of broadcast schemes (i.e. lower reachability values for the same rebroadcast rate). Conversely, the use of random delays makes broadcast schemes approach their ideally modeled behavior. These two aspects constitute the core of the present work.

Chapter 4 provides an analysis of the Signal to Interference ratio seen by nodes during the propagation of a broadcast packet. The analysis illustrates how critical the geometry of node layouts can be to allow broadcast packets to reach most nodes in each dissemination event (i.e. the impact of geometry on reachability). The use of Signal to Interference analysis considers realistic channel conditions of wireless networks, which is seldom a concern in the evaluation of probabilistic broadcast proposals. In this context, the contributions of Chapter 3 are as follows:

- The incorporation of realistic assumptions to analyze the causes of the limitations in reachability of transmitted messages in probabilistic broadcasting schemes.
- The analysis of the impact of the geometry of network layouts on the reachability of probabilistic broadcasting protocols and the stronger influence of the node geometry when compared to the impact of the position of the source node.

Chapter 5 provides an analytical approach to assess the change in the probability distribution function of random delays, which are a common feature of probabilistic broadcast schemes. The analysis shows that, by changing the probability distribution of random delays, a significant reduction in average broadcast latency is achieved. The contributions of Chapter 5 are as follows:

- A strategy to reduce broadcast end-to-end latency that can be applied to any probabilistic broadcast protocol.
- A proposition to prove the negligible impact of exponentially distributed random delays on collision events at the MAC layer.

In Chapter 6, a set of discrete probability distribution functions that describe the number of hops traversed by packets in an ad hoc network is derived for various scenarios, considering the geometry of node layout. This set of discrete distributions is combined with the distribution of random delays to compute the probability density function (pdf) of broadcast latency (a known parameter of protocols). The contributions of Chapter 6 are as follows:

- Computation of probability density functions (pdf) of broadcast latency.
- Evaluation of broadcast latency considering random-delay distributions different from the Uniform distribution.
- Latency models that consider node-position geometries different from random Uniform.

The content of Chapter 4 has already been published in the Ad Hoc Networks journal and can be cited as:

Forero F., Peña N. M., da Fonseca Nelson L.S., Geometric aspects of probabilistic broadcasting in ad hoc networks. *Ad Hoc Networks*, vol. 87, pp. 146–156, 2019. DOI: 10.1016/j.adhoc.2018.11.014.

The content of Chapter 5 has been accepted for publication in the *IEEE Wireless Communications Letters* journal and can be cited as:

Forero F., Peña N. M., da Fonseca Nelson L. S., Latency Reduction in Probabilistic Broadcast Protocols for Ad Hoc Networks. *IEEE Wireless Communications Letters*, vol. 8, pp. 1268–1271, Aug. 2019. DOI: 10.1109/LWC.2019.2915077.

The content of Chapter 6 constitutes original material for future publications.

Chapter
TWO

Concept Framework

This chapter introduces the fundamental ideas of *percolation*, which constitute the theoretical grounding of probabilistic broadcasting schemes. First, a basic description of the process of dissemination of packets over a multi-hop ad-hoc network is presented. This description familiarizes readers with the most frequent concepts that appear throughout the development of this thesis, namely the concept of *reachability* and the use of *tiers* to characterize the propagation of broadcast packets away from a source node.

Then, a formal description of the problem is presented by providing references to seminal work in the literature as well as by giving examples of the type of percolation principles applied by the authors of probabilistic broadcasting protocols. Finally, a technical description of the problem completes the framework by incorporating the elements of existing wireless technology (i.e. network, data-link and physical layer considerations) into the context of multi-hop ad hoc broadcasting protocols.

2.1 Basic Description

The process of disseminating a packet in a multi-hop environment begins with an arbitrary source node transmitting a new broadcast packet to its neighbors. Figure 2.1a shows a situation in which the transmission radius (r) of the source is just enough to reach 5 nodes around it. Since the packet is intended for dissemination, *some* of the 5 nodes (2 nodes in red in Figure 2.1b) that received a copy of the packet directly from the source should forward it so that copies of the packet propagate away from the source. Figure 2.1b shows how 4 aditional nodes receive a copy of the original broadcast packet. In Figure 2.1a, the 5 receivers can be considered as a *first tier* that forms around the source, whereas the 4 additional receivers in Figure 2.1b form a *second tier*. The concept of tier formation assumes that all copies of the packet are received successfully; then, the tier of a given node, with respect to a source, represents the minimum number of hops necessary for packets to travel from the source to the intended node. Thus, under ideal conditions (no packet errors, collisions, or packet losses), a node in Tier 2 should receive a copy of the broadcast packet from the source should receive a in Tier 1 and, in general, nodes at any distance from the source should receive packets from other nodes in the previous tier.



Figure 2.1: Example of tier formation around a broadcasting source.

The concept of tier formation around a broadcast source establishes a topological distance, namely a *hop distance*, that allows characterizing the propagation of broadcast packets. Indeed, after defining subsets of nodes into tiers, it is possible to study the random paths taken by broadcast packets when reaching nodes at different distances from the source (see Chapter 6).

Figures 2.2–2.5 show a sequence of tiers forming. As the transmissions in a tier pass the copies of the broadcast packet to the next tier, more nodes are *reached*. At the end of the sequence (i.e. after a hop-distance equal to 4), 21 nodes out of 500 have been reached, then it can be said that a fraction of the network equal to 0.042 has been reached.

Figures 2.6–2.9 show four non-contiguous snapshots of different tiers for the same 500node network. Specifically, Figure 2.6 shows a saded area covering the nodes that have received a copy of the broadcast packet after 5 forwarding events (when the nodes in Tier 5 have received the packet). Figures 2.7–2.9 show the same type of shaded area for Tiers 10, 14 and 19, respectively. In each figure, the fraction of the network that has been reached is indicated in the caption and is called *reachability*. Therefore, at the end of a dissemination event, the value of reachability is a metric of the success of a broadcasting protocol. In the literature, the average reachability value is computed for several replications of a dissemination process; hence, such an average value becomes an estimation of the probability that nodes receive a broadcast packet. In this thesis, the term *reachability* is used for the *average* values of reachability, and this metric is also applied to subsets of nodes in the same tier so that the behavior of reachability as a function of the hop-distance from an arbitrary source can be studied.

From the previous description, it is important to notice that:

- The study focuses on multi-hop dissemination of packets.
- Dissemination should not rely on a previously established topology.
- Packets reach most, not necessarily all, nodes (i.e. the concept of reachability).
- Not all nodes should forward all packets (only a subset of nodes).



Figure 2.2: Tier 1 around a broadcast source; 5 out of 500 nodes reached. Fraction of the network reached = 0.01.



Figure 2.3: Tier 2 around a broadcast source; apart from the source, 9 out of 500 nodes have been reached (gray area). Fraction of the network reached = 0.018.



Figure 2.4: Tier 3 around a broadcast source; apart from the source, 15 out of 500 nodes have been reached (gray area). Fraction of the network reached = 0.03.



Figure 2.5: Tier 4 around a broadcast source; apart from the source, 30 out of 500 nodes have been reached (gray area). Fraction of the network reached = 0.06.



Figure 2.6: Nodes in black representing Tier 5 around the broadcast source. The gray area shows the fraction of the network reached; reachability = 0.11.



Figure 2.7: Nodes in black representing Tier 10 around the broadcast source. The gray area shows the fraction of the network reached; reachability = 0.48.



Figure 2.8: Nodes in black representing Tier 14 around the broadcast source. The gray area shows the fraction of the network reached; reachability = 0.70.



Figure 2.9: Nodes in black representing Tier 19 around the broadcast source. The gray area shows the fraction of the network reached; reachability = 0.93.

• Subsets of nodes sharing the same minimum hop-distance from the broadcas source can be seen as tiers.

The following section introduces a formal version of these ideas by presenting the theoretical formulations that have motivated the use of probability to design protocols (schemes) for packet dissemination in multi-hop ad-hoc networks.

2.2 Formal Description

The formal description used in the literature to analyze the multi-hop dissemination of packets in an ad hoc network corresponds to the so called Geometric Graph on a random point configuration also called Random Geometric Graph (RGG) [1,2]. Such a description defines a graph, in an *d*-dimensional space, in terms of the transmission radius of nodes, r, as follows:

Let $\|\cdot\|$ be some norm on \mathbb{R}^d and let r be some positive parameter. Given a finite set $X \in \mathbb{R}^d$, G(X;r) denotes the undirected graph with vertex set X and edges connecting all those pairs x_i, x_j with $||x_i - x_j|| < r$, $i \neq j$.

Figure 2.10 shows an example of a Random Geometric Graph (RGG) in the twodimensional space (d = 2). Every edge of the graph represents a communication link between pairs of nodes. Notice that most of the nodes belong to a connected component, and only a few nodes remain disconnected. This property of RGGs supports the idea that most nodes, not necessarily all nodes, may receive the broadcast packet in a single dissemination process (a single realization of a random graph).



Figure 2.10: Random Geometric Graph. The source node is pointed at using a red arrow.

Once the graph is computed, an optimal subset of edges that maintains the connected component can be found. This means that the dissemination process can be completed with a minimum set of forwarders, producing only the necessary links for connectivity in the graph. Figure 2.11 shows an example of what is called a Broadcast Graph or Broadcast Tree rooted at a source node. This example results from applying a tree algorithm to the RGG realization shown in Figure 2.10.



Figure 2.11: Broadcast Tree obtained from a Random Geometric Graph. The source node is pointed at using a red arrow.

Although the resulting graph (tree) is an optimal solution (i.e. same broadcast reachability with minimum subset of forwarders), the algorithm to find such a graph assumes previous knowledge of the existing topology, that is, knowledge of all existing edges in the original RGG, which is unfeasible for ad hoc broadcasting protocols. Moreover, the solution to the tree is found using the so called Minimum Spanning Tree (MST) algorithm, which is known to be an NP-hard problem [3].

Since optimal solutions require global knowledge of the graph as well as NP-hard algorithms, other solutions are envisioned. The main idea is to avoid building a pre-established broadcast topology, or broadcast underlying backbone, since it is impractical due to the dynamic nature of the topology of ad hoc networks (nodes joining/leaving the network at any time as well as mobile nodes). Moreover, a broadcast tree depends on the root node; then, optimal solutions assume that there is a tree for every possible broadcast source. In this context, the properties of RGGs can be studied to come up with a type of solution that suits the needs of ad hoc broadcasting protocols (i.e. fully-distributed, lightweight solutions). A property called *percolation* is formally introduced for RGGs in [2] and applied to the problem of ad hoc broadcasting in [1]. The fundamental idea of percolation for RGGs reads as follows [1, 2]: Let the set $X_{\lambda} \in \mathbb{R}^d$ be a Poisson Point Process with density λ . Consider the graph $G(X_{\lambda}; r)$, then the percolation probability $\theta(\lambda)$ is the probability that the node at the origin is connected to infinitely many nodes of the graph $G(X_{\lambda}; r)$. Again r is the transmission radius of nodes.

Percolation in RGGs is a phase transition phenomenon that depends on a node density threshold, that is, below a critical value of density no percolation occurs; formally:

The percolation threshold density is $\lambda_c = \inf\{\lambda > 0 : \theta(\lambda) > 0\}$. With ϕ denoting the node degree (the number of neigbors), then the relation $\phi = \lambda \pi r^2$ makes $G(X_{\lambda}; r) = G(\phi)$. Therefore, percolation is determined by a critical average node degree $\phi_c = \lambda_c \pi r^2$ with percolation probability $\theta(\phi)$. It is important to notice that $\theta(\phi) > 0$ only for $\phi > \phi_c$. Moreover, above ϕ_c , $\theta(\phi)$ is an increasing function of ϕ .

Figure 2.12, taken from [1], shows an example of $\theta(\phi)$. Although infinite networks are unfeasible in practice, the curves in Figure 2.12 were obtained for different types of *large networks*. For a random network contained in a squared area of side length L, a size parameter α is defined such that $\alpha r = L$. Then, for an average degree (neighbor density) of $3 \le \phi \le 8$, $\alpha = 20$ yields networks of 400–1000 nodes; $\alpha = 40$ yields networks of 1500–4000 nodes; and $\alpha = 80$ yields networks of 6000–16000 nodes.



Figure 2.12: Sample curves of percolation probability, $\theta(\phi)$, obtained from simulation in [1].

As expected from the theory of percolation (i.e. infinite networks), larger networks exhibit a faster transition for $\theta(\phi)$ from $\theta(\phi) \to 0$ to $\theta(\phi) \to 1$. The value of ϕ at which such a transition occurs is then ϕ_c .

Percolation also occurs in *regular* geometric graphs, that is, in grids. Since grids have a constant node degree (number of neighbors), percolation is not associated to a critical node degree but to a critical probability of edges/nodes existing in the resulting graph. Specifically, when edges (bonds) appear in a graph with probability p, the process of forming a large connected component (connected graph) over the grid is called bond percolation; instead, when nodes (sites) appear in a graph with probability p, the process is called site percolation.

Formal definitions of percolation in regular graphs are provided in [4,5]. For example, bond percolation is defined as follows:

Let \mathbb{Z} be the set of all integers and \mathbb{Z}^d the set of all vectors $x = (x_1, x_2, \dots, x_d)$. Then, \mathbb{Z}^d can be turned into a graph (*d*-dimensional cubic lattice), by adding edges between all pairs of points $(x, y) \in \mathbb{Z}^d$ with $\delta(x, y) = 1$, where $\delta(x, y) = \sum_{k=1}^d |x_k - y_k|$.

The lattice (graph) is defined with vertex and edge sets as $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$. Each edge of \mathbb{L}^d will be open with probability p and closed otherwise. Percolation probability $\theta(p)$ is the probability that a given vertex belongs to an infinite open cluster. There exists a critical value $p_c = p_c(d)$ of p such that:

$$\theta(p) \begin{cases} = 0, & \text{if } p < p_c, \\ > 0, & \text{if } p > p_c; \end{cases}$$

 p_c is called the *critical probability* and is defined formally by $p_c(d) = \sup\{p : \theta(p) = 0\}$.

As mentioned above, in grids, the value of p_c plays the role of ϕ_c in RGGs. A theoretical representation of $\theta(p)$, which is equivalent to the empirical representation of Figure 2.12, can be seen in Figure 2.13 (taken from [4]). As in the case of RGGs, the function $\theta(p)$ is also increasing and shows a phase transition at a critical value (a sudden change from $\theta(p) \to 0$ to $\theta(p) \to 1$ at p_c).



Figure 2.13: Hypothetical function $\theta(p)$ proposed in [4].

The most representative examples of bond and site percolation applied to ad hoc broadcasting schemes appear in [6,7]. In these studies, percolation is applied to ad hoc broadcasting schemes to show that redundancy and energy consumption can be reduced by allowing nodes to forward packets with a value of probabily p (probabilistic forwarding). In [7], directional antennas are simulated in nodes so that the probabilistic forwarding of packets can be mapped to the bond percolation model, whereas omnidirectional antennas are mapped to the site percolation model.

Figure 2.14, taken from [7], is equivalent to Figure 2.12 for the case of percolation in RGGs. In this case, the percolation probability is a function of p, and the various values of p_c for the two-dimensional case can be observed for different grids and different models. The curves obtained with grids of degree 3, 4, 6, and 8 are labelled using the initials OMN (omnidiectional, site model) and DIR (directional, bond model).



Figure 2.14: Sample curves of percolation probability in grids; grids of degree 3, 4, 6, and 8; taken from [7].

Figures 2.15–2.17 show examples of the resulting graphs in small regions of 50×50 grids with different node degrees. All graphs were obtained from networks with omnidirectional antennas with site percolation probabilities above their corresponding critical values. It can be seen that the graphs are highly connected, leaving only a few nodes out of the large component.

In summary, the theory of percolation provides solid grounding for ad hoc broadcasting schemes to make use of a random selection of the forwarding set (forwarding probability) and still disseminate broadcast packets onto a large fraction of the network (a fraction near unity, associated to a large connected component). In the case of RGGs, nodes may randomly decide to forward broadcast packets as long as the resulting average node



Figure 2.15: Example of a square grid with p = 0.65 using site percolation (omnidirectional antennas), $p_c = 0.593$.



Figure 2.16: Example of a hexagonal grid with p = 0.8 using site percolation (omnidirectional antennas), $p_c = 0.696$.



Figure 2.17: Example of a triangular grid with p = 0.65, using site percolation (omnidirectional antennas), $p_c = 0.5$.

degree, namely the degree of the subgraph formed by forwarders only, is higher than the critical degree ϕ_c . In the case of grids, it suffices with having a forwarding probability value greater than the value of p_c .

In principle, by allowing nodes to forward broadcast packets independently with probability p, it is possible to have an effective dissemination process (i.e. high reachability) that is based on a lightweight, fully distributed broadcasting scheme.

2.3 Technical Description

This section introduces technological aspects that need to be considered in the development of any packet dissemination protocol intended for wireless multi-hop ad-hoc scenarios. The ideas summarized in this section are taken from the description of the *Broadcast Storm Problem* first introduced in [8]. However, this section focuses exclusively on the role of the data link layer (Medium Access Control–MAC) and the physical layer (Wireless medium) of current technology (IEEE 802.11 protocol suite) leading to essential problems such as redundancy, contention and packet collisions.

The characteristics of an ad-hoc broadcasting process define the needs that protocols should meet. Since ad hoc broadcast relies on a wireless channel employing the Carrier Sensing Multiple Access with Collision Avoidance mechanism (CSMA/CA in IEEE 802.11 standard [9]), the following features have guided the desing of protocols [10].

- Broadcast is spotaneous; therefore, any sinchronization of nodes and/or preparation of global topology knowledge are prohibitive.
- There is no collision detection mechanism available. Once a collision begins, transmissions continue, so energy and bandwidth are wasted with no oportunity to recover colliding packets.
- Broadcats is unreliable. Since broadcast packets are intended for all nodes, the use of confirmation (ACK) or dialogue (RTS/CTS) packets would saturate the wireless media around transmitters, impeding communication altogether.
- In many applications, such as route-finding, 100% reliable broadcast is not necessary (i.e. a few nodes missing some of the broadcast packets).

The simplest of all ad-hoc broadcasting protocols consists in allowing every node to forward broadcast packets once only (a scheme called *flooding*). With only one forwarding event per node, a broadcast packet should propagate to all connected nodes. This protocol is totally distributed and lightweight, and meets all the conditions stated above. However, when the size of the network increases, multi-hop communication and an increasing node density reduce the applicability of flooding. The main problems analyzed in [8] are redundancy, contention and packet collisions.

The redundancy analysis suggests that many of the forwarding transmissions of broadcast packets are unnecessary in ad hoc networks. Specifically, the analysis focuses on computing the Expected Additional Coverage (EAC) area of a forwarding transmission. Figure 2.18 shows that for two nodes, lying in each other's transmission radius, the maximum additional area covered by forwarding a packet is $0.61\pi r^2$. This means that, once a packet is received, no more than 61% additional coverage can be reached; on average, the value of EAC for the same two nodes is only 41%.





Figure 2.19 shows how the values of EAC (i.e. average additional area) decrease rapidly with only a few nodes sharing their transmission radii. The first value of the curve corresponds to the case of one forwarder (i.e. two nodes in Figure 2.18). With as few as 4 forwarders, the value of EAC is nearly 0.05. This analysis illustrates that, by allowing all



Figure 2.19: Curve of Expected Additional Coverage area (EAC) as a function of the number of forwarders; taken from [8].

nodes to forward every broadcast packet, flooding leads to many unnecessary forwarding transmissions.

During a broadcasting process, contention for the wireless media occurs when the same copy of a broadcast packet is received by several nodes and some of the receivers are located close enough so that they can detect each other's forwarding transmissions. In Figure 2.20, when node B receives a packet from node A, three sets of nodes can be established, namely the nodes that are neighbors of A and B (nodes in $A \cap B$), neighbors of A only (nodes in the area $A - A \cap B$), and neighbors of B only (nodes in the area $B - A \cap B$). Contention occurs for the nodes in $A \cap B$ since these nodes received the same copy of the broadcast packet from A and will surely detect the transmission of B. Based on this observation, simulations were conducted in [8] to estimate the probability of having k, out of n, receivers of A free of contention (a contention-free function cf(n, k)). Figure 2.21 shows that the probability of having zero contention-free receivers (i.e. contention) is always higher than 0.5. Indeed, when A has more than 5 neighbors the probability of contention (i.e. zero contention-free neighbors) remains above 0.8. Considering that the critical average number of neighbors is larger than 5 for percolation in finite networks, contention can be considered a certain event.

Given that contention is unavoidable, packet collisions are very likely to occur during the propagation of broadcast packets. Considering again the description of Figures 2.20 and 2.21, and the fact that nodes have no previous knowledge of their local topology, if all contending nodes run the same carrier sensing algorithm (e.g. 802.11 MAC), they should start their forwarding transmissions almost simultaneously. Since carrier sensing is not immediate, several packets will collide with no collision detection mechanism to recover from the loss of packets. Such a collision event results in several nodes not receiving broadcast packets (corrupted/interfered packets), which significantly reduces the reachability of the dissemination process.



Figure 2.20: Three sets of nodes for contention analysis after node B receives a packet from node A.



Figure 2.21: Estimation of a contention-free function cf(n, k), where n is the number of neighbors of a transmitter from which k remain free of contention when forwarding a broadcast packet; taken from [8].
2.4 Framework Summary

Table 2.1 gathers the most relevant ideas of the three types of description provided in the present chapter. These ideas are used in the rest of the thesis to make assumptions for modeling and also to assess the performance of probabilistic broadcasting schemes.

Basic concepts	Formal concepts	Technical concepts
Multi-hop communication.	Percolation.	CSMA/CA technology.
Most (not all) nodes reached.	Random Geometric Graphs.	Synchronized transmissions un-
		feasible.
Not all nodes need to forward	Critical average degree ϕ .	No global information avail-
packets.		able.
Reachability.	Regular graphs (grids).	Flooding.
Hop distance from source.	Critical bond/site probability	Redundancy.
	(p).	
Tier formation.	Trees are NP problems.	Contention.
	Practical networks are finite,	Packet collisions.
	but large.	

Table 2.1: Concept framework summary

In summary, the problem of disseminating a packet over a multi-hop ad-hoc network reduces to finding a subset of forwarders to deliver broadcast packets to most nodes while avoiding the redundancy of *flooding*, which in turn reduces contention and packet collisions in the wireless media. Figure 2.22 further illustrates this idea and offers a comparison between deterministic and probabilistic solutions that justifies the use of probabilistic schemes. Specifically, although deterministic solutions (e.g. broadcast trees) produce optimal subsets of forwarders, probabilistic solutions do not require a pre-established backbone topology to be built before the beginning of the dissemination process. At its simplest, if every node forwards broadcast packets independently according to a random variable locally computed, the properties of percolation will apply and dissemination will take place. Probabilistic broadcasting schemes do not require pre-established topologies based on global knowledge and provide an implicit rotation of the subset of forwarders in every broadcasting event, which leads to an even distribution of workload and energy consumption. Let us toss the coin!



Figure 2.22: A diagram comparing the advantages and disadvantages between deterministic and probabilistic ad-hoc broadcasting solutions.

Chapter THREE

Literature Review

The purpose of this chapter is to present the mainstream of probabilistic broadcast solutions proposed over a 15-year period before the beginning of this thesis, and then show current research trends, including the contributions made in subsequent chapters. This chapter begins by classifying the fundamental principles that have motivated the existence and evolution of particular groups of probabilistic broadcast schemes. For each group, the studies that provide analytical foundations are prioritized and discussed. In each section, other studies that have explored the impact of adding heuristic complements are presented according to their specific target applications (e.g. MANET, WSN or VANET). This chapter should provide readers with a structured framework to conveniently address the fundamental issues of probabilistic broadcast, or else, to quickly find the group of techniques that suits a particular need.

Figure 3.1 corresponds to the broad-classification groups (sections) contained in the present chapter. The figure suggests that the two fundamental ideas motivating most probabilistic broadcast studies lie in the topological (*neighbor-based*) and positioning (*area-based*) features of ad hoc networks. Only a few proposals address probabilistic broadcast solutions with *alternative* principles. However, regardless of the solution, several *trade-offs* prevail when looking at the performance of probabilistic broadcast schemes in the literature. It is worth mentioning that some authors have combined their solutions in order to achieve more robust and reliable performance, but in every study, the main idea effortlessly stands out from the subsidiary (complementary) ideas. Therefore, the broad classification of Figure 3.1 still applies.

Both the intuitive and analytical aspects of neighbor-based and area-based solutions are presented in subsequent sections, followed by a small collection of alternative probabilistic broadcast solutions. The chapter unfolds by explaining the causes and implications associated to the trade-offs pointed out in Figure 3.1. Some issues identified as common to all of the studies presented herein are discussed at the end of the chapter.

3.1 Neighbor-based probabilistic broadcast

The use of neighbors information to estimate the best possible probabilistic broadcast scheme can be classified according to two principles. First, there is an intuitive belief (well



Figure 3.1: Broad classification of probabilistic broadcats schemes, including a list of trade-offs that are common to all schemes. In the figure, *Rt* stands for *Route* and *A.I.* stands for *Artificial Intelligence*.

supported by theoretical grounding) in which nodes in populated regions of a network should forward broadcast packets with low probability, whereas nodes in sparse regions should forward packets with higher probability. Second, when considering information of the two-hop neighborhood centered at the forwarding node, it seems that the opposite principle holds, namely a node with more neighbors than the other nodes nearby should be given higher priority (i.e. a higher forwarding probability and a shorter forwarding delay). Figure 3.2 includes two representative studies for each of the allegedly conflicting principles, namely the *Controlled Gossip* and *Rapid* schemes represent the more intuitive principle whereas the *LPR* and *NCPR* schemes support the ideas of the less-intuitive apporach.

These two principles, and their detailed characteristics, some of which are explicitly stated in Figure 3.2, are explained in the following two sections in greater detail. For the moment, the figure suffices to illustrate that the ideas behind probabilistic broadcast schemes using information about the one-hop and two-hop neighborhood can be very different.

3.1.1 One-hop neighborhood

Neighbor information has been used to allow nodes to calculate their forwarding probability in different ways. However, the most widely accepted, intuitive idea states that nodes in a highly populated area of the network (i.e. nodes with many neighbors), should forward broadcast packets with low probability, while nodes with very few neighbors should forward the packets with high probability. The theoretical support for this idea can be found in [1] (a paper called *Controlled Gossip*), in which the authors build upon the theory of Random Plane Graphs [2] to prove that, in a network with nodes placed according to a



Figure 3.2: Two fundamental ideas behind neighbor-based probabilistic broadcast schemes.

two-dimensional Random Uniform distribution, the forwarding probability at every node can be set to the inverse of the node's number of neighbors, namely $p_i = \beta/n_i$, where p_i and n_i are the forwarding probability (p_f) and number of one-hop neighbors (i.e. degree) of the *i*-th node, respectively, and β is a constant that determines the *coverage level* (i.e. the average fraction of nodes receiving each broadcast packet: the term coined *reachabil*ity). The work in [1] provides theoretical and simulation-based evidence supporting the aforementioned intuitive idea. What may be argued about the analysis in [1] is whether a universal value for β can be found. The authors acknowledge that there is no known analytical expression for β as a function of the coverage level, and therefore, since the theory applies to an asymptotic behavior where the number of nodes in the network tends to infinity, a simulation-based estimation of the behavior of β is necessary for finite networks of interest. At this point, it is necessary to mention that all nodes are assumed to have the same transmission range, and, according to the theory, the value of β is the minimum average degree (i.e. number of nodes within such homogeneous transmission range) that yields the expected coverage level (reachability). In other words, the theoretical grounding of the simple relation $p_i = \beta/n_i$ suggests that β exists as a function of the expected coverage level, as shown by simulation.

Also theoretically supporting the ideas of *Controlled Gossip*, the work in [11], [12] contains an analytical boundary on the probability of broadcast packets failing to reach any arbitrary node. Since nodes receive and forward packets independently, the probability of not reaching any node can be associated to the fraction of the network that is not reached by broadcast packets, using $p_i = \beta/n_i$ as the forwarding probability. The analytical boundary in [11], [12] suggests that if β is the amount of one-hop neighbors allowed to forward broadcast packets, a computation of worst-case reachability (maximum lack of

reachability) is immediate. In this regard, the function derived for the upper bound on broadcast failure plays the role of the unknown analytical expression for β (mentioned in [1]) as a function of the so called coverage level (i.e. reachability). This is to say that β (called *reliability factor* in [11], [12]) is equivalent to the constant that produces a specific coverage level in [1].

Perhaps the best coincidence of the analyses in [1] and [11], [12] is that the use of a broadcast probability inverse to the number of one-hop neighbors scales well with the number of nodes and depends mainly on the size of the area where nodes reside. Specifically, the theory proves that for networks of a fixed area, the expected number of transmissions, in the case of a successful broadcast process, does not depend on the number of network nodes, but rather on the network area (i.e. the number of transmissions is constant regardless of the number of nodes and node density).

A topological interpretation of this mapping between the analyses from [11], [12], and [1] can be seen in Figure 3.3. In the figure, the nodes that decided to forward the broadcast packet according to $p_i = \beta/n_i$ are represented by dots, whereas the remaining fraction of nodes (not forwarding the packet) are represented by little crosses. In the figure, as long as the dot-nodes form a connected component that includes a specific fraction of the dot-node set (i.e. a particular coverage level or reachability), the whole network (dot-nodes and cross-nodes) will receive the broadcast packet with the same coverage level. What is more important is that the average number of forwarding transmissions (i.e. number of dot-nodes) should remain constant even if the number of cross nodes grows arbitrarily within the same area (see Figure 3.3b. The previous observation further supports the claims in [12] suggesting that this scheme easily overcomes node failure and/or selfish behavior of some nodes, as long as the percentage of selfish/failing nodes does not affect the connectivity of dot-nodes. Notice that Figure 3.3 shows a single realization of the random variables that selected the dot-nodes. However, from theory and simulations, it has been shown that this behavior holds for a fraction of realizations at least equal to the coverage level (reachability) of successful broadcast. Therefore, with different random sets of dot-nodes, a high percentage of broadcast packets reaches most of the network nodes, and only occasionally a broadcast packet reaches only a few nodes. This behavior is also called *bi-modal* and is associated to the theory of percolation and phase transition phenomena [4], [2].

Another positive coincidence in the analytical studies with forwarding probability $p_i = \beta/n_i$ is that the different authors seem to be well aware of the connectivity conditions for their simulation settings (the Critical Transmission Range presented in [3]) and so, based on their analyses, it should be possible to calculate some minimum feasibility conditions that must be met to successfully implement the formula $p_i = \beta/n_i$ as a valid probabilistic broadcasting technique.

Other studies working on exactly the same idea, although without the rigor of theoretical analysis, used simulations to establish the value of β when computing $p_i = \beta/n_i$ at every node. For example, in [13], simulation results show that a value of β around 9 offers good performance, namely reducing the amount of forwarding events (Saved Rebroadcast) while reaching almost all nodes (Reachability) in a variety of network settings.

Continuing with the same idea of adjusting the forwarding probability to the inverse of the number of one-hop neigbors, several authors have proposed different heuristic functions to calculate p_i . The exact functions to obtain the value of p_i range from threshold functions



Figure 3.3: Topological view of the agreement between the ideas in [1] and [11], [12]; β corresponds either to the fraction of nodes that are dot-nodes or to the minimum degree that guarantees a particular reachability in a graph made of dot-nodes only. (a) sparse network leads to a higher value of p_f ; (b) denser network leads to a much lower value of p_f keeping the average number of dot-nodes constant.

on the number of neighbors, to the use of the geometric series [14] and even to the adoption of an *infection rate* metric from other problem domains [15]. To focus on the comparative aspects of the proposals, rather than on the heuristic details that motivated the choices of every author, Figure 3.4 summarizes the behavior of probability as a function of the number of one-hop neighbors (i.e. node density).



Figure 3.4: Comparison of the value of probability as a function of the number of one-hop neighbors in Gossip2 [16], BNR mode2 [13], Hyper Gossip [15], Adjusted Flooding [17], Rapid [12], Dynamic Probability [14], 2P and 3P [18], AP algorithm [19].

The values of thresholds and/or constants involved in the curves depicted in Figure 3.4 are those identified by each author to give the best performance. In this regard, the curves provide a *fair* comparison by selecting the probability functions that provided the best results in each case.

From Figure 3.4, it is easy to identify which proposals were based on thresholds since their functions simply fall showing a staircase shape. A salient feature of the figure is that, for most schemes, whenever a node has more than 10 one-hop neighbors, forwarding probabilities fall below 0.5, as if the value of 10 neighbors could be generalized to represent the case of a dense neighborhood. Another interesting coincidence in the figure is that the curve for the *BNR mode2* scheme follows very closely the trend of the *Hyper Gossip* curve, even though these two curves have very different heuristic motivations behind their calculations. Overall, despite the common trends observed in Figure 3.4, it can be stated that no definite consensus about the function $p_i(n_i) \propto 1/n_i$ has been achieved. However, the curves suggest there may be a unifying criteria to find the best probability function, perhaps based on network connectivity theory and not in extensive simulation studies.

3.1.2 Two-hop neighborhood

As mentioned above, when considering the two-hop neighborhood of nodes, the literature suggests that nodes with more neighbors should be given higher priority, namely a higher forwarding probability and shorter transmission delays. As shown in Figure 3.2, schemes like LPR [20] and NCPR [21] support this idea. In LPR, only analytical results are provided to indicate that the nodes of an ad hoc network can be easily organized into four groups (levels), and so the nodes in each group should forward broadcast packets with a predefined probability. By exchanging Hello packets, every node decides which group it belongs to by comparing its number of neighbors with the number of neighbors of the other nodes nearby. If a node has more neighbors than all other nodes in its neighborhood, it forwards broadcast packets with the highest probability (the highest level with $p_f = 1$, referred to as *definitely forwarding nodes*); conversely, if a node has less neighbors than all of the others in the one-hop neighborhood, it belongs to the lowest level (level four with $p_f = 0$, referred to as *silent nodes*). Nodes in the second level must have more neighbors than half their nearby nodes (called *forwarding nodes with high probability*). Finally, the third level includes the remaining nodes (called forwarding nodes with low probability). Based on the analytical curves, the authors show that nodes in the highest and lowest levels are very infrequent, and also that nodes in the second and third levels are evenly distributed. Consequently, this organization of nodes is expected to form a spontaneous backbone structure, similar to the backbone structures of the deterministic approaches, but with a much simpler formation algorithm based on probability. One of the main advantages of implementing this scheme is that the information required in Hello packets only includes the number of nodes (not their identity, or their neighbors list) and so Hello packets can be kept short, which is convenient in mobile scenarios due to the reduced overhead.

In the *NCPR* scheme, broadcast packets must include a neighbor list so that a receiving node knows which of its neighbors have not been covered. Then the proportion of uncovered neighbors and the number of one-hop neighbors determine the value of forwarding probability. Higer priority is given to nodes with more neighbors in common by assigning lower delays to their transmissions. Although it can be stated that redundancy increases because this scheme appears to be delivering more copies to the same nodes (i.e. nodes in common), the authors argue that, in this way, more nodes can exploit the neighbor knowledge to adjust their *Uncovered Neighbor Sets*. Considering the previous argument, it would be interesting to have an analytical perspective of this idea, as in [20]. Other studies that have considered two-hop information to compute the forwarding probability of nodes have used the ratio between the number of one-hop neighbors and the number of two-hop neighbors. For example, denoting $n_{i(2h)}$ as the number of two-hop neighbors and n_i as the number of one-hop neighbors of node i, simulation results are shown in [22] using different formulas for the values of the forwarding probability (p_f) : either $p_i = n_i/(n_i + n_{i(2h)})$ or $p_i = n_{i(2h)}/(n_i + n_{i(2h)})$. In [23], the forwarding probability values (p_f) vary by adding or subtracting a fraction of probability determined by the hopby-hop variations of a so called *Expansion Metric (EM)*, which is equal to $EM = n_{i(2h)}/n_i$. It is interesting to see that in [24] the value of p_f is computed as the average of the different estimations of p_f found in [22]. This average is simply applied in the context of VANETs and it produces good results even though there is no analytical rationale to support the success of this approach.

On the other hand, proposals that require neighbors identities (e.g. nodes IDs such as a list of IP or MAC addressess) can be further divided into schemes that simply compare nodes IDs to verify the set of uncovered nodes and schemes that establish more elaborate relations among neighbor nodes. Examples of ID verification for uncovered nodes include [25] and [26] (scheme NCPF only). The more elaborate schemes found in [27] and [28] establish very precise relations among nodes in which forwarding probabilities depend even on the ID of the source node and the expected reliability associated to application-layer requirements. In these studies, the interactions between the nodes and their two-hop neighbors are tailored to suit the needs of WSN applications and therefore performance may be affected when applying the same principles in other contexts. Such is the case of *Smart Gossip* [28], which seems to perform very well in WSN, with limited mobility and a reduced number of broadcast sources, but, when tested in a VANET scenario [29], [30], major adjustments have to be incorporated to obtain acceptable performance.

In summary, the schemes that gather nodes IDs and perform more complex comparison processes exhibit longer delays and can be significantly affected by mobility. From the experience gathered in the work with two-hop neighbor information, it appears that more analytical work is necessary to have a better insight into the proper use of two-hop neighborhood information, especially to find the simplest scheme that will surely adapt to a wider range of ad hoc scenarios.

3.1.3 Timeline of neighbor-based schemes

Figure 3.5 shows all neighbor-based studies (schemes) reported in [10] published in the period 2001–2013. The studies at the bottom of the figure use the relation f(n) to reffer to the schemes that used one-hop neighbors information to compute the forwarding probability. The studies at the top of the figure use the relation f(2n) to represent the proposal in which the forwarding probability was computed using information about two-hop neighbors.

Some of the schemes compute the forwarding probability as a function of two variables. In the case of one-hop schemes, the relation f(n, c) means that the one-hop neighbor information is complemented by counting the number of copies of the same packet received by the nodes. Specifically, nodes should *always* forward packets if there is no reception of an additional copy, even if the probabilistic decision had initially prevented the node from forwarding. In the case of two-hop neighbor-based schemes, the second variable, N,



Figure 3.5: Chronological diagram of all neighbor-based schemes reported in [10]. The convention f(n) means that the forwarding probability is obtained from information about one-hop neighbors whereas f(2n) represents forwarding probabilities derived from information of two-hop neighbors.

refers to knowledge about the network size (i.e. the total number of nodes in the network); these studies serve as benchmark, but their dependence on global information makes them impactical for ad-hoc broadcast applications.

Figure 3.5 shows that there are twice as many proposals using one-hop-neighbor information. However, one-hop neighbor-based schemes were last proposed in 2011, suggesting that the idea was extensively explored, which made authors look for other type of probabilistic schemes.

3.2 Area-based probabilistic broadcast

Area-based probabilistic broadcast has its foundations in the seminal analysis of the Broadcast Storm Problem [8]. The analysis showed that, upon receiving a broadcast packet, the maximum additional area possibly covered by a packet retransmission (i.e. a forwarding event) is approximately $0.61\pi r^2$, and, if all possible locations for potential receivers were occupied with equal probability, in the limit, the average value of such additional area is approximately $0.41\pi r^2$. Figure 3.6 clearly illustrates the three predominant metrics considered by authors to calculate an appropriate forwarding probability (p_f) , always aiming at avoiding the redundancy caused by the nodes for which the expected additional coverage area is small. The first idea (shown at the top of Figure 3.6) suggests that if the coordinates of every node are known, it is possible to calculate the exact additional area (shaded area in the figure). However, as the exact computation of many intersections among several circles can be cumbersome, the fact that a transmitting node lies inside/outside the convex polygon formed by the coordinates of other nodes can be used as a quicker estimation of how much additional area will be covered. The second idea in Figure 3.6 shows that the additional coverage area when forwarding a broadcast packet is proportional to the distance from the broadcast source. However, as suggested in the figure, such an estimation fails to consider that additional areas may spread on different directions, and the view of such a distance estimation is made as though transmitting nodes were lying along a straight line. Finally, the figure shows the case in which the number of nodes within the surrounding area of transmission is known. The example depicts a simple case in which 10 nodes lie in the white area (as neighbors of the node in the center of the white circle) and 6 nodes lie in the additional coverage area (gray area also related to a central node). By simply dividing the number of nodes in the additional coverage area into the number of nodes in the white area, a good approximation of the 0.61 fraction (discussed above) woud be possible. In this last case, it is very important to notice how crucial a uniform distribution of the nodes is, otherwise the estimation of the additional coverage area can be very inaccurate.



Figure 3.6: Fundamental principle of area-based probabilistic broadcast. Probability is proportional to the additional coverage area (gray area), and such an area is known to be at most 0.61 the area of the original circle, and 0.41 on average. Three common types of additional converage area estimation are depicted.

What is common to all the studies categorized in this section is the use of a function that relates the forwarding probability to the decision metrics (either the estimated distance or area). Regardless of the scheme, the values for maximum distance and maximum additional area are known, namely r (the transmission range) and $0.61\pi r^2$. Hence, every function used by the authors of area-based schemes to calculate the forwarding probability of nodes spans the domain [0, 1], which represents the fraction of maximum uncovered area or maximum distance from the source. Additionally, different authors [26], [13] agree on using an exponent to shape the functions that yield p_f and thus tailor their approaches to suit particular needs. Figure 3.7 shows four examples of functions that yield a higher value of probability when the expected additional area or distance from the source is larger. The three monotone increasing functions represent the crucial role of the aforementioned exponent. Specifically, assuming k as an arbitrary exponent of any forwarding probability function (i.e. p_f as a function of additional area), it is easy to see that, for k > 1, significantly higher probabilities should be assigned to the nodes located near the border of the transmission range (r), i.e. nodes expected to cover more additional area. This positive exponential shape is suggested by most authors to increase the coverage area on each hop, finding shorter routes and reducing latency. However, authors also warn that, if k is too large, the strategy fails because almost all nodes forward with very low probability and the small fraction of nodes right at the border, forwarding with high probability, are not enough to maintain connectivity in the network.

In Figure 3.7, the decreasing curve represents the case in which probability is a function of the number of packet copies received by an arbitrary node. This curve is equivalent to the others because, as shown in [8] (also explained below), the expected additional area shrinks as the number of packet copies grows, therefore the decreasing curve in 3.7 can be seen as equivalent to the curve for k > 1 but with the x axis inverted.



Figure 3.7: Four functions that assign higher probability to higher values of the estimated area/distance. The decreasing function represents the proposals where the area is estimated by the number of copies of the same broadcast packet that nodes receive.

After describing the concepts that underpin the mainstream of area-based probabilistic broadcast schemes, the following sub-sections are intended to classify specific studies according to their central metric for area estimation (e.g. by estimating Euclidean distances, the number of nodes in the transmission area, or the number of packet copies).

3.2.1 Area Estimation by Number of Nodes

The number of neighbors to estimate the additional coverage area was used in [13] (mode3 and mode4) exactly as depicted in Figure 3.6, namely p_f was estimated as the ratio between the number of nodes in the new coverage area and the number of nodes in the source transmission area. Additionally, an exponent is applied to the calculated ratio, like the exponent k presented above (called coefficinet of convexity by the authors). The results in [13] are particularly valuable because the authors compare this strategy with a neighbor-based strategy and also with a hybrid improvement to the area-based solution that includes counting packet copies as a back-up mechanism to guarantee high reachability of every broadcast packet.

An interesting approach that also uses estimation of the additional coverage area based on the number of neighbors can be found in [31], where an upper bound on the uncovered area is taken as the metric to decide the expected number of retrasmissions within the range of the transmitting node. Figure 3.8 shows the upper bound for the case of four expected retransmissions (forwarding events). With such an estimation, the forwarding probability value (p_f) corresponds to the ratio between the number of expected retransmissions and the actual number of neighbors. Something peculiar in this approach is that the forwarding probability is *designated* by the source and not decided by the forwarding nodes, that is, in Figure 3.8, the node in the center decides the forwarding probability of the receivers. Then, the forwarding probability value is carried in the header of the broadcast packet and is computed by the preceeding node, not by the actual forwarding node. This subtle difference with most schemes can also be observed in [16] (*Gossip2*) and it is not very common. It might be interesting to find out whether a *designated* probability is better than a *self-decided* probability to adapt the ad hoc broadcast process to the variations in node density found in realistic ad hoc networks.



Figure 3.8: Proposed geometrical upperbound to determine the number of retransmissions and thus the forwarding probability of nodes according to [31]. In this ideal example, the transmitting node (in the center) determines that four nodes (on average) should forward the broadcast packet.

A more recent contribution to the estimation of uncovered areas based on the number of nodes is presented in [32], in which a metric of similarity that is used in the statistical analysis of data clustering and linear regressions, namely the *Jaccard Distance* between two data vectors, was shown to be correlated to the Euclidean distance among nodes in an ad hoc network. This idea fits perfectly with the proposals in [13] and represents a motivation to use the Jaccard Distance in situations in which no location service is available to estimate Euclidean distances.

3.2.2 Area Estimation by Distance Estimation

The studies that use estimations of the Euclidean distance to determine forwarding probabilities can be further classified as *traditional* and *routing-oriented*. The traditional proposals are those that follow the principles of area estimation presented in Figure 3.6. In these proposals, authors also include exponential shape factors to compute the values of p_f , as explained in Figure 3.7. For example, in [26] and [33], exponential functions of the form $p = e^{k(d/r)}$ are used to compute the value of p_f as a function of distance, where d/r corresponds to the fraction of the maximum distance (or maximum coverage area) associated with consecutive forwarders. What makes these schemes different is the complementary strategies added by authors to fulfill the concerns of each study. For example, in [26], the probabilistic scheme is complemented by verification of covered nodes (as in the neighbor-based schemes), whereas in [33] the scheme is complemented by using a different function to compute p_f depending on whether the forwarding decision is prompted by a packet received for the first time or not. Also in [33], the authors use a deterministic delay that prioritizes the transmissions of nodes separated by larger relative distances. Other more traditional proposals make use of a simple linear function to compute p_f as a function of distance (area) [34], [35]. In [34], the main contribution is to show that cross-layer capabilities make it possible for probabilistic broadcast schemes to use the received signal power information from the MAC layer to estimate relative distances between nodes and to avoid the use of location services. In [35], the main contribution lies in showing how serious the broadcast storm problem is in VANETs by using a case study of a four-lane highway scenario. In the VANET case study, programmed delays also play an important role (as in [33]).

The less traditional, routing-oriented schemes take advantage of the relative distances between nodes to produce a directed broadcast wave that propagates in the direction of a particular destination node. In [36], a location service is assumed (e.g. a GPS service) to provide nodes with their coordinates. Nodes use distance information to decide whether or not to participate in the broadcast process. Nodes will forward broadcast packets only if they belong to the ellipse whose foci are the coordinates of the source and destination nodes. Based on properties of ellipse construction, a node that receives *routing* packets towards a particular destination checks its distance to both the source and destination nodes, and if the sum of the distances is less than the ellipse factor, the node forwards the packet with a high probability value (typically p = 0.7); otherwise, the node discards the packet. This scheme keeps the broadcast process concentrated in an ellipsoidal region between source and destination, saving additional efforts (in processing and forwarding) for the nodes far away from the line between sender and receiver. Although the assumption of nodes knowing the location of the destination is the most difficult to satisfy, the authors argue that this scheme can be used in a WSN collecting data, a procedure in which the destination node is often fixed. In this context, the main concern is the communication between pairs of nodes, rather than network-wide dissemination.

Another *routing-oriented* probabilistic broadcast scheme that uses distance estimations to make forwarding decisions can be found in [37], [38]. In these studies, the authors take advantage of a metric called the hint of a node [39]. Such hints capture historical information about the connectivity of nodes, which can be interpreted as a very good estimation of how far from a point a particular node is. Specifically, all nodes use a periodic beacon (i.e. short Hello packets) so that nearby nodes can record the presence of each other wherever they go. Therefore, after a while, mobile nodes can estimate their relative distance to a specific node using the records on the last time they *heard* a beacon from that node as well as on the length of the last connectivity event. These records (*hints*) are like an indicator of how quickly a node left certain area, and therefore serve to estimate which nodes are farther away from a sender than other nodes. However, the most appealing feature of the work in [37], [38] is not the use of *hints* per se, but the way authors take advantage of the error rate of hint-based estimations to achieve a directional broadcast wave towards the destination (as in [36]). The main idea is that nodes should forward a *routing* packet only if they estimate that they are closer to the destination than the sender. However, if the estimation is wrong (e.g. with probability q = 1 - p), this is equivalent to having a probabilistic broadcast scheme forwarding packets with probability p only in the direction of the receiver. Hence, if p is large enough, successful probabilistic (and directed) broadcast will dominate the route finding process in the network. What is interesting of this solution is that it poses the forwarding probability as an *implicit* feature of the system (the error rate of distance estimations) and at the same time reduces the amount of packets that are unnecessarily propagated away from route paths without the need for location systems in mobile scenarios.

In summary, distance estimation schemes are mostly dependent on location services (with very few exceptions). This trend is justified by the lower delays in the overall performance of broadcast, which is very useful in the context of VANETs. However, more analytical studies might be convenient to reduce the excessive calibration efforts necessary to adapt schemes to the specific application scenarios (e.g. WSN vs. VANET).

3.2.3 Area Estimation by counting copies

Counting the number of copies of broadcast packets has been the technique to prevent broadcast propagation from dying out early due to an unlucky realization of the random variables that drive forwarding decisions. Similar to the case of distance estimation, it has been shown that the additional coverage area of a forwarding event is drastically reduced as the number of packet copies *heard* by the forwarding node increases. Figure 3.9 shows a topological representation of the simulation results first introduced in [8], in which the expected additional area covered by a potential transmitter (the gray node in the figure) drops from 0.41%, when hearing one packet copy, to less than 5% (on average) after receiving four copies of the broadcast packet.



Figure 3.9: Drastic reduction of the average additional coverage (avAC, shaded regions) when forwarding a packet as a function of the number of *listenable* nodes from which copies of the same broadcast packet can be received (black nodes): (a) receiving one copy, (b) receiving two copies (c) receiving three copies (d) receiving four copies.

The seminal work in [16] (Gossip3) provided the basic counting scheme in which the nodes that refrain from forwarding (due to a first realization of the random variable determined by p_f) should wait for a random period of time (known in the literature as Random Assessment Delay or simply RAD). Upon expiration of RAD, the packet is forwarded deterministically if no additional copies are received. A slightly different idea is proposed in [40] (PCBR scheme), but unlike Gossip3, forwarding events are always delayed. This subtle difference may lead to a better-informed forwarding decision but at the cost of longer delays.

3.2.4 Timeline of Area-based schemes

Figure 3.10 shows a chronological diagram including the proposals that compute the forwarding probability of nodes from area estimations published in the period 2002–2014 (proposals described in [10]). Mainly, the schemes focus on the estimations of the distance between nodes, d, to find the forwarding probability of nodes, that is p = f(d). The notation $f(\cdot)$, showing d as the first variable (i.e. $f(d, A, \ldots)$), indicates that the scheme estimates the areas using an estimation of the distance. Conversely, the notation in which A appears as the first variable (i.e. $f(A, d, \ldots)$) denotes studies which attempt to estimate areas using the coordinates of the nodes; so these schemes use estimation of the distances based on coordinates (obtained from location services) in order to compute estimations of the additional coverage area for finding the forwarding probability value.



Figure 3.10: Chronological diagram of the schemes that use estimations of the distance between nodes and the additional coverage area of transmissions to find the forwarding probability.

Other variables used by area-based schemes include the Global Positioning Systems (GPS) value, the distance estimated using the Received Signal Strength (RSS) indicator of the data-link layer, the use of probability functions of distance between nodes (proDist in the diagram), additional control packets (CtrPack) and Angle-of-Arrival (AoA) estimations. Other additional variables to determine the forwarding probability values in area-based schemes include the number of copies (c), two-hop (2n) and one-hop (n) neighbor information as well as the total number of nodes in the network (N).

It is important to notice that the schemes at the top of Figure 3.10 use several variables (hybrid schemes) to come up with the value of the forwarding probability, and most of these schemes assume a GPS service. These are the schemes intended for vehicular networks (VANETs). In vehicular networks, the battery of mobile nodes is not an issue and vehicles are commonly assited by road infrastructure; hence, the proposals in this category tend to use a lot of resources to provide robust broadcasting schemes at the cost of protocol complexity. The schemes at the bottom of Figure 3.10 are a lot simpler and can be applied to mobile and sensor networks.

Figure 3.11 shows the time diagram of the studies that focused on the copy-counting principle only. Since these sheemes rely heavily on random delays, they have been associated to larger values of end-to-end broadcast latency and so authors prefer not to use them for vehicular networks. The notation f(c) means that the number of copies of a broadcast packet is the only variable necessary for computing the forwarding probability. The variable called Colors indicates that the concept of colors, from graph theory, was used to produce communication graphs with different properties; meanwhile, the notation noRAD represents a study in which there is no use of Random Assessment Delays despite being a copy-counting scheme.



Figure 3.11: Chronological diagram of area-based schemes that rely only on the number of packet copies received by nodes.

As can be seen in Figures 3.10 and 3.11, area-based schemes based on distance estimations constitute most of the proposals, mainly applied to vehicular networks.

3.3 Alternative probabilistic broadcast schemes

A small set of studies can be highlighted for their unusual, yet innovative way of approaching the problem of finding the value of the forwarding probability (p_f) for nodes during the ad hoc broadcast process. As pointed out in Figure 3.1, *alternative* proposals bring ideas from Game Theory, Artificial Intelligence optimization, energy conservation, the use of Directional Antennas, and broadcast restarting processes (i.e. retrasmissions) to solve the problems of redundancy and acceptable reachability levels that pertain to ad hoc probabilistic broadcast.

In [41], Game Theory is used to derive the function that yields p_f . In this work the authors combine the concept of Nash Equilibrium with a recommended value of the minimum node degree that guarantees network connectivity in order to derive the optimal values for the *Gain Factor* of a game (a game played among the neighbors of a broadcast transmitter/forwarder upon receiving a broadcast packet). In this regard, the problem translates from finding the optimal value of p_f to finding the optimal *Gain Factor*. Nonetheless, the most interesting contribution of this paper is that, by using a completely different paradigm (Game Theory on its own), the authors produce very flexible functions to compute p_f that resemble the functions depicted in Figure 3.4 (i.e. p_f is the inverse to the number of one-hop neigbors), yet with an asymptotic behavior that preserves connectivity. However, the paper focuses only on the routing capabilities of the Game Theoretical scheme and not as much on the resulting broadcast performance per se. Of course, great advantages are observed for very dense networks (as in the other schemes), hence it would be interesting to study this proposal in more detail and look at its broadcasting properties alone, perhaps in sparser networks or including node mobility.

In [42] and [43], Genetic Algorithms are applied to specific ad hoc scenarios. The idea is to monitor the *output* of the simulation environment (many times) to obtain its statistical behavior. The *output* of each set of simulation replications is mapped to objective functions for optimization. Objective functions must be maximized/minimized according to the need of the specific target application. For example, in [42], three conflicting objectives, namely the Number of Collisions, the Propagation Time and the Number of Transmissions during the simulation are optimized by adjusting input variables such as the forwarding probability of nodes (p_f) , the number of times each node retransmits a packet, the delay between two successive retransmissions, and the TTL of broadcast packets. In this work, the target scenarios correspond to five VANET situations ranging from cars traveling in a very sparse rural road to dense urban traffic. The work in [43] is an equivalent version of [42], but using different evolutionary algorithms and applied to a completely different scenario, namely a disaster zone divided into specific context-based areas. The three conflicting objectives for optimization in this work are very similar to those in [42], but the variables adjusted to produce optimal behavior are different, namely a dissimilarity metric (the *Jaccard Distance* discussed in Section 3.2.1), the exponent on the function that yields p_f (as explained in Section 3.2), and a coefficient that controls the impact of the dissimilarity metric on the Random Assessment Delay of broadcast packets. The most remarkable contribution of these studies is that the authors offer a methodology for decision makers to conveniently obtain a tailored performance of an ad hoc network in very specific target scenarios based on optimal broadcast performance.

Another alternative approach to probabilistic broadcast can be found in [7], where the authors provide a very accurate mapping of the broadcast process with onmidirectional and directional antennas to the site and bond percolation models [4], respectively. What is unique about this work is that there is a clear motivation to reduce the number of duplicate packets *received* (not only sent) by nodes, while achieving the same broadcast coverage (Reachability). This work offers a wide range of experiments on how the forwarding probability value changes according to the specific geometrical features of node placement (e.g. Square, Triangular and Hexagonal grids are compared to their theoretically expected behavior and experiments are extended to scenarios with random node placement). The authors combine directional antennas with counter-based, neighbor-based and hybrid approaches and agree that their directional-antenna schemes may not work with mobile nodes.

More alternative work on probabilistic broadcast has included the level of remaining energy in nodes to adjust their forwarding probability [44]. The scheme is basically a distance-based scheme aided by a counter-based mechanism to prevent network partitioning (see Section 3.2.3). Although in this work the authors explicitly integrate the conditions under which a node refrains from forwarding based on its remaining energy and a local estimation of the average remaining energy in its neighborhood, the idea of including energy constraints in computing p_f is not exclusive of this study. For example, the authors in [26] had already pointed out that the exponent of the functions for computing p_f (shape factor in Section 3.2) could be modified by a constant (called *passivity parameter*) so that, as part of the broadcast protocol, nodes with lower levels of energy could self-adjust their functions to yield lower forwarding probability values and preserve their remaining energy. However, the proposal in [44] is valuable to extend the lifetime of WSNs.

Finally, the authors in [45] offer an analysis of the impact of retransmitting the same broadcast packet several times (broadcast restart) as an alternative to overcome the undesirable, but accepted, reachability limitations of probabilistic broadcast (i.e. unfeasible 100% reachability level). The study focuses on adjusting the retransmission timeout to an optimal value such that maximum reachability can be obtained. This technique takes advantage of the fact that the specific set of unreached nodes can vary from one broadcast process to the next. The experiments are validated in a testbed. The results reveal that an optimal timeout period depends on a wide range of parameters, namely the applicationdependent dead line imposed on broadcast response, the traffic conditions (number of broadcast sources) and the payload size of broadcast packets. Overall, the study clearly illustrates that the adjustment of the timeout duration in protocols is a sensitive variable for probabilistic broadcasting schemes.

3.3.1 Timeline of Alternative schemes

Figure 3.12 shows a chronological diagram of the schemes that find forwarding probabilities using alternative ideas, different than the conventional area-based or neighbor-based principles. Schemes that focus on the conservation of energy are grouped at the bottom of the figure; schemes using the convention f(Enrg).



Figure 3.12: Chronological diagram of probabilistic broadcasting schemes that propose alternative ideas to find the forwarding probability of nodes. The schemes that focus on saving the energy of nodes are grouped at the bottom of the diagram.

The schemes at the top of Figure 3.12 use various conventions that can be easily associated to the descriptions above. For example, schemes with the convention f(d, hints) and f(d, region) clearly refer to the *hint-based* and *regional gossip* studies presented in Section 3.2.2. Likewise, the convention f(speed) refers to the study in which the forwarding probability is a function of the speed of nodes and the remaining studies are those based on Genetic Algorithms and the supressed bond-site percolation model.

Figure 3.12 shows that the concern for energy in nodes, applied to probabilistic broadcast, appeared only until 2006, and the use of Genetic Algorithms to callibrate broadcasting protocols is a more recent idea. While energy conservation schemes are commonly intended for sensor networks, the other alternative schemes apply to mobile networks in general.

3.4 Recent Studies

Since probabilistic broadcasting is an active research area, several studies have been published in the period 2014 - 2019. Recent publications address topics such as congestion in the wireless channel, routing, energy conservation, evaluation in VANET scenarios, and analytical modeling. This section summarizes the details of every proposal.

3.4.1 Congestion Awareness in Studies

In this set of studies, authors show how the conditions of the wireless channel (e.g. noise and co-channel interference) affect the performance of probabilistic broadcasting schemes.

In [46], authors combine a neighbor-based scheme and a copy-counting scheme. Three values of the forwarding probability are used by nodes depending on the number of heighbors (as in [18]). Additionally, nodes use RAD (Random Assessment Delay) to count packet copies. If the number of copies is less than a threshold, a packet must be forwarded regardless of the probabilistic forwarding decision. Congestion is considered by adapting the duration of RAD as a function of the average number of packets received per second at each node. Although authors claim that the counting threshold depends on the number of neighbors, the relation between these two variables is not clearly stated.

A study dedicated to the analysis of the effects of congestion is presented in [47,48]. The authors propose the use of mediation analysis, a technique used in social sciences, to evaluate the effects of congestion (the mediator variable) in the relationship between a dependent variable (such as reachability or the packet delivery ratio on the application layer) and an independent variable (the forwarding probability, p_f). The results indicate that the values of p_f that increase rechability have a negative impact on application layer metrics. Therefore, reductions in reachability result in better performance on the application layer as congestion increases.

The number of one-hop neighbors and the resulting channel conditions, such as thermal noise and co-channel interference, are considered in [49] to determine the forwarding probablity of nodes. Based on previous information about the Packet Error Rate (PER), nodes that receive a broadcast packet with no error under harsh conditions should forward the packet with high probability since it is likely that other forwarders received a corrupted packet. Likewise, when noise and interference are low, forwarding probability values should decrease given that many forwarders received the packet successfully and are ready to forward it. This scheme has been evaluated [50] in the context of Low Rate WSN (LR-WSN) for static nodes using the Zigbee protocol suite. The analysis of thermal noise and co-channel interference is also used in [51] to show that noise and interference act as natural limiters of channel congestion. When the value of p_f increases arbitrarily, the total number of broadcast packets in the network does not increase as much since the resulting increase in interference leads to higher rates of packet loss.

Packet Error Rate (PER) is also considered in [52] to compute the forwarding probability of broadcast packets. In this study, the value of PER is associated to the distance between nodes on a per-hop basis. To avoid congestion of the wireless channel in the network, the forwarding probability decreases with the number of hops. This scheme is justified for VANET scenarios since the importance of traffic information for vehicles is dependent on how far the recipient is from the source. A co-channel interference analysis is provided in [53] to show that the geometry of node layout has a considerable impact on reachability. The analysis indicates that nodes placed at random facilitate the propagation of packets by providing more stable values of average Signal to Interference Ratio, which results in higher reachability levels than those obtained with node layouts that resemble grids.

3.4.2 Routing-oriented Studies

The studies considered in this section focus on the route-finding functionality of broadcast and relax reachability requirements. These studies stem from the idea that a route to a destination of a data packet can be found with no need for perfect dissemination of route request packets. What is important is the choice of stable routes once they are found. Most of these proposals take the two-hop neighbor-based scheme proposed in [21] as benchmark.

In [54], the forwarding probability is computed as the product of a link stability metric (the ratio between the estimated connection time and the residual time if a link) and the expected neighbor coverage. The analysis is based on finding the subset of neighbors with probability of communicating higher than a threshold, a subset called Predictable Rendezvous Nodes (PRN). The links established between a node and its PRN set are considered as stable links, resulting in routes that last longer and require less broadcasting of route request messages.

Routing is improved in [55] by applying Particle Swarm Optimization (PSO). As the scheme proposed in [21] (NCPR scheme) is being used to disseminate route request packets, an iterative process based on PSO selects optimal routes. The results show a reduction in routing overhead and end-to-end delay.

A combination of the Load Balanced Routing (LBR) protocol (an existing AODVbased protocol) and NCPR [21] is used in [56] to reduce routing overhead. Using a metric called Cumulative Active Path Count (CAPC – average number of active routes per node in the path), which is piggy-backed in route request packets, destination nodes take the decision about which path is to be selected for data transmission (the path with lowest CAPC). The routing overhead of this proposal is almost 30% lower than that of the original NCPR.

A cross-layer scheme is proposed in [57] to choose stable routes based on the values of the RSS indicator of the MAC layer. The forwarding probability is proportional to the RSS values. The authors consider that interference is not necessary for the analysis since its effects are implicitly reduced by selecting the links that offer better signal strength.

Similar to [52], the study in [58] proposes to reduce the forwarding probability as the distance (in hops) from the broadcast source increases. The routing-wise motivation of this scheme is the following: as route request packets propagate farther from the source, covering a larger fraction of the network, it is more likely that the destination was already found; then, route-finding packets become less necessary with distance. Inspired in wave equations for signal strength attenuation, the computation of the forwarding probability also considers the local density of neighbors on a per-hop basis, avoiding the propagation of packets from dying out early.

3.4.3 Studies on Energy Conservation

Studies that include energy conservation in their probabilistic broadcasting schemes use hybrid schemes (e.g. neighbor-based schemes combined with counter-based schemes). In these proposals the remaining energy in nodes affects the values of the forwarding probability.

Simulation tests are conducted in [59] to evaluate the energy savings in MANET when using different values of the forwarding probability. Nodes use AODV as their routing protocol and move according to the Random Way Point model. Energy consumption is evaluated as a function of the forwarding probability and the speed of nodes.

In [60], energy constraints are added to the NCPR scheme [21]. Apart from the forwarding probability, two conditions must be met to forward a broadcast packet: there must be sufficient energy (battery lifetime) in the node, and node density around the forwarded must exceed a specific threshold. This scheme is also evaluated using AODV as the routing protocol and the Random Way Point mobility model.

The hybrid scheme porposed in [61] combines the pure-probabilistic with the distancebased and counter-based schemes (integrated in AODV). Probabilistic broadcast is applied when a node receives the packet from a forwarder located closer than a distance threshold (otherwise $p_f = 1$). The objective is to prioritize the transmissions of nodes with larger inter-node distance. The number of packet copies is used as the exponent of the forwarding probability value obtain from the distance estimation. Therfore, the forwarding probability decreases rapidly with the number of packet copies.

The copy-counting technique is used in [62] to extend network lifetime. Nodes will drop packets when the number of copies is equal to their counting threshold. Nodes with more energy are called Dependable Nodes (DN) and use larger values on their copy-counting threshold. When the copy counter is less than the threshold, p_f is high, otherwise p_f is low. The energy level that determines the increase in the counting threshold of nodes is given by the average energy of the path form by previous forwarders. The energy level of a path is an average energy value computed by piggy-backing the values of the energy of nodes in route request packets. This scheme is modified in [63] to assign short delays with low counting thresholds to Dependable Nodes. In the new version of the broadcasting scheme, the values of p_f are recomputed during the delay using $p_f = p/C$, where C is the number of packet copies and p is the value of probability assigned to Dependable Nodes. Alternatively, the same authors proposed an energy-aware probabilistic broadcasting scheme that employs a Fuzzy Logic algorithm to compute p_f [64]. The membership functions and the set of rules are obtained from the number of 1-hop neighbors, the bandwidth available and the remaining energy at each node.

Probabilistic broadcast has also been integrated into a reactive routing protocol called DSR (Dynamic Source Routing) to balance the energy consumption of nodes [65]. In the route discovery process, intermediate nodes receive reply packets from neighbors with information about energy so that routes are established using the nodes that meet specific energy criteria, namely the nodes with an energy level near the mean remaining energy of their neighbors. The probabilistic scheme chosen for this study was NCPR [21].

3.4.4 Schemes applied to VANETs

In the context of VANETs, probabilistic broadcasting has been evaluated mainly in scenarios that resemble a highway. Most studies compare their proposals to distance-based schemes such as weighted p-persistent [35], *n*th-powered probabilistic broadcast [66] and Irresponsible Forwarding [67–69]. Only ocasionally, scenarios based on city roads are used. The main idea is to assess the effectiveness of probabilistic broadcasting to disseminate emergency messages within a target area. As in most proposals for VANETs, the following schemes assumed that Hello packets and beaconing messages maintain updated neighbor information, such as the ID, the locations, the speed of nodes and other data necessary for computing the forwarding probability.

By considering the maximum speed of vehicles in a highway, the forwarding probability in [70] is computed as the ratio of the node's speed and the maximum speed (i.e. $p_f = V/Vmax$). This scheme was evaluated assuming an error-free wireless channel with no initial contention phase, and the scenario corresponds to a 3-line highway. The forwarding probability is proportional to the speed of vehicles since speed is proportional to intervehicle distance.

The studies in [71,72] also use the speed of vehicles to compute the forwarding probability. By identifying node density as the factor that determines the critical value of the forwarding probability [1,6], the study presents a linear regression model that relates the varying vehicle density in highways (in Los Angeles) with the speed of vehicles. As a result, the forwarding probability of vehicles is $p_f = 0.22v + 0.042$.

A hybrid scheme for VANETs is introduced in [73], combining the distance-based and neighbor-based approach. Two values of the forwarding probability (i.e. $p_f = p_{high}$ and $p_f = p_{low}$) are assigned to the vehicles that meet specific conditions. To forward packets with $p_f = p_{low}$, vehicles must be close to the previous forwarder and they must have more neighbors than the average degree of the network. Vehicles at a distance larger than half the transmission range, and with less neighbors than the average degree, should forward packets with probability $p_f = p_{high}$. Vehicles that do not meet these conditions must drop the broadcast packets.

Several existing distance-based schemes [35, 66, 67] are adapted in [74] so that the use of Euclidean distances is avoided. Stemming from [32], it is shown that the similarity in the set of neighbors of two nodes is correlated to their Euclidean distance. Then, the use of similarity metrics between pairs of nodes can substitute the use of *estimations* of Euclidean distances, which can be misleading. Existing similarity metrics such as Jaccard, Dice, Kulczynsky, Folwkes-Mallows and Sokal-Sneath are used in the study. The use of similarity metrics is extended in [75] to produce new metrics obtained by applying genetic programming (GP). The new metric has a higher correlation with the Euclidean distance than the existing similarity metrics. The adapted schemes in [75] were evaluated using realistic VANET scenarios based on real city maps, whereas the evaluation of the schemes in [74] was conducted using the Manhattan mobility model.

The scheme proposed in [67–69] is evaluated in three urban scenarios [76], namely scenarios that resemble pedestrian mobility, pedestrian–vehicular mobility, and vehicular mobility. The evaluation is conducted from a routing perspective (focus on delay and throughput) by integrating the broadcasting scheme into the AODV protocol.

Two porposals for VANET have combined probabilistic broadcasting schemes with clustering [77, 78]. In [77], intermediate nodes use probabilistic broadcasting when requiring more than one hop to reach a cluster head. Nodes that are within the transmission range of cluster heads do not forward packets. The metric proposed to choose cluster heads is the product of the probability of successful transmission and the expected link connection time. In this scheme the traveling direction of vehicles is considered to increase link stablity so that links are formed between cluster heads traveling in the same direction. The forwarding probability is set to the inverse of the square of the number of packet copies (c): $p_f = 2/c^2$; a counter-based scheme. Conversely, in [78], cluster heads use probabilistic broadcasting. Cluster heads receiving the same transmission of a broadcast packet are numbered depending on their distance from the transmitter. The forwarding probability is computed based on the numbers assigned to cluster heads so that the farthest cluster head is always assigned the highest probability value regardless of its actual Euclidean distance. This makes the scheme robust against variability in node density and channel conditions. In both schemes [77,78] there is a concern about using stable links and cluster heads re-transmit packets several times to ensure reliability.

Similar to [41], a probabilistic broadcasting scheme based on Game Theory is proposed in [79] to determine the forwarding probability of nodes in VANETs. Unlike [41], this study uses the distance between nodes (instead of the number of neighbors and connectivity reasoning) to determine the cost and gain functions in the payoff matrix of the game. The results of the study focus on reducing redundancy (saved rebroadcast) and delay; no attention is devoted to reachability results.

3.4.5 Studies with Analytical Models

As ponted out in [10], analytical models for probabilistic broadcasting schemes are scarce in the literature. Therefore, the following studies are valuable to gain a better understanding of both previous and future proposals in the field.

The neighbor-based, self-pruning scheme in [21] is mathematically analyzed in [80,81] to quantify the forwarding probability (p_f) of a node located at a certain distance from its previous forwarder. The theoretical expression found for p_f requires nodes to know the distance to the previous forwarder, the average node density of the network and the transmission range. This computation of p_f is evaluated by simulation. Therefore, by using mathematical modeling, the original neighbor-based scheme in [21] can be implemented as a distance-based scheme. Simulations are conducted in C/C++ assuming an ideal MAC layer. The authors continue working on the analysis of their proposal to include a MAC layer such as the IEEE 802.11.

A remarkable proposal to improve the reachability of probabilistic broadcasting schemes makes use of multiple transmissions and coding-based redundancy [82]. Redundancy is introduced in the form of coded packets; that is, the source node has k broadcast packets to disseminate, which are encoded into $n \ge k$ coded packets, such that any k of these coded packets are sufficient to recover the original k data packets. The authors show the benefit in terms of a reduction in the overall number of transmissions needed for a successful dissemination of the original k packets. Although p_f can always decrease as n increases (reducing the set of forwarders per packet), arbitrarily large values of n counterbalance the benefit of coding by requiring more packets to be transmitted. Therefore, there is an optimal value of n that reduces the total number of packets transmitted in the network over time. Furthermore, the study points out that the success of packet coding in probabilistic broadcasting depends on the underlying network topology. With highly connected topologies (grids and random networks), multi-path diversity allows succesful dissemination. This idea is supported by analysing and simulating a network with a tree topology, in which the proposal failed.

3.4.6 Other Studies

The counter-based scheme is improved in [83] by comparing the inter-arrival time of packet copies with the interval in which Random Assessment Delays (RAD) are defined. A counter keeps track of the number of inter-arrival time values that are larger and smaller than RAD/2. At the end of RAD, a node that recorded more time values smaller than RAD/2 infers that it is in a dense area and chooses p_f randomly in the interval [0, 0.5]. Otherwise, the node infers that it is in a sparse area and p_f takes a random value in [0.5, 1]. Although this scheme relies on random delays, the actual value of RAD used in the simulations is not reported.

Another improvement to counter-based schemes is proposed in [84]. The scheme called $GOSSIP \ 3$ [6] is studied and its parameters are modified based on curve fitting techniques. Specifically, the value of the forwarding probability, which was constant in the original scheme, is now computed using a weibull function that was fitted from node-density data obtained after extensive simulation. Also, the counting threshold of $GOSSIP \ 3$ is modified according to node-density. It is worth mentioning that the authors use a technique based on the packet reception rate, seen by a node from its different neighbors, to estimate node density so that the node-density information required by the proposal does not imply the use of hello messages.

Regarding neighbor-based schemes, recent schemes continue to explore different functions to establish a relation between node density and the forwarding probability. The principle is still the same, namely assigning lower values of p_f to nodes in dense areas and increase the values of p_f for nodes in sparse areas. In [85], an exponential function of the form $p_f = 1 - e^c$ is used, where c is the fraction of uncovered neighbours (i.e. neighbors of the current forwarder that are not in the list of neighbors of the previous forwarder). In [86], however, a non-linear, square-root function is used to establish the relation between the number of neighbors and the forwarding probability. The authors define a metric called drop factor (DF) which is inversely proportional to the forwarding probability. The results show that the function defined for DF (i.e. an indirect calculation of p_f) reduces routing overhead, delay, and increases the data packet delivery ratio when comparing these metrics to the original version of AODV and a modified version of AODV based on NCPR [21] in dense networks.

As in [74, 75] for VANETs, in [87], similarity metrics are incorporated into existing schemes [13, 35, 67]. However, the adapted schemes are now evaluated in a testbed with more than 50 static nodes. In this study, the p-persistence scheme [35] based on similarity metrics showed the best results, (i.e. high reachability and reduced number of redundant packets).

Finally, probabilistic broadcasting has been applied to Autonomous Underwater Vehicle (AUV) networks [88]. Although probabilistic broadcasting performs well under the conditions of the underwater acoustic channel (low data rate, limited bandwidth, and large propagation delay), recent studies have shown that simple heuristics outperform probabilistic schemes when considering sparse AUV networks with non-isotropic wave propagation patterns [89].

3.5 Tradeoffs and Issues

This chapter ends by elaborating on the tradeoffs presented in Figure 3.1. Thus, a summary of the challenges that have been identified as common issues in the field, regardless of the scheme, is provided.

- * Reachability vs. Savings Most results sections in the literature measure the saved rebroadcast and reachability of probabilistic broadcast schemes by increasing node density. Moreover, teoretical analysis, with the assumption of ideal packet delivery, indicates that p_f should decrease monotonically as node density increases. However, studies rarely evaluate reachability in sparse networks, and when simulating models with realistic assumptions (e.g. 802.11 MAC layer), reachability is maintained with p_f showing a lower-bound at which no further savings are possible despite increasing node density. More studies are needed to evaluate the limits of saved rebroadcast with sustained reachability, especially in sparse networks with realistic wireless channel assumptions.
- ★ Overhead vs. Savings Most proposals have proved to be very effective at reducing the number of forwarded packets compared to *flooding*. However, from the assumptions made by the proposed schemes (e.g. ideal MAC layer, neighbor information availability, location services in nodes), it is difficult to know how much true overhead would be generated in a real implementation. In this regard, it is possible that the overhead of some schemes is comparable with the reduction in forwarded packets. Quantifying the true overhead of probabilistic broadcast protocols would constitute an interesting research endeavour, especially for neighbor-based schemes.
- ★ Information quality vs. Overhead It is necessary to devote specific efforts to specific target scenarios (e.g. WSN, VANET, MANET). The evidence suggests that the scenarios demanding higher quality information are those found in VANETs while in WSN energy is a priority. In VANETs, for example, high quality information (e.g. broadcast packets carrying the location of nodes or a list of neighbors IDs) is worth a reduction in latency. In WSN, latency is not as important as reducing the overhead of protocols. Therefore, each setting (e.g. WSN, VANET, MANET) can have its own overhead limits.
- * Route Length vs. Route Stability Aiming at covering the network area with a reduced number of forwarding events, area-based schemes prioritize the transmissions of nodes with the largest inter-node distance (i.e. higher values of p_f and shorter delays). However, relying on the most distant nodes to forward packets at every hop may result in short-lived or unstable propagation paths, especially when considering node mobility or realistic wireless channels. Since broadcast is the mechanism that provides routing protocols with forwarding paths, the routes relying on distant nodes

are very likely to disappear with minimum movement of nodes or due to fading and shadowing. Then, further attention should be given to the functions that determine the forwarding probability of nodes (p_f) for the benefit of routing protocols.

3.6 Locating the Contributions of this Thesis

According to the 5 areas identified in Section 3.4, the following chapters make contributions in the *awareness of the wireless channel* and its impact on reachability, and the understanding of probabilistic broadcasting through *analytical modeling*.

Recalling Figure 1.1 in Chapter 1, Figure 3.13 clearly points to the gaps in which the present thesis is making contributions.



Figure 3.13: Location of the gaps for the contributions made in the present thesis in the context map proposed in Section 1.

Chapter 4 focuses on analyzing the impact of realistic assumptions on reachability, namely the geometry of node layouts and the use of the 802.11 MAC layer with physical channel considerations (Signal to Interference ratio). Chapter 5 proposes a strategy to reduce broadcast latency and maintain reachability for the same value of p_f by modifying the parameters of the random delays used in probabilistic broadcasting protocols; a strategy that entails minimum overhead. Finally, Chapter 6 provides a theoretical framework to model the probability density function of broadcast end-to-end latency in various ad hoc scenarios with different node layouts and types of random delays.

	Chapter
	FOUR

Geometric Aspects of Probabilistic Broadcasting in Ad Hoc Networks

This chapter studies pure probabilistic broadcast in ad hoc networks under a variety of topological scenarios, offering a comparison of the performance of broadcast in lattice-like geometric node layouts (e.g. nodes arranged in triangular, square and hexagonal grids) with that in randomly placed nodes. Results suggest that the geometry of the position of nodes has an impact on the success of probabilistic broadcast techniques. Specifically, networks with randomly-placed nodes exhibit a near-ideal (collision-free) behavior, whereas the grid layouts are extremely sensitive to the impact of collisions and interference. To account for the unreliable behavior of broadcast under certain node-distribution geometries, this chapter provides an analysis of Signal to Interference ratio for ad hoc networks.

4.1 Introduction

In wireless ad hoc networks, nodes have no knowledge about the network topology, which makes broadcasting a fundamental building block for topology discovery and other network functions. Existing standards for broadcasting in ad hoc networks rely on a broadcast technique called *Flooding* [9]. In *Flooding*, every node receiving a broadcast message for the first time simply retransmits a copy of the message; under reliable (or almost ideal) channel conditions, this guarantees that messages reach all nodes in the network (i.e. 100% reachability). However, it has been shown that, in realistic wireless channels (i.e. when interference, delays, transmission errors, noise and radio-wave propagation features are considered), *Flooding* may not perform well [8], allowing the transmission of many redundant messages, leading to energy waste and channel contention.

To discuss the feasibility of using other protocols instead of *Flooding*, various broadcasting techniques, based on deterministic and probabilistic principles, have been assessed [90]. Despite producing minimum redundancy of broadcast messages, deterministic techniques depend on collecting topological information (i.e. transmitting additional control messages) and they increase the complexity of the broadcast process, which can be more expensive than *Flooding* in dynamic topologies [91] and large-scale networks [92]. Recent studies [10] have illustrated the advantages of forwarding broadcast messages at random to reduce redundancy and produce low overhead. This is known as *probabilistic broadcast*. *ing* and consists of assigning probability threshold values (i.e. the so called *forwarding probability*) to individual nodes to decide whether a copy of the message received should be forwarded or not.

Performance evaluations of the simplest probabilistic protocols, conducted in testbeds [45,93–95], confirm the advantages of probabilistic broadcasting. However, the results suggest limitations in the average reachability of transmitted messages, as well as the dependence of the reachability of transmitted messages on the position of the broadcasting source; these issues are not anticipated by existing models of probabilistic broadcasting.

Most probabilistic broadcasting models assume an *ideal* wireless channel [1, 96-98], and only a few models consider a *given* value for the probability of unsuccessful message reception [12, 45]; little attention has yet been given to the incorporation of the *causes* of unsuccessful reception (as in [99]) in these probabilistic broadcasting models.

Moreover, in most studies (both in test-beds or based on simulation) [10], the limitation in reachability of transmitted messages and the efficiency of probabilistic broadcasting protocols under realistic conditions are associated with the calculation of the forwarding probability value (i.e. the strategy used to avoid the redundancy of *Flooding*). Performance limitations are rarely associated with other factors, such as the position of nodes in the network.

The present study was designed to help understand the causes of the limitations in the reachability of transmitted messages, especially the impact of the position of nodes, while taking into consideration realistic assumptions about channel conditions.

In line with this, Signal to Interference Ratio (S/I), which has been successfully applied to the design of Cellular Networks [100], has been employed in this chapter to help analyze the impact of the geometry of node distribution on the protocol performance. Similar to [101] for improving network lifetime in large-scale sensor networks, the geometrical features of network layouts are identified here as a factor that significantly affects the success of probabilistic broadcasting protocols in static networks with a few thousand nodes. Simulation results indicate that networks with randomly-placed nodes exhibit a near-optimum (near collision-free) behavior, while grid layouts are extremely sensitive to the impact of node interference.

The contribution of this chapter is twofold:

- The incorporation of realistic assumptions (i.e. S/I model) in the analysis of the causes of the limitations in reachability of transmitted messages in probabilistic broadcasting schemes.
- An analysis of the impact of the geometry of network layouts on the reachability of probabilistic broadcasting protocols, as well as the greater strength of the influence of the node geometry in comparison to the impact of the position of the source node.

The present analysis is important for an understanding of applications involving hundreds of static nodes in which the geometry of network layout is a key factor in network performance. Results are relevant for grid-like geometries [102–105], for random geometries [106–108], and, in general, for terrestrial static sensor networks and environmental applications [109, 110].

The rest of this chapter is organized as follows. Section 4.3 summarizes previous evaluations of existing probabilistic broadcasting techniques. Section 4.4 introduces a

Signal to Interference analysis that illustrates the significant impact of node placement distributions on the reachability of ad hoc broadcasting. Section 4.5 presents two sets of simulation results, and finally, Section 4.6 concludes the chapter presenting a summary of the most relevant findings of the study.

4.2 Table of Chapter Variables

Table 4.1 summarizes and describes the variables used in this chapter.

Variable	Description
r:	Transmission radius of nodes. It is assumed to be equal for all nodes.
S/I:	Signal to Interference ratio.
d:	Euclidean distance. Regardless of the subscript (e.g. d_1), this variable always
	refers to an Euclidean distance between nodes
P_r :	Power of the signal received by nodes (in Watts).
P_t :	Power of the signal transmitted by nodes (in Watts).
G_t :	Transmitting antenna gain.
G_r :	Receiving antenna gain.
h_r :	Height of the receiving antenna above ground level.
h_t :	Height of the transmitting antenna above ground level.
$I_{\{X < p\}}$:	Indicator function with parameter p ; $I = 1$ when $X < p$, $X = 0$ otherwise.
n:	Number of nodes in the previous tier.
k:	Number of hops for which a packet is forwarded with probability $p = 1$
	regadless of the broadcasting scheme; $k = 4$ in this chapter.

Table 4.1: Variables used in Chapter 4.

4.3 Related Work

The purpose of the present chapter is to discuss some previously unexplored causes for failure of models proposed for probabilistic broadcasting protocols which assume the existence of an ideal channel. This section focuses on a set of studies that have already assessed the performance of probabilistic ad hoc broadcasting schemes and have observed existing models lacking precision under realistic settings using both simulation and experimentation in testbeds.

One of the first studies addressing the success of probabilistic broadcasting [111] identified that the network size and node degree impact on the avoidance of the *phase transition phenomenon* (i.e. a sudden transition from reaching less than 20% of the nodes to over 90% of the nodes with a small change in the forwarding probability). However, the collisionprone conditions were only implemented for randomly placed nodes. Grid scenarios were employed primarily to show the absence of the phase transition. Despite the inclusion of a moderate number of nodes in the random-placement scenarios, the authors clearly showed that, under realistic MAC-layer conditions, forwarding probability values should be carefully chosen so as to obtain the most from the pure-probabilistic broadcasting principle. Specifically, it was shown that, in more realistic, collision-prone environments, there is a maximum broadcasting success ratio when the probability of forwarding a message is as low as p = 0.1. As the probability increases, a larger number of collisions occur and, as a consequence, very few nodes can receive broadcast messages. The experiments designed to assess the effects of packet collisions in [111] were conducted using various simultaneous broadcasting sources, which makes it difficult to identify whether or not most of the collisions occurred as a consequence of self-interference of a message with its own copies or as a consequence of random collision of different messages.

More recent studies [45] introduced a retry-based scheme to improve reachability of the broadcast messages. By using both testbeds and simulation, the authors corroborated that interference and collisions must be included in probabilistic broadcasting models to allow for accurate predictions. The authors even proposed mapping between a more elaborate percolation model [112] and their retry-based scheme that could be used in future work to capture the behavior of the experiments adequately. Again, the conditions for inducing a high-interference environment were oriented by the use of many simultaneous broadcast sources. The interference caused by this large number of simultaneous sources was only tackled experimentally by adjusting the frequency of retries and little attention was devoted to the possibility of a broadcast message interfering with its own copies. Since the diameter of the network allowed good reachability with just a few hops, the number of nodes causing interference with the same copy of a broadcast message decreased very rapidly with the distance (in hops) from the broadcasting source.

The same authors of [45] have also reported empirical results suggesting that special attention should be given to the effects of channel interference on the expected behavior of probabilistic broadcasting [95], especially when multiple broadcasting sources transmit simultaneously. Evidence from the test-bed experiments indicated that probabilistic broadcasting techniques do not avoid the collision of broadcast messages and excessive overhead. This observation was supported by the fact that certain reachability limits could not be surpassed during the evaluation in the test-bed. These findings are consistent with more recent studies [113], in which the performance of a completely different family of probabilistic broadcasting schemes (based on dissimilarity metrics and Euclidean distances) also evinced the reachability limits associated with the same test-bed.

What is common to all of the studies described here is that the limited reachability and efficiency of probabilistic broadcasting protocols under realistic conditions are associated to the way the calculation of the forwarding probability is carried out (i.e. the strategy used to avoid the redundancy of *Flooding*); and performance limitations are rarely associated with the geometrical properties of node placement in the network. Since little attention has been devoted to addressing the impact of the positions of nodes and the incorporation of realistic channel conditions in models, the following sections discuss the advantages of integrating the expected geometrical properties of network-wide node placement distributions into the assessment of probabilistic broadcasting techniques.

4.4 Impact of Geometrical Node Distribution on Probabilistic Broadcast

The analysis in this section shows that the geometrical distribution of nodes is crucial for the success of the delivery of messages in ad hoc networks with static nodes employing a probabilistic broadcasting protocol. It also shows the difference in broadcasting conditions existing in grid layouts and those in a random layout. The assumption of nodes being static makes possible the evaluation of the impact of the geometry of nodes on broadcasting, while excluding potential effects due to node mobility. Moreover, the present analysis can be considered as a first step towards the understanding of the impact of node geometry resulting from mobility on broadcasting, since the random layout used in this chapter captures the asymptotic behavior of some mobility models (an example can be found in [114]; the interested reader can also refer to [115] for details on mobility models that converge to asymptotic node spatial distributions).

The same principles in [100] for the analysis of co-channel interference in Cellular Telephone Systems are employed for the study of variation in the Signal to Interference Ratio (S/I) as a function of the position of nodes with respect to the broadcasting source. The calculations of S/I serve to evaluate the impact of self-interference of a broadcast message with its own copies in a single broadcasting wave, which is independent of collisions caused by high-rate traffic patterns. Specifically, the values of S/I in two scenarios are calculated and compared, the first scenario corresponding to a regular node layout of degree 4 (i.e. square grid), and the second to a layout where nodes are placed at random.

Figure 4.1 illustrates a broadcasting source node located to the left of a square grid where the transmission range of the nodes (r) can cover four neighbors (i.e. a regular geometric graph of degree four). Diagonal dashed lines show groups of nodes associated with the same hop distance from the broadcasting source (i.e. the 3 nodes at a one-hop distance from the source are Tier 1, and the 5 nodes at a two-hop distance are Tier 2). The figure also shows three double arrows explicitly indicating the Euclidean distance of one node in Tier 2 (the receiving node) with respect to the three nodes in Tier 1. This situation corresponds to a worst-case scenario in which all the nodes in the previous tier (i.e. Tier 1 of the figure) transmit at exactly the same time, causing maximum interference. In any situation, the Collision Avoidance procedure in the medium access protocol (CSMA/CA in Layer 2) reduces such interference. Indeed, the inclusion of such a worst-condition assumption leads to a lower-bound model of the impact of network geometry on broadcasting without needing to consider a detailed time-dependent analysis.

In the traditional analysis of collision events, only the two nearest nodes (nodes at distances d_1 and d_s) would be considered to determine the probability of successful message delivery, but the analysis carried out here considers the calculation of S/I at the receiving node, including the interference of all nodes located in the previous tier. This makes possible the inclusion of the impact of node distribution in the analysis.

Without loss of generality (as in Chapter 3 of [100]), the S/I calculation assumes a two-ray pathloss model, in which the power of the signal received at a distance d from the transmitter can be estimated using the transmitted power P_t together with the gain and height of both the transmitting (G_t, h_t) and the receiving antennas (G_r, h_t) as follows:



Figure 4.1: S/I for a receiving node in the second tier, where interference comes from two of the three nodes in the first tier.

$$P_r = P_t G_t G_r \frac{{h_t}^2 {h_r}^2}{d^4}$$
(4.1)

Consequently, for the receiving node in the second tier of Figure 4.1, the value of S/I corresponds to:

$$\frac{S}{I} = \frac{P_t G_t G_r h_t^2 h_r^2 \frac{1}{d_s^4}}{P_t G_t G_r h_t^2 h_r^2 \left(\frac{1}{d_1^4} + \frac{1}{d_2^4}\right)}$$
(4.2)

where d_s is the distance between the expected transmitter from the previous tier and the receiving node, and d_1 and d_2 are the distances from the receiver to the other two transmitters from the previous tier (i.e. Tier 1). Assuming homogeneous characteristics of all nodes in the network (i.e. similar antenna heights, antenna gains and transmission power), the previous expression reduces to:

$$\frac{S}{I} = \frac{\frac{1}{d_s^4}}{\left(\frac{1}{d_1^4} + \frac{1}{d_2^4}\right)}$$
(4.3)

By considering again Figure 4.1, and the forwarding probability p in the analysis, equation (4.3) becomes:

$$\frac{S}{I} = \frac{\frac{1}{d_s^4}}{\left(I_{\{X_1 < p\}} \frac{1}{d_1^4} + I_{\{X_2 < p\}} \frac{1}{d_2^4}\right)}$$
(4.4)



Figure 4.2: A zoom-out representation of Figure 4.1 for 10 rows and 13 columns of a square grid with the source node on the left (row 5 column 1 of the grid). Solid lines indicate the shape of the expected tiers (angle-shaped tiers) propagating the message from left to right. The minimum number of hops to reach the node in the top-right corner is 16 (i.e. Tier 16).

where $I_{\{X_i < p\}}$ is the *indicator function*, and X_i is a random value drawn at the *i*-th node (i.e. $X \sim U([0, 1])$), such that I = 1 when $X_i < p$, otherwise I = 0. This includes, in Equation (4.3), the Bernoulli variables that indicate whether or not nodes in the previous tier forward a broadcast message. Notice that the numerator has no *indicator function* because the intention is to calculate the average signal quality (S/I) only when broadcast messages are *received*, i.e. the calculation of *probabilistic interference* given that nodes receive messages from their expected previous-tier transmitter. The value of p, thus, dictates the reduction in interference as p decreases.

Finally, for an arbitrary receiver with n interferences (i.e. n nodes associated with the previous tier, excluding the expected transmitter), the calculation of *probabilistic* S/I corresponds to:

$$\frac{S}{I} = \frac{\frac{1}{d_s^4}}{\sum\limits_{i=1}^n I_{\{X_i < p\}} \frac{1}{d_i^4}}$$
(4.5)

where d_s is the distance between the receiver and the expected transmitter whereas d_i is the distance between the receiver and the *i*-th interferer.

Indeed, Equation (4.5) is a ratio of distances, for a single interferer $S/I = (d_s/d_1)^{-4}$. Therefore, what matters is the positions of the nodes relative to each other, and not the absolute distances (in meters). This highlights even more the fact that the geometry of nodes is a key issue in the S/I analysis.

Equation (4.5) is used in the remainder of this section to calculate the S/I values for all nodes in the network. Notice that this calculation does not depend on the network layout (i.e. it is not restricted to the example in Figure 4.1) and can be used to estimate the values of probabilistic S/I in any network as long as tiers can be established with respect to a broacasting source. Besides that, other models of propagation different from the two-

ray pathloss model would result in a similar equation. For example, for a more general propagation model such as the *log-distance path loss model with log normal shadowing* the only change in Equation 4.5 would be the power of the Euclidean distances (known as the path-loss exponent), which would not be 4. Moreover, measurement studies [100] have reported that for urban environments (path-loss exponent between 3 and 5) as well as for indoor environments with obstacles (path-loss exponent between 4 and 6) the value 4 for the path-loss exponent can be used. A good example can be found in [116], using a path-loss exponent value of 4.2. Thus, results obtained using Equation 4.5 are still general enough for the use of a more elaborate propagation model.

Suppose now that the square grid topology in Figure 4.1 is part of a larger square-grid network (i.e. regular layout of degree 4). Figure 4.2 illustrates a *larger-scale* view of the way tiers are created as a function of the position of the source. This 10-row, 13-column square grid illustrates how tiers take an angular shape, resembling two plane wavefronts that propagate away from the source (up to tiers 16 and 17 for nodes in the right-hand corners of the rectangular region). If this pattern is maintained for a 20×50 square grid, the farthest node would be seen in Tier 60. Figure 4.3 illustrates the calculation of the probabilistic S/I values averaged for the 60 tiers of such a 20×50 grid with a source positioned in row 10 column 1. The x-axis indicates the tier, i.e. the hop-distance from the broadcasting source. The values of probabilistic S/I were calculated 100 times and the average per node was calculated. Finally, each point in Figure 4.3 shows the average of probabilistic S/I values per tier. For example, for Tier 2, one value of S/I is obtained from 5 nodes; for Tier 3, the result is obtained from 7 nodes, and so on. The main idea here is to observe how the average value of S/I varies as the broadcast message propagates away from the source. Figure 4.3, thus, represents the variation in the quality of the received wireless signal as the distance, in hops, to the source node increases, that is, as the message propagates from left to right throughout the 20×50 grid. The values of the first two tiers were not included in the figure because the nodes in Tiers 1 and 2 produced very few interferers and several values of S/I were infinity.

Indeed, the S/I calculation gives the average quality of the signal for a worst-case broadcasting scenario in which transmitters are totally synchronized, leading to a maximum number of closest interferers (i.e. all interferers from the previous tier). Unlike some deterministic broadcasting algorithms [117], synchronized nodes are not essential for the operation of the protocol, but make a worst-case assumption. For any other scenario, in which heterogeneous delays in nodes imply transmitters desynchronization, the Collision Avoidance procedure of nodes reduces interference, and as a consequence, the S/I values are expected to be higher than those in Figure 4.3. Thus, the curves in the figures in this section represent lower bounds for the S/I values.

Higher values of S/I imply greater chances of successfully propagating broadcast messages over additional tiers. Moreover, nodes near the farthest corners of the region would have a low rate of successfully received messages, since these nodes lack neighbors, rather than because of the low values of S/I. Such reachability in the last tiers is independent of the geometry, as shown in Section 4.5.

The role of the geometry of nodes can be better emphasized when comparing the S/I values in Figure 4.3 with those presented in the measurement-based studies in [118]. For example, in the first 10 tiers, an S/I value around 10dB would result in nearly 100% of successfully received messages at low rates of 1Mbp, which is a standard broadcast
rate [9], while for nodes located at more than 10 tiers from the source, with S/I values between 2dB and 6dB, the percentage of successful transmissions would not exceed 50%. Moreover, regardless of the forwarding probability, the curves in Figure 4.3 clearly reveal three network segments, namely Tiers ≤ 10 (receiving the highest S/I values), $11 \leq$ Tiers ≤ 49 and $50 \leq$ Tiers. These three segments are the points at which the propagation of broadcast messages reaches the edges of the rectangular area. This may indicate that the relation between the geometry of node position (grid of degree 4) and the shape of the area where the nodes reside (a rectangle) also has an impact on the average quality of the channel to receive broadcast messages. This impact is as significant as the impact of the forwarding probability value, since the difference between the curves for p = 1 to p = 0.65is similar to that between S/I values in tiers 1-10 and tiers 10-50 for all the three curves.

It is worth mentioning that the values of p chosen for the S/I curves in Figure 4.3 are well supported by the literature on probabilistic broadcasting [10]. In general, values between 0.6 and 0.75 are of great interest since they lead to minimum levels of redundancy (i.e. maximum Saved Rebroadcast, SRB in the literature) while maintaining the reachability of transmitted messages at over 90%. In the majority of investigations on probabilistic broadcast the optimal value of p is around 0.7, except for studies in which the average node density is over 10 nodes within the transmission range, for which the optimal values of p are far lower than 0.7.



Figure 4.3: S/I profile for nodes in the 20 × 50 grid of degree 4 as a function of the tier (i.e. hop distance to the broadcasting source). Three curves corresponding to three values of probability. Notice that lower probability leads to less interference, that is, better average S/I.



Figure 4.4: S/I profile for random node placement as a function of the hop distance from the source. Lower probability yields higher S/I values. This profile is more stable than that for the grid.

Now the calculation of probabilistic S/I in Equation 4.5 is applied to another rectangular region, also containing 1000 nodes, but with these nodes located *at random* according to a uniform distribution (see Figure 4.5). The transmission range of the nodes (r) was set at 250m and the dimensions of the rectangle at $7500m \times 3000m$. As in [16], this setting leads to a high probability of having a connected network [2,3]. This means that, if there is a smaller value of r, fewer nodes, or a larger region, a considerable proportion of random layouts would result in disconnected networks, which impacts the performance of broadcast. Additionally, this setting has the same ratio of height to width for the sides of the rectangular area (i.e. 2 : 5 ratio as the 20×50 square grid). To make a fair comparison with the results of the grid, 100 different random layouts were derived. For each random layout, tiers were established with respect to the nodes located closest to the point (0, 1500) (left center nodes) so that left-to-right propagation of broadcast messages was maintained. Results for random placement are shown in Figure 4.4.



Figure 4.5: Large-scale view of the positions of 1000 nodes for a) the random scenario and b) the 20×50 square grid in two rectangular regions with equivalent height-width ratio.

Results in Figures 4.3 and 4.4 indicate that the random setting exhibits higher values of S/I and far more stable conditions for broadcast messages to propagate from the source node towards the end of the network. For example, in the square grid, with the lowest interference conditions (p = 0.65), the S/I remained below 8dB from tier 10 onwards. Meanwhile, with the random layout and the highest interference (p = 1), S/I was always above 7dB. Based on the results, it is important to point out that interference has a stronger impact on the grid, where nodes have fewer neighbors, than on the random layouts, in which nodes have an average of eight neighbors. This is equivalent to saying that nodes with fewer neighbors suffered more interference because they are distributed in a grid, whereas nodes with a higher number of neighbors (on average) experience less interference due to the randomness of their relative positions.

The analysis in this section yields three important observations. First, there are aspects specific to the geometry of ad hoc networks (namely, the distribution of node placement) that may significantly affect the performance of probabilistic broadcasting techniques. Second, it seems that random settings offer more stable conditions for the propagation of broadcast messages. Finally, by integrating the Euclidean distance in the analysis, the S/I calculation can be used to predict the expected success of ad hoc probabilistic broadcasting messages as a function of the network-wide node distribution.

4.5 Simulation Experiments

4.5.1 Random and Regular Placement of Nodes

The previous section showed that, for realistic wireless channels, ad hoc networks with random layouts facilitate the propagation of broadcast messages more than do grid layouts. To further verify this conclusion, the present section shows results about the performance of probabilistic broadcast derived via simulation. The simulator used was QualNet(R)v7.3 [119]. Table 4.2 displays the values of the simulation parameters. The two network scenarios correspond to the square grid and the network with random placement of nodes described in the previous section. Nodes in the first four tiers forward broadcast messages with a probability of p = 1. This difference in scenarios was adopted for comparison with the results in [6], which used a technique called GOSSIP1(p,k). The work in [6] was taken as benchmark because the authors offered a good comparison of probabilistic broadcasting performance employing both grids and random node layouts. The parameters of *GOSSIP1* are p for the forwarding probability, and k for the number of tiers where nodes use p = 1(i.e. the simulations were run with GOSSIP1(p,4)). The results in [6] were derived with the assumption of an ideal MAC layer, so the following comparison illustrates the impact of realistic conditions on the performance of probabilistic broadcast, which reinforces the observations in Section 4.4. The values of p in the present study are the same as those used in [6], since values for p around 0.7 provide maximum reachability with the minimum number of message copies, especially for the scenarios considered in this section. Values higher than 0.75 would lead to a higher probability of collisions and waste of energy and bandwidth, whereas values lower than 0.6 would result in a considerable reduction in reachability (reachability near 20%) [93,111].

Parameter	Value
Number of nodes	1000
Area of Uniform Random	$7500m\times 3000m$
Area of $20\times 50~{\rm grid}$	$12500m\times5000m$
Transmission Range	250m approximately
Path-loss model	Two-ray
Mobility	none
Mac Layer	802.11
Technique	GOSSIP1(p,4)
Source Position	left-center
Runs	120 replications

Table 4.2: Simulation settings

Figures 4.6 and 4.7 show the impact on probabilistic broadcasting performance when employing realistic channels instead of ideal ones. These figures measure the percentage of nodes reached by broadcast messages as a function of the node distance, in hops, from the broadcast source (i.e. percentage of nodes of the same tier). When realistic conditions affect the performance, the most noticeable difference is for the case of the square grid with forwarding probability p = 0.72 (Fig. 4.6), in which reachability decreases by more than 15% after 30 hops. As the probability decreases in the grid scenario, the



Figure 4.6: Comparison of probabilistic broadcasting success as a function of node distance to the broadcasting source, for three values of p. Collision-less, ideal conditions vs. realistic conditions for the 20×50 grid.

gap between results for the *ideal* and *realistic* channels also decreases, since an *emptier* channel resembles an ideal channel.

The realistic curves for p = 0.72 and p = 0.65 in the square grid (Figure 4.6) show a sudden drop after Tier 10, as predicted by the decrease in the average S/I values shown in Figure 4.3. Such a drop can be associated with the interplay between the rectangular shape of the area containing the nodes and the position of the source node; indeed, for the left-hand centered position of the source in the 20×50 grid, the propagation of the message meets the first border of the network at hop 10.

In [6], a slightly higher probability value (0.75 > 0.72) for the random scenario was employed to achieve a reachability similar to that of the grid scenario. When realistic conditions are considered, however, the curves for p = 0.65 show that the random scenario does not require a higher probability value in order to match (or even outperform) the reachability of grid scenarios, as confirmed in Section 4.5.2.

What is common for all the probability values evaluated in Figure 4.6 is that reachability values in the *ideal* curves show a smoother decrease than their *realistic* counterparts. In fact, the realistic curves oscillate more, reaching more nodes than do the ideal curves in some of the tiers (as is the case of p = 0.65 between hop-distance 12 and 30). Oscillating curves may arise when many nodes of one tier successfully receive a broadcast message, so that transmission to the next tier has to tolerate maximum interference. However, when only a few nodes receive the message, their transmissions to the next tier will produce low values of interference, increasing the chances of successful broadcast propagation in subsequent tiers.

Figure 4.7 shows that the percentage of nodes receiving the message in different tiers (hop-distance) is almost the same for both the *ideal* and *realistic* scenarios. In fact, for p = 0.65, the curves after hop 22 are almost identical, whereas before hop 22, the realistic curve reached a higher fraction of nodes. Such a trend is consistent with the S/I analysis in Section 4.4, shown in Figure 4.4. The broadcasting propagation offered by the



Figure 4.7: Comparison of broadcast reachability under ideal conditions vs. more realistic conditions for Uniform Random Node Distribution with p = 0.75 and p = 0.65.

randomly-placed node layouts is stable, even for the worst-case scenario of the interference analysis.

These results reinforce the need for analyzing the impact of the geometry of ad hoc networks on the performance of probabilistic broadcasting, as described in Section 4.4, especially when observing that:

- Random network-wide node distributions promote the success of probabilistic broadcasting transmissions in realistic scenarios more than do grid layouts.
- The geometry of node distribution should be considered in the performance of ad hoc probabilistic broadcasting.

In order to draw more general conclusions, the following section explores the consistency of these findings. By changing the position of the source node, different patterns of broadcast transmission should be seen. The following section also shows the impact of using other geometries such as a triangular grid, a larger square grid, and hexagonal grid, as well as variations in random distributions (e.g. Uniform, Normal, Poisson).

4.5.2 Other Grid Geometries and Node Distributions

This section provides a broader view of the relation between probabilistic broadcasting and node placement by exploring more general scenarios that involve various different grid layouts, as well as random node distributions. The following simulation results also extend the discussion about the impact of the position of the broadcasting source on broadcasting reachability, motivated by previous work [93,94] in a test bed, which showed that results are strongly dependent on the position of the source node. The simulation parameters are those in Table 4.2. However, in this section, the number of nodes and the shape of the area vary. The following results were obtained with 2500 nodes within a square region (e.g. a 7000×7000 square region for the random settings). For the random settings employing the Normal and the Poisson distributions, the number of nodes is on average 2500 since the number of nodes is also a random variable. Every single replication of the simulation was run using a different set of random numbers for both the forwarding decisions and the random node positions. Therefore, the results capture the average behavior of node distributions and avoid the potential bias of using a single geometrical realization for all replications. Having such a variety of scenarios (120 for each random distribution) using a test bed would be prohibitive, but it is feasible using simulations, as has been done here. Moreover, the different random positions of nodes in each replication can be seen as snapshots of a mobile network; since the density functions of the random variables that determine the positions of nodes at different times converge to a unique density function in the long run (cf. [114]). A mobility model with such a property is said to be *stationary* [115].



Figure 4.8: Comparison of the shape of tiers for the Hex-grid and Tri-grid for sources in the bottom left corner. The Tri-grid covers more nodes in each tier. Consequently, for grids of the same dimensions the Tri-grid results in less tiers than both the Hex-grid and the Sq-grid.

As in the previous sections, the figures show the reachability profile of probabilistic broadcast, namely the fraction of nodes that successfully receive broadcast messages as a function of the node distance, in hops, from the source (i.e. the tier). Moreover, the figures are organized to illustrate the impact of changing the position of the source node, which in turn changes the conditions for the propagation of messages.

Figures 4.9 to 4.11 show the behavior of probabilistic broadcasting in 50×50 grids in which nodes (except the nodes forming the sides of the grid) have three, four, or six neighbors. Hereinafter the prefix Tri will corespond to the regular grid of degree six (i.e. triangular grid), the prefix Sq to the grid of degree four (i.e. square grid), and the prefix Hex to the grid of degree three (i.e. hexagonal grid or honeycomb grid). Figure 4.9 focuses on all scenarios with broadcasting sources located in one of the four corners of the square region. Figures 4.10 and 4.11 show the curves for scenarios with broadcasting sources on one side and at the center of the grids, respectively. In the figures, the curve length depends on the grid. The Tri-grid curves are always the shortest. In these figures, the longest curves always correspond to the grid of degree three (i.e. Hex-Grid), which has 123 tiers for the corner source, more than 110 tiers for sources on the side of the grid, and 63 tiers when the source lies at the center of the grid. Figure 4.8 shows how the topology of an Hex-grid leads to fewer nodes per tier than does its Tri-grid counterpart. As a consequence, the Tri-grid needs less tiers to cover the same number of nodes. This figure can be compared to Figure 4.2 to further illustrate how the number of tiers (x-axes in the figures) was obtained for each simulation scenario.



Figure 4.9: Reachability when the source node is located in one corner of the grid. The x-axis shows the hop distance from the source while the y-axis indicates the average percentage of nodes that received the message at particular distances (i.e. on average, the percentage of the tier that received broadcast messages).

For geometrically regular node placement, the Tri-grid is consistently the most favorable setting for the success of probabilistic broadcasting. For example, with the source located in the corner, for p = 0.6, the broadcast messages reached slightly fewer than 80% of the nodes from tier 15 to 55, while for p = 0.65 and p = 0.72, messages reached over 90% of the nodes. This occured because the triangular grid has more *collision-free* nodes receiving messages in every tier than the other grids. For example, when the source is at the center of the grid, in the worst case scenario in which all nodes from the same tier transmit at the same time, the triangular grid still guarantees that at least six nodes can propagate the message in all directions with no risk of collision. For a source node at the center, the square grid offers only four collision-free receivers per tier and the hexagonal grid only three collision-free nodes per tier. Similar geometrical limitations can be shown for sources on the side and in the corner of grids. However, for the hexagonal grid, collisions due to time-correlated propagation of broadcast messages occur only every two tiers. This lack of redundancy means that, for every collision-free tier, the use of probabilistic broadcasting simply reduces reachability, similar to the propagation of broadcast for nodes lying in a straight line. This explains the negative-exponential shape of the curves (curves of the Hex-grid), almost resembling the graph of a geometric sequence for the value of p(recall that 0).

Moving the source to one side, and then to the center, always improves reachability. In the first tiers, a sudden decrease in reachability occurs, even before the fourth tier. This is somewhat unexpected since we know that broadcast packets are forwarded with p = 1for the first four tiers. This early drop in the curves is more accentuated when the source



Figure 4.10: Reachability when the source node is located on one side of the grid (geometrically regular node placement scenarios only).

is in the corner (Fig. 4.9). As the message propagates over the tiers, either the same reachability is kept until the end of the network or reachability increases, suggesting that, under *realistic* collision-prone conditions, the unreliable delivery of the MAC layer occurs mostly in the first tiers, due to the highly correlated times when forwarding messages from the same source. After the first collisions occur and the message propagates farther from the source, forwarding times gradually lose correlation and facilitate the operation of the CSMA/CA mechanism, which explains the shape of the curves.

The number of collision-free nodes per tier is extremely important in providing robustness. Namely, for the triangular grid, the curves for p = 0.72 show almost the same performance as the curves for p = 0.65, which represents a 10% reduction in the number of messages forwarded, but the reachability is still the same. Moreover, the curves for the same probabilities differ significantly in other grids.

The impact of having a finite network is noticeable in Figure 4.9. Specifically, for sources in the corner, there is a consistent *tweak* at around tier 50 for all curves, which is the tier where the sides of the network end. This *tweak* is even more noticeable with the broadcasting source on the side (Fig. 4.10), where a marked change can be observed in all curves at around tier 25 (the tier at which the broadcast propagation wave encounters the firts borders of the square region). In these two figures, there is always a segment of the curves, which corresponds to the transmission encountering the border, when the broadcasting process simply recovers, reaching a larger percentage of the nodes than in previous tiers. Conversely, when the source is at the center of the grid (Fig. 4.11), the curves always decrease, although some slight changes can also be observed at tier 25, where the number of nodes per tier begins to drop. These patterns can be explained by the way the number of nodes changes on a tier as the message propagates. For example, the square grid with *the source on the side* has 3 nodes in Tier 1, 5 nodes in Tier 2, 7 nodes in Tier 3, and so on, until tier 25. As the size of the tier increases, the number of nodes in the border space of the square region represents a smaller percentage of the tier. At hop 25, the



Figure 4.11: Reachability when the source node is located at the center of the grid (geometrically regular node placement scenarios only).

size of the tier and the proportion of its border nodes stabilizes (remaining 51 nodes with two nodes on the border of the region) until hop 49, where the size of the tier decreases continuously to the edge of the network. It can be affirmed that a stable behavior of the probabilistic broadcasting phenomenon can only be seen for 25 tiers (from tier 25 until tier 49). After this, the shape of the curves will be closely related to the proportion of border nodes per tier.

For the three random network scenarios (Figures 4.12 - 4.14), the overall performance is more stable than that of grids, as suggested by the analysis in Section 4.4. Even with the lowest probability, when the broadcasting source is in the corner (Fig. 4.12 with probability p = 0.6), the fraction of nodes reached by broadcast messages was always over 80%. As the source moved to the side (Fig. 4.13), and then to the center of networks (Fig. 4.14), reachability always improved. It is important to mention that all random scenarios were set using the theoretical Critical Transmitting Range for Connectivity, which is defined as the minimum value of r (i.e. the transmission range) such that the resulting communication graph is connected. It is important to recall that a graph is connected if and only if there exists at least one path connecting any two nodes in the graph. For the particular case of the random Normal distribution, a detailed view of the first values of the graph was necessary (see the right side of Figures 4.12 - 4.14), since these networks had a very small diameter. This situation results from maintaining connectivity when employing the Normal distribution, since for this distribution the Critical Transmitting Range for Connectivity [120] is greater than those for the Uniform and Poisson [3] distributions. Despite the reduced number of observable tiers for the Normally distributed node layouts, reachability was even greater than in the scenarios with Uniform and Poisson distributions.

What is common in these figures is the impact of the position of the source node on broadcast reachability. Regardless of the situation (whether grids or random placement), it is clear that when the source is in the corner, the lowest reachability is obtained in all scenarios. In fact, results for sources located in a corner can be taken as a lower bound



Figure 4.12: Reachability for Geometric random networks as a function of hop distance from the source (source node located in one corner of the network). These Figures include a plot zoom on the right so that the curves for the Normal distribution can be observed in the same scale as the other six curves.

for performance, which is important for design purposes (cf. test-bed results in [93,94]). Indeed, results confirm that any other position for the source node, other than the corner of the region, would result in greater reachability. Meanwhile, when the source is on the side, the wave of broadcast seems to pass through segments of recovery where reachability is greater than for previous tiers. The figures also show that centered sources always result in the greatest reachability, despite showing monotone decreasing reachability values (i.e. no segments of recovery or stability).

Finally, the random layouts reduce the highly-correlated forwarding times that degrade the performance of probabilistic broadcasting in geometrically regular layouts, facilitating the operation of the CSMA/CA mechanism to avoid collisions, especially in the first tiers of a single-message broadcast propagation event. This property provides random layouts with *robustness* against the effects of collisions when applying probabilistic broadcasting in ad hoc networks.

4.6 Chapter Summary and Future Work

The present chapter has highlighted the impact of node placement on the success of probabilistic broadcasting in large-scale static ad hoc networks when realistic wireless channels are considered. The calculation of S/I levels revealed that random placement of nodes facilitates propagation of broadcast messages, while nodes arranged in grids face reachability limitations. S/I analysis (borrowed from the co-channel interference analysis of cellular networks) has made possible the integration of the Euclidean geometry of node position with the benefits of probabilistic broadcasting, and this should be considered for future development of ad hoc probabilistic broadcasting protocols.

The observations obtained in the analysis have been further confirmed by running a comprehensive set of complementary simulations that included more and larger grids



Figure 4.13: Reachability for Geometric random networks as a function of hop distance from the source (source node located on one side of the network). Notice the zooming on the right for visualization of the curves associated with the Normal distribution.

with different geometries, as well as more geometrically random scenarios. The evidence clearly indicates that grid layouts render ad hoc networks extremely sensitive to the impact of realistic conditions. Conversely, when nodes are placed at random, the behavior of probabilistic broadcasting appears to reflect more ideal conditions.

The results also reveal that the use of tiers to characterize the propagation of broadcast messages facilitates the analysis of broadcast reachability profiles, providing detailed information about the broadcasting process e.g. the interplay between the shape of the area surrounding the network and the position of the source node. Furthermore, realistic reachability profiles show an oscillating pattern as a function of the distance in hops from the broadcast source, with tiers receiving a large percentage of broadcast messages being followed by tiers with low reachability, and vice versa. This suggests that, after a series of collision events, there will be a series of successful broadcast events as a consequence of the previous reduction in potential broadcast interferers. Under ideal conditions, these collisions would not occur and the oscilating pattern is not observed.

The benefits of studying non-conventional causes of pure-probabilistic broadcast failure (e.g. geometrical features rather than node density or broadcasting rates) under non-ideal wireless conditions are two-fold. First, the simplicity of pure-probabilistic broadcasting leads to faster broadcasting waves, which is convenient for overcoming the effects of node mobility and should have a positive impact on higher-level network functions, such as Admission Control when applied to different routing protocols [121]. Second, for scenarios where grids are relevant [102–105], simple, yet effective, countermeasures can be adopted to avoid poor performance of their ad hoc broadcast protocols.

Future work should include the study of latency and mobility under various geometrical conditions. Simulations of large-scale networks like the one in [102] should be analyzed in order to assess the feasibility of real probabilistic-broadcasting implementations. The analysis in this chapter, including the ideas in [111], about changes in the behavior of



Figure 4.14: Reachability for Geometric random networks as a function of hop distance from the source (source node located at the center of the network).

broadcast reachability depending on network size and node degree, should be conducted in order to evaluate the impact of geometry on small networks. Also, probabilistic broadcasting and the impact of geometry can be integrated in the performance evaluation of MAC layer protocols in a multi-hop context [122].

4.7 Table with Related Work

Table 4.3 shows a comparison of the studies that have focused on the assessment of reachability using probabilistic broadcasting schemes (related work). The columns of the table compare the network size, the geometry, and the average node degree of the networks in which the protocols were tested.

Network size ranges from some tens of nodes (i.e. $\leq 10^2$) to a few thousand nodes ($\geq 10^3$). In simulation studies, both grids and random networks of different sizes were used whereas in test-bed studies the size of the network appears explicitly (105 or 107 nodes).

Some of the results were reported for grids or random networks while assuming ideal conditions of the wireless channel (no packet loss or interference, that is, no packet collision events). Scenarios using a realistic MAC layer are indicated using the acronym MAC.

The average node degree has three possible ranges. The range 5-8 is the minimum range for percolation to occur (recall Chapter 2). The ranges 8-15 and > 15 appear in most of the literature as dense networks [10].

	Size (N)			Geometry			Av. degree		
	$\leq 10^2$	$10^2 - 10^3$	$\geq 10^3$	Grid	Rnd	T.bed	5-8	8-15	> 15
Sasson [111],	grid,	grid	grid	ideal	MAC		\checkmark		\checkmark
Multi-source	rand.								
Blywis [45],		107				\checkmark	\checkmark		
Multi-source									
Blywis [95],		105				\checkmark	\checkmark		
Multi-source									
Haas [6],		grid,	grid,	ideal	ideal,		\checkmark	\checkmark	
Single-source		rand.	rand.		MAC				
Lysiuk [1],		rand.	rand.		ideal		\checkmark	\checkmark	
Single-source									
Forero [53],			grid,	ideal,	ideal,		\checkmark		
Single-source			rand.	MAC	MAC				

Table 4.3: A comparison of studies about reachability and percolation-property evaluation in the literature.

Chapter
FIVE

Latency Reduction in Probabilistic Broadcast Protocols for Ad Hoc Networks

Broadcasting protocols in ad hoc networks usually employ uniform random delays on a per-node basis, which increases end-to-end latencies but helps reduce collisions at the MAC layer. Reducing random delays in these cases can reduce broadcast reachability. This chapter proposes the employment of the truncated-exponential distribution instead of the uniform distribution, which has been proved to reduce significantly the broadcast latency and yet maintain broadcast reachability.

5.1 Introduction

In ad hoc networks employing probabilistic broadcast protocols, nodes forward broadcast packets after an interval of random duration starting at the time of the packet arrival at the node [10].

Probabilistic broadcast protocols are designed to decrease the forwarding probability (p_f) as node density increases. Protocols estimate node density from neighboring nodes either by collecting copies of the already-transmitted broadcast packet or by receiving special control packets sent periodically for that purpose (i.e. Hello packets) [10]. A random uniformly distributed delay (defined in an interval [0,T]) is employed in these two types of protocols. In the former, it defines an interval for nodes to receive copies of transmitted packets. In the latter, waiting before transmission, decreases the probability of nearby nodes forwarding broadcast packets simultaneously.

After waiting a random delay, the broadcast protocol passes the broadcast packet from the network layer to the MAC (data link) layer. This contributes to reduce the collisions at MAC layers employing Carrier Sensing (CS) access mechanisms since packets are more sparsed in time. Reducing collisions is of paramount importance in ad hoc networks since collisions can lead to disconnected network segments, which affects the percentage of nodes covered by packet propagation (reachability).

However, the use of random delays results in slow packet dissemination. Therefore, there is a tradeoff between reducing collisions or reducing random delays, which impacts on the end-to-end latency [27] (latency for short). Since most of the protocols employ uniform distribution, changing the domain range of the uniform distribution has been considered the way to reduce latency, but at the expense of reducing reachability [37] [12].

This chapter shows that changing the distribution of random delays, from uniform to truncated-exponential, reduces the latency in probabilistic broadcast protocols without compromising reachability. Actually, for the same interval [0, T], the mean of random delays is reduced by using a truncated-exponential distribution when compared to using a uniform random delay. Given a multi-hop transmission in which successive random delays (one at each hop) occur along an end-to-end path, employing a distribution with a smaller mean results in lower latencies (i.e. the time elapsed between transmission and the packet reaching the last node). The effect of not compromising reachability is a consequence of the higher coefficients of variation (CV) of truncated-exponential random delays, which reduce the synchronization of transmissions at the MAC layer [123], avoiding collisions.

Reducing latency is essential to delay-sensitive applications [121]. Moreover, fast broadcasting of information compensates the uncertainty resulting from topological changes in the analysis of Vehicular Ad Hoc Networks (VANETs), which allows making the assumption that VANETs have a static random distribution of nodes [124].

This chapter is organized as follows. Related work is summarized in Section 5.3. Section 5.4 introduces an analytical framework to calculate the probability of collision events and assess the effects of using the proposed distribution. Finally, simulation results show that reductions in latency are achieved and yet reachability is maintained.

5.2 Table of Chapter Variables

Variable	Description
<i>T</i> :	Upper limit of the interval in which a random variable is defined; in this chapter either
	uniform or truncated exponential variables are defined in $[0, T]$.
p_f :	Forwarding probability.
V:	Length of the backoff window (in seconds) in the IEEE 802.11b standard.
$[V_{i-1}, V_i]:$	An interval of time of length V .
F(t):	Cumulative Distribution Function (cdf) of random delays; defined in $[0, T]$ in this chapter.
<i>I</i> :	Number of intervals $[V_{i-1}, V_i]$ in $[0, T]$; in this chapter T is always larger than V.
p_i :	Probability that the random variable with cdf $F(t)$ (defined in $[0, T]$) lies in $[V_{i-1}, V_i]$.
R_1 :	An event of the birthday paradox problem in which there are k distinct outcomes, and the
	first repeat occurs exactly in the $k + 1$ outcome.
f(t):	Probability density function (pdf) of random delays; defined in $[0, T]$ in this chapter; $f_e(t)$
	for truncated-exponential and $f_u(t)$ for uniform.
g(t):	Probability density function (pdf) of random delays conditioned on $[V_{i-1}, V_i]; g_e(t)$ for
	truncated-exponential and $g_u(t)$ for uniform
P_C :	Probability of two transmissions beginnig in the same slot, on the MAC layer, given that two
	random delays ended in the same interval $[V_{i-1}, V_i]$.
μ_e :	The mean of a truncated exponential random variable defined in $[0, T]$.
μ_u :	The mean of a uniform random variable defined in $[0, T]$.
α :	The ratio between μ_e and μ_u , namely $\alpha = \mu_e/\mu_u$.

Table 5.1 provides detailed description of the variables and functions that are used in the present analysis of collision events.

Table 5.1: Variables used in Chapter 5.

5.3 Related Work

In [12], a hybrid scheme (requiring Hello packets and collecting copies of broadcast packets) is proposed to avoid low reachability by using two uniform random delays with different ranges. In [27], the upperbound value of the interval (T) is different for each node and is defined as a function of the number of expected simultaneous forwarding nodes at each hop. The hybrid scheme in [125] uses a sequence of three random delays and increases the forwarding probability (p_f) as a function of the waiting time to receive a copy of a packet. However, these techniques rely on hello messages and are not applicable to other protocols.

Another strategy associates the duration of random delays to the inverse of the forwarding probability [126]. This strategy produces almost the same delay for nodes located in a neighborhood, allowing the use of strong assumptions about information availability, such as the Euclidean distance between nodes or the exchange of packets that contain the IDs of two-hop neighbors.

Overall, most of the strategies based on the use of uniform random delays are too specific to certain protocols, making them unsuitable to other broadcast protocols. Additionally, changes in the type of distribution of random delays have never been proposed in the literature [10].

5.4 Impact of the employment of exponentially distributed random delays on collision events

This section analyzes the effect of employing truncated-exponential random delays on packet collision events. It will be shown that the reduction of the mean of random delays leads to small variability in the probability of collision events.

One type of collision event that reduces reachability occurs when broadcast packets reach regions of the network which have not received any copy of this packet (unseen packet) and several nodes receive copies of this packet caused by the same wireless transmission [8]. The probability of collisions of nodes receiving the same unseen packet, virtually at the same time, can be computed considering two simultaneous events: i) the probability of having random delays at the network layer on two nodes ending with a time difference smaller than the duration of the backoff window at the MAC layer (overlapping backoff periods); and ii) the probability of two backoff timers at the MAC layer expiring in the same transmission slot.

For the analysis in this chapter, we use the IEEE 802.11b Distributed Coordination Function (DCF) MAC Layer [10]. Thus, we assume a minimum backoff window size of 32 slots. No acknowledgement mechanism is assumed [8]. Therefore, no retry counter exists to increment the size of backoff windows. We also use a value of T = 10ms, chosen for intervals of random delays (network layer) since it is the most frequent value used in various studies [10, 127], which helps the comparison of our results to other studies.

5.4.1 Probability of two overlapping backoff periods

Let [0, T] be the time interval in which the distribution of random delays is defined, and let V be the duration of the minimum backoff window at the MAC Layer ($V = 32 \times 20 \mu s$). With negligible differences in the propagation and processing times of the same unseen packet reaching several nearby forwarders, the random delays of these forwarders can be assumed to start at the same time. Dividing [0, T] into intervals of length V (hereinafter V-intervals), backoff intervals will overlap at the MAC layer if there are random delay values falling in a comom V-interval. Figure 5.1 shows the random delays of forwarders FW_1 and FW_2 ending in $[V_{i-1}, V_i]$ at the network layer. Even if FW_1 pasess the packet to its MAC Layer at the very beginning of $[V_{i-1}, V_i]$, a backoff window of length V will overlap with the beginning of the backoff window at the MAC layer of FW_2 . If two random delays end in adjacent V-intervals, their backoff windows may overlap. In this section, we focus on backoff windows that will overlap (Figure 5.1) and later show that the random delays of additional forwarders ending in adjacent V-intervals are negligible for the analysis.



Figure 5.1: Two forwarders passing broadcast packets to their MAC Layers in the same V-interval.

The probability of a random delay ending in $[V_{i-1}, V_i]$ is given by $p_i = F(V_i) - F(V_{i-1})$, where F(t) is the cumulative distribution function (cdf) of the random delays. For an integer number I of V-intervals, $I = \lceil T/V \rceil$, p_i is the probability mass function (pmf) of having a random delay ending in the *i*-th V-interval.

For an arbitrary distribution p_i over the set of V-intervals in [0, T], the probability of two forwarders choosing the same V-interval (R_1) , with exactly k + 1 forwarders, is obtained from the solution of the Birthday Paradox problem [128]:

$$P(R_1 = k) = \sum_{|A|=k} k! \Pi_A p_A,$$
(5.1)

where $\Pi_A = \prod_{i \in A} p_i$, $p_A = \sum_{i \in A} p_i$, and the summation spans all subsets A of kV-intervals. Then, when k + 1 forwarders contend, the probability of having any pair of random delays ending in a common V-interval is $P(R_1 \leq k) = \sum_{n=1}^k P(R_1 = n)$. Figure 5.2 compares the behavior of $P(R_1 \leq k)$ for T = 10ms and different distributions (p_i) obtained from uniform and truncated-exponential random delays. In Figure 5.2, exp07 μ and $\exp 05\mu$ indicate that p_i was obtained from truncated-exponential distributions. These distributions reduce the mean of random delays to 70% and 50% when compared to the mean delay given by the uniform distribution.



Figure 5.2: $P(R_1 \leq k)$ uniform vs. truncated exponential random delays

Figure 5.2 shows two aspects about the probability of having overlapping backoff periods. First, the cumulative probability distribution, $P(R_1 \leq k)$, reaches almost 1 with less than 10 forwarders. This means that overlapping backoff periods is almost a certain event in ad hoc broadcast since the average number of neighbors that leads to a connected network (99% of the times) is > 7.5 for nodes placed uniformly at random [1]. Second, as the number of forwarders grows, the relative increase in $P(R_1 \leq k)$ is reduced compared to the use of uniform random delays. For example, with 2 forwarders and $\exp 05\mu$, the probability of overlapping backoff periods almost doubles (an increase of 89%) whereas the increase is only 20% with 8 forwarders. With a higher number of forwarders, such increase tends to vanish, leading to a 50% reduction in the mean of random delays. Moreover, for exp07 μ the impact is even smaller (less than 6% increase for 8 forwarders).

When using the uniform distribution, the reduction in the mean of random delays yields higher values of $P(R_1 \leq k)$ compared to the use of the truncated-exponential distribution, i.e. the truncated-exponential p_i always produces a smaller increase on the number of collision events for the same reductions in the mean of random delays.

It is possible to extend this analysis to consider two or more V-intervals [128] (e.g. multiple pairs of random delays ending with time differences smaller than V). However, $P(R_1 \leq k)$ is sufficient to assess the impact of the truncated-exponential distribution since we are interested in the *relative* increase of collision events compared to uniform random delays. Moreover, the extended analysis including multiple overlapping V-intervals leads to the same conclusions.

5.4.2 Probability of two transmissions beginning in the same slot

We focus on the arrival of two packets in the same V-interval (Figure 5.1). In this section, the backoff timer assigned to packets by the MAC protocol is added to the arrival time of packets at the MAC Layer to determine the slot in which the transmission of a packet begins. As pointed out in [124], backoff timers should be non-persistent in ad hoc broadcast, therefore our analysis adds a backoff timer to every packet.

In an interval $[V_{i-1}, V_i]$, the probability density function (pdf) of the arrival time of packets at the MAC Layer is $f(t \mid V_{i-1} < t \leq V_i)$, where f(t) is the pdf of random delays in [0, T]. Denoting $g_u(t) = f(t \mid V_{i-1} < t \leq V_i)$ for uniform random delays, and $g_e(t)$ for truncated exponential random delays, we have:

$$g_u(t) = \frac{I}{T}, \qquad 0 < t \le T/I, \forall i$$
(5.2)

$$g_e(t) = \frac{\lambda e^{-\lambda t}}{1 - e^{-\lambda(T/I)}}, \qquad 0 < t \le T/I, \forall i$$
(5.3)

Notice that, for any i, $g_e(t)$ is the same since a truncated exponential distribution conditioned on equal-length subsets yields the same distribution for all subsets.

Since the MAC layer uses time slots of fixed duration s, $g_e(t)$ can be expressed as a truncated geometric distribution $g_e(s = s_a)$, i.e. the probability mass function (pmf) of a broadcast packet arriving at the MAC layer at $V_{i-1} + s_a$ is:

$$g_e(s=s_a) = \frac{p(1-p)^{s_a-1}}{1-(1-p)^S}, \qquad s_a = 1, 2, \cdots, S$$
 (5.4)

where S = 32, $p = 1 - e^{-\lambda s}$, and s = V/32 is the length of a slot. For uniform random delays, the equivalent pmf is $g_u(s = s_a) = 1/32$.

Upon arrival at slot s_a , the packet receives a random backoff timer of s_b slots, uniformly distributed on the backoff window. Then, the probability of two transmissions beginning in the same transmission slot (collision event) is the probability of having the same result $s_a + s_b$ for packets with overlapping backoff intervals (Section 5.4.1). This collision event is illustrated in Figure 5.1 at the MAC Layer for the packets of forwarders FW_1 and FW_2 .

Denoting $h(s = s_b)$ as the uniform pmf of backoff timers, the convolution $w_s = g(s) \circledast h(s)$ yields the pmf of the sum $s_a + s_b$; g(s) being either g_e or g_u .

Finally, collision events have probability $P_C = \sum_s w_s^2$, which considers all the cases the sum $s_a + s_b$ is the same for the two forwarders.

For T = 10ms and $\exp 05\mu$ at the network layer, the value of P_C (at the MAC layer) is $P_C = 0.0209$. This value of P_C is only 0.084% higher than the value of P_C obtained considering uniform random delays. The following proposition explains this small increase in the value of P_C considering reductions of up to 50% in the mean of random delays. The proposition leads to the conclusion that the impact of the change in the distribution can be estimated directly from $P(R_1 \leq k)$, defined in section 5.4.1, since collision probability values at the MAC layer remain virtually unaffected.

Proposition 1: Let $f_e(t)$ and $f_u(t)$ be the truncated-exponential and the uniform probability distribution functions, respectively, both defined in [0, T], with mean μ_e and μ_u . If $\mu_e = \alpha \mu_u$ for $0.5 \le \alpha < 1$, and $I \gg 4$ (I is the number of V-intervals in T), then, for (5.2) and (5.3), $g_e(t) \to g_u(t)$.

Proof. From $\alpha \mu_u = \mu_e$, we have:

$$\frac{\alpha}{2} = \frac{1}{\lambda T} - \frac{e^{-\lambda T}}{1 - e^{-\lambda T}},\tag{5.5}$$

which is a function of the product λT . As λT increases, $e^{-\lambda T}/(1 - e^{-\lambda T})$ decreases much faster than $1/\lambda T$ in (5.5), hence $\alpha/2 < 1/\lambda T$. Then, for $0.5 \le \alpha, \lambda T < 4$. Using the series expansion of the exponential function, and considering $I \gg 4$, (5.3) becomes:

$$g_e(t) = \frac{\lambda [1 - \lambda t + \frac{1}{2!} (\lambda t)^2 - \cdots]}{\frac{\lambda T}{I} [1 - \frac{1}{2!} \frac{\lambda T}{I} + \frac{1}{3!} (\frac{\lambda T}{I})^2 - \cdots]}$$
(5.6)

$$pprox rac{I}{T} = g_u(t), \qquad 0 < t \le T/I$$

Since $g_e(t) \rightarrow g_u(t)$, more-detailed analysis about increasing collision events at the MAC layer would be redundant (recall the collisions caused by random delays ending in adjacent V-intervals).

5.5 Simulation Results

Figure 5.3 shows simulation results, derived by using the QualNet(\mathbb{R})v.7.3 simulator to illustrate our claims. The network scenarios for comparison were taken from [6], which specifies: i) a square grid compared to nodes placed uniformly at random. Since a packet is first received from one of the neighbors, the square grid would have at most 3 forwarders, whereas the random scenario has 8 neighbors on average (number of forwarders); ii) the forwarding probability (p_f) values chosen produced sustained reachability (above 95%) to most nodes in the two scenarios (good reachability); iii) packets require at least 59 hops and more than 40 hops to reach the farthest nodes in the grid and random scenarios. With a higher number of hops, the impact of the reduction in the mean of random delays is more noticeable (multi-hop); iv) results in [6] assume no collision events. Our results are compared to an upper-bound of reachability (ideal conditions).



Figure 5.3: Reachability for different random-delay distributions. (a) Scenario with forwarders ≤ 3 . (b) Scenario with 8 forwarders on average.

The simulation scenarios in [6] were used to assess the performance of probabilistic broadcast scheme GOSSIP 1, with forwarding probability $p_f = 0.72$ for a grid, and

Table 5.2: Latency in two multi-hop scenarios, uniform vs. truncated-exponential random delays

	grid $p_f = 0.72$		random $p_f = 0.75$
Rn. delay	latency (in ms)	Rn. delay	latency (in ms)
distribution	with $95\%~{\rm CI}$	distribution	with 95% CI
uniform	402.9 ± 8.4	uniform	265.6 ± 5.1
$\exp 0.7 \mu$	307.9 ± 8.0	$\exp 0.6 \mu$	182.0 ± 4.3

 $p_f = 0.75$ for a random network. These fixed values of p_f guarantee that our results are compared involving the same average number of forwarders in each broadcast event (same rebroadcast).

This confirms that the mean of random delays can be reduced significantly with minimum impact on reachability by using truncated exponential random delays. In Figure 5.3a, the mean was successfully reduced to 70% when compared to the mean given by the uniform distribution, whereas in Figure 5.3b the mean was reduced to 60%, yet maintaining reachability. These results were not achieved using uniform random delays. The random scenario (Fig. 5.3b) leads to a greater reduction in the mean of random delays since the relative impact on reachability decreases as the number of forwarders increases, as explained in Section 5.4. [0,T] The case of $\exp 0.5\mu$ shows the extent to which reachability is compromised when the mean of random delays is reduced to 50%. However, if the reachability obtained with $\exp 0.5\mu$ were tolerable, our proposal would be effectively applied by decreasing the value of T and using truncated exponential delays to maintain reachability, reducing broadcast latency even further.

Table 5.2 shows the average time elapsed until reaching the last node (latency) in simulation runs with reachability over 0.85. The table compares uniform vs. the truncatedexponential random delays that maintained the same reachability level in Figure 5.3. The reductions in latency were smaller than the expected (23.6% instead of 30%, and 31.5% instead of 40%). This indicates that, when reducing the mean by using truncated exponential random delays, the Carrier Sensing (CS) function of the MAC layer deffers more transmissions due to detection of a busy medium. Indeed, a higher number of successful detections of a busy medium contribute to sustained reachability, as was anticipated in Section 5.1.

5.6 Chapter Summary and Future Work

The present work showed a way to reduce latency when using probabilistic broadcast protocols. Results show that latency was reduced and reachability maintained for the same rebroadcast, which is not possible by using the wide-spread uniform distribution. Such reduction was achieved by decreasing the mean of random delays by using the truncatedexponential distribution. Moreover, such distribution can be applied to any probabilistic broadcast protocol for ad hoc networks, including VANETs. Future work should focus on testing the use of truncated exponential random delays in a different probabilistic broadcasting schemes.

5.7 Table with Related Work

Table 5.3 summarizes the studies which have explicitly mentioned the reachability-latency tradeoff and which have proposed a strategy to circumvent the low reachability when random delays are defined in intervals of few milliseconds.

The table shows the schemes that use Hello packets, that is, schemes based on neighbor information (neighbor-based). These schemes employ random delays to avoid packet collisions, increasing reachability.

Also, the table indicates the studies in which broadcasting schemes heavily rely on the target application. For example, in [126], the forwarding probability is obtained from estimations of the distance between nodes that assume location services as well as the exchange of node-state information. These assumptions are justified specifically for vehicular networks and would not be applicable to other types of ad hoc networks.

	Hello	Specific	Strategy
	Pkt.	App.	
Drabkin [12]	Yes	No	Two Uniform rnd. delays
Hyocheol [27]	Yes	WSN	Variable interval $[0, T]$ with number of neigh.
Nourazar [125]	Yes	WSN	A sequence of 3 random delays
Mostafa [126]	No	VANET	Variable interval $\left[0,T\right]$ as the inverse of p_{f}
Forero [129]	No	No	Truncated exponential random delays in $\left[0,T\right]$

Table 5.3: A comparison of studies adressing the reachability-latency tradeoff in the literature.

Chap	ter
\mathbf{SI}	Χ

Distribution of Latency in Ad Hoc Probabilistic Broadcast

This chapter computes the probability density function of broadcast latency in probabilistic broadcast schemes. The probability density functions of latency are obtained from the random delays used in protocols and the geometry of the position of nodes. The resulting probability density functions are compared to simulation results to test the accuracy of the analysis.

6.1 Introduction

Broadcasting packets in an ad hoc network is a challenging task due to the absence of a central control in the shared wireless media. The simplest approach to ad hoc broadcasting consists in allowing nodes to forward every received packet once. This technique, known as *flooding*, successfully delivers broadcast packets in small networks, where the number of forwarded copies of a single broadcast packet is not enough to saturate the wireless channel. However, in large, multi-hop, ad hoc networks, flooding leads to inadequate redundancy [8], increasing packet collision events and compromising packet dissemination.

One way to avoid the redundancy of *flooding* is to use a probabilistic approach to reduce the set of forwarders; that is, nodes should forward broadcast packets according to a random variable drawn on a per-packet basis [10]. Unlike deterministic techniques [130–132], probabilistic broadcast schemes yield intrinsic rotation of the subset of forwarders at every broadcasting event, evenly distributing the workload of packet dissemination. Furthermore, probabilistic protocols demand low overhead and yet produce a fully distributed, dynamic selection of the set of forwarders.

Probabilistic broadcast protocols are designed to decrese the probability of forwarding a packet (p_f) as node density increases; therefore, node density around forwarders must be estimated. Forwarders estimate node density either by waiting to *hear* copies of broadcast packets from neighbors [6, 40] or by using periodic control (Hello) packets [1, 12]. Waitbased probabilistic protocols introduce a random delay after receiving a broadcast packet so that contending forwarders can hear each other properly [96]. In protocols that use Hello packets, nodes maintain updated information from neighbors so that probabilistic forwarding can happen immediately; however, these protocols also use random delays before transmissions to avoid collisions due to synchronized forwarding [12]. The main tradeoff in probabilistic broadcast lies in using the values of p_f that minimize the subset of forwarders (rebroadcast) and maximize the number of nodes receiving every broadcast packet (reachability). Thus, most of the literature explores different ways in which node-density estimations can be used to compute p_f [10]. However, only few studies have focused on providing analytical models that capture the relation between protocol parameters and broadcast metrics.

Existing analytical models for probabilistic ad hoc broadcast protocols include estimation of the value of p_f for both wait-based schemes [96] and schemes based on hello packets [1,12]; closed-form expressions to compute the reachability, saved rebroadcast and average latency of wait-based protocols [97,98]; and estimations of reachability as a function of broadcast rates [99]. These models have revealed the impact of protocol parameters on the reachability-rebroadcast tradeoff as well as on the convergence of the values of p_f as a function of node density. However, most studies have focused on finding average values. Only in [1], a discussion about random geometric graphs [2] is provided to point out that the probability distribution function of broadcast reachability for probabilistic protocols can follow a bi-modal pattern [6] in which either most nodes or very few nodes are reached at every broadcast event (i.e. probability accumulates near 0% and near 100%, but not in the middle of the reachability range).

The present chapter provides analytical expressions to compute the probability density function (pdf) of broadcast latency for probabilistic broadcast protocols. The resulting models permit observation of the way broadcast latency distributions change as a function of p_f , network geometry and different types of random delays. Two types of geometries are considered, grids and static nodes placed at random; then, the results are mostly relevant for terrestrial static sensor networks [102, 103] and environmental applications [109, 110]. However, since probabilistic broadcast can be much faster than the speed of nodes [124], the analysis on random networks can be used to estimate the distribution of broadcast latency in mobile scenarios [115]. Also, the analysis allows computing the pdf of latency for any distance between the source and the farthest node, which is equivalent to computing latency for networks of different size.

The probability density function of broadcast latency provides the necessary information to predict the level of service that an ad hoc network can meet, especially considering that time constraints on route discovery are fundamental in ad hoc networks to meet the demands of applications [133]. Furthermore, computation of the broadcast-latency distribution at an arbitrary distance from the source can be used to callibrate routingparameter values such as time-outs and TTL, e.g. in the evaluation of reactive ad hoc routing protocols supporting delay-sensitive services [121].

The present chapter is organized as follows. Related work on analytical models of probabilistic broadcast is summarized in Section 6.3. Section 6.4 is concerned with deriving and validating the probability distribution of the path length of broadcast packets. Section 6.5 uses the results in Section 6.4 to yield the pdf of probabilistic-broadcast latency. Percentiles obtained from the models are compared to simulation results in Section 6.6 and conclusions are drawn in Section 6.7.

6.2 Table of Chapter Variables

Table 6.1 provides detail description of the variables that appear in this chapter. Some of the variables may have a continuous and discrete interpretation. For example, x represents discrete distance (in hops) and continuous (Euclidean) distance. Readers should notice that variables with discrete and continuous interpretation correspond to the analysis of grids and random networks, respectively.

Variable	Description
p_f :	Forwarding probability.
T:	Upper limit of the interval in which a random variable is defined; in this chapter either uniform or
	truncated exponential variables are defined in $[0, T]$.
x:	Represents distance. For the discrete model (grids) it is the minimum distance in hops from a
	roadcast source; it is equivalent to the value of a state in the random-walk model. In the continuous
	model x represents Euclidean distance.
p:	Probability of packets moving one tier forward. It is equivalent to the probability of a step forward
	in a one-dimensional random walk model.
q:	1-p in the one-dimensional random-walk model.
y:	Number of hops taken by a broadcast packet to reach a distance $x = x_0$ from the source. It is
	equivalent to the index of the random walk (random process).
k:	Number of steps forward in a one-dimensional random walk model for a given index (y) of the random
	process.
N:	Total number of nodes in an ad hoc network (network size).
r:	Transmission radius of nodes. All nodes are assumed to have the same value of r .
n_r :	Average number of nodes in the transmission radius of a node.
l:	Euclidean distance between the broadcast source and distant nodes (i.e. nodes reached by multi-hop
	communication.
h_{av} :	Average number of hops between nodes at Euclidean distance l .
<i>h</i> :	Regardless of the subscript, h represents the number of hops.
n:	Any discrete random variable with probability mass function (pmf) $f(n)$. The random variable for
	the distribution of the number of hops in the random sum computation

Table 6.1: Variables used in Chapter 6.

6.3 Related Work

An analytical model to find the value of p_f in wait-based schemes was proposed in [96]. The authors showed the way p_f decreases as the number of nodes increases in a given area. The formulas to solve for the forwarding probability (p_f) were of the form $p_f = f(p_f, N)$, where N is the number of nodes in the network; thus a computer package was necessary to obtain the values of p_f .

Also for wait-based schemes, in [97,98], the authors provide expressions for the expected values of performance metrics such as reachability, saved rebroadcast and average broacast delay (i.e. latency) as a function of protocol parameters (e.g. the number of packet copies that should be heard to refrain from forwarding).

Analytical models were presented in [99] to predict the reachability of *flooding* and a scheme with a fixed value of p_f . The study analyzes probabilistic broadcast considering the effects of the carrier sensing (CSMA) mechanism of the 802.11 wireless standard [9] and the traffic rate. The probability of a broadcast source releasing packets into the wireless channel is calculated. The study shows the limits of traffic-driven probability of transmission of packets; above these limits, the excess of traffic reduces reachability to arbitrarily low levels either using flooding or probabilistic broadcasting.

For schemes relying on Hello messages, in [1], the theory of Random Plane Graphs [2] and Continuum Percolation [134] is used to prove that, in a network with nodes placed according to a two-dimensional Random Uniform distribution, the forwarding probability at every node can be set to the inverse of the node's number of neighbors, namely $p_f = \beta/n$, where n is the number of one-hop neighbors, and β is simply a constant (called coverage level) that is proportional to the expected reachability. Since the theory applies to networks in which the number of nodes tends to infinity, a simulation-based estimation of the behavior of β was necessary to model finite networks.

The work in [12] presents an analytical bound on the probability of messages failing to reach any arbitrary node in the network. In line with the ideas in [1] (i.e. $p_f = \beta/n$), the analytical bound suggests that if β is the average number of one-hop neighbors allowed to repeat broadcast messages, a calculation of worst-case reachability is immediate, and this constant determines the expected value of reachability, as in [1].

Besides providing *average* values of broadcast metrics, common assumptions in the studies above include nodes placed at random according to the Uniform distribution, and random delays Uniformly distributed in an interval [0, T]. In this regard, the contribution of the present work is threefold:

- Computation of probability density functions.
- Evaluation of broadcast latency considering random-delay distributions different from the Uniform distribution in [0, T].
- Latency models for node-position geometries different from random Uniform.

The following section derives probability mass functions (pmf) of the path length of broadcast packets for different geometrical scenarios (grids and nodes placed at random). The results are used in Section 6.5 to compute the probability density functions of broadcast latency including different types of random delays.

6.4 Path Length According to Node Layout

The purpose of this section is to derive the probability mass function (pmf) of the path length that is traversed by a broadcast packet. The resulting distribution depends on the geometrical features of node layout as well as on the forwarding probability (p_f) . First, path-length probability mass functions are derived for regular geometries, and then for random node layouts.



Figure 6.1: In grids, the random process that determines the path length of broadcast packets is the same regardless of the direction of packet propagation.

6.4.1 Regular Geometries (Grids)

6.4.1.1 Path-length Model

Let x be an integer variable that represents the graph distance between the broadcast source and any other node in a grid. Then, regardless of the position of the source, sets of nodes sharing the same value of x will form tiers that expand away from the source until reaching the farthest nodes.

Ideally, the path length of broadcast packets should be equal to the graph distance However, the present model considers the non-ideal path length produced by the randomness of probabilistic broadcast schemes. For the model, the ideal graph distance between nodes and the broadcast source becomes the minimum number of hops and indicates the tier of each node.

In grids, the transition of a broadcast packet from one tier to the next occurs under the same topological conditions regardles of the value of x and the direction of propagation. In this context, we propose to model the distribution of the path length of a broadcast packet as a one-dimensional (1D) random process conditioned on a specific value of x; formally, a random variable Y = y conditioned on $x = x_0$. Figure 6.1 summarizes this idea.

In Figure 6.1, tiers are separated by dashed lines, shown as concentric circles arround the source, and the same random process is shown to occur in three different directions for x_0 taking integer values i, j and k. The area of the grid is a square. To the right of the square region, a diagram presents tiers separated by straight dashed lines, emphasizing that the 1D random process depends mainly on the probability of packets going forward in one tier (with probability p); the random variables X and Y are mapped onto the tier of interest and the number of hops (arrows), respectively.

Given the discrete nature of tier formation and the 1D consideration, the distribution of Y will depend mostly on the probability (p) of moving forward one tier. Due to the regularity of grids, the value of p will be assumed constant and can be estimated from the local topology of nodes (i.e. a single tier-transition pattern to determine the value of p). Based on the previous description, we map the variables x, y and p to a one-dimensional random walk [135,136] with discrete state space (range of X) and discrete index (value of Y).

The state of the random walk is described by $X = \sum_{i=1}^{y} Z_i$, with $P(Z_i = 1) = p$ and $P(Z_i = -1) = 1 - p = q$, and Z_i independent and identically distributed random variables (i.i.d. r.v.). Given an index Y = y, the distribution of X is obtained from [135, 136]:

$$P(X = x \mid y) = {\binom{y}{k}} p^k q^{y-k}, \qquad y \ge x_0 \tag{6.1}$$

where k is the number of times $Z_i = 1$, leading to x = k - (y - k) = 2k - y; hence k = 1/2(y + x) in (6.1). In this model, P(X = x | y) exists for even values of x when y is even, and the same relation holds for odd values of y. In our mapping (Figure 6.1), at tier x, the path length corresponds to the value of y, then the pmf of the path length can be obtained from:

$$P(Y = y \mid x) = \frac{P(X = x \mid y)P(y)}{\sum_{y \ge x} P(X = x \mid y)P(y)}$$
(6.2)

In (6.2), P(y) = 1 for any value of y assuming that broadcast is never stopped by time, or by a limitted number of hops, i.e. as long as there are nodes receiving packets, packets continue propagating and the index of the radom walk shall increase.

The following sub-sections analyze the local topology of various grids to determine the corresponding value of p and validate the use of (6.2) to estimate the distribution of the path length of broadcast packets in grids.

6.4.1.2 Square Grid

Figure 6.2 shows the topological tier-transition pattern for packets to move forward in a square grid. Figure 6.2a shows the case in which all wireless links (arrows) occurred from left to right i.e. all nodes forward the broadcast packet. Figure 6.2b shows the situation in which node **A** does not receive any copy of the packet because its two *expected* transmitters in tier *i* refrained from forwarding due to p_f . Figure 6.2c shows what may happen if node **A** receives the packet from one of the nodes one tier ahead, a packet *hops back* from Tier i + 2 to Tier i + 1, resulting in a path that can go back and forth between tiers.

Considering probabilistic decisions as the only cause of packets hopping back to the previous tier, the value of p in (6.1) corresponds to $p = 1 - (1 - p_f)^2$. Notice that in a square grid each node has only two expected transmitters in the previous tier; then, for a collision event, the two colliding packets should have the same effect shown in Figures 6.2b and 6.2c. Hence, collisions can be included in the model by decreasing the value of p with an estimation of the probability of collisions.

Packets going back two (or more) tiers would be restricted in realistic scenarios due to the sequence numbers of broadcast protocols, namely in Figure 6.2c the two nodes that refrained from forwarding in tier *i* will not repeat the packet forwarded by **A** because they recognize it as previously seen. However, since $p_f > 0.7$ for good reachability in square grids [6], consecutive occurrences of $Z_i = -1$ in the model would introduce a small distorsion.



Figure 6.2: Topological pattern of the propagation of a broadcast packet over a square grid regardless of the position of the source. a) Ideal pattern with all links going forward one tier. b) Node A in Tier i + 1 does not receive the packet due to probabilistic decisions in tier i. c) A path that goes back one tier is formed.



Figure 6.3: Histogram vs. pmf of path length at Tier 10 in a square grid.

Figures 6.3–6.5 show the results of using (6.2) for different values of x (different tiers) in a 50 × 50 square grid. The results are compared to the histograms obtained from simulations using QualNet(\mathbb{R} v.7.3. [119]. Simulation settings appear in Table 6.2. Each graph shows the value of p_f and the sample size (*SmplSz*) of the histogram. For 120 simulation runs, histograms combined the data from two nodes in the same tier with symmetric positions so that sample sizes could grow up to 240 samples.

The values of x in Figures 6.3–6.5 were chosen to illustrate that the model can be applied to networks of different sizes. Since the source was placed on one of the sides (first column, center row), a value of x = 10 is equivalent to having a network with 121 nodes; similarly, x = 30 involves 925 nodes and x = 50 involves 1924 nodes. Notice that both the histogram and the pmf exist for even values of the path length, given that x is even, which supports the assumptions of the model.

Figures 6.3–6.5 show a good fit of the analytical model with the simulation results. The qualitative behavior of the path-length pmf as a function of p_f was properly captured by the model. When p_f is low, longer paths occur, shifting the probability mass function



Figure 6.4: Histogram vs. pmf of path length at Tier 30 in a square grid.



Figure 6.5: Histogram vs. pmf of path length at Tier 50 in a square grid.

Parameter	Value
Grids	50×50 square, hexag-
	onal and triangular
Broadcast Scheme	GOSSIP1(p,4) [6]
Transmission Range	250m approximately
Transmission Rate	2Mbps
Path-loss model	Two-ray
Transmission Rate	2Mbps
Mac Layer	802.11b
Source Position	row 25 column 1
Runs	120 replications

Т	abl	le	6.2	:	Simulation	settings.
---	-----	----	-----	---	------------	-----------

Square Grid pmf GoF								
p_f	Tier	50	Tier	30	Tier 10			
	P-value Bins		P-value	Bins	P-value	Bins		
0.77	0.63	6	0.11	4	0.02	3		
0.72	0.67	6	0.01	4	0.10	2		
0.67	0.35	8	_	4	0.06	3		

Table 6.3: Chi-square goodness of fit (GoF) test; pmf of path length in 50×50 square grid.

(pmf) to the right and also producing a longer tail. Also, the model captures the changes in the pmf of path length for nodes at different tiers (or networks of different sizes).

Since the graph distance is a discrete random variable, the Chi-square goodness-of-fit (GoF) test was applied to validate the analytical model. The results for different values of p_f and x (tier), appear in Table 6.3. According to the P-values of the test, the analytical pmf represents the data collected from the simulation experiments. Considering a 0.01 significance level, the tests fail to reject the proposed model [137].

Table 6.3 also shows the number of bins in the GoF test for each case. Since GoF tests must have a minimum number of expected observations in each bin [137], the number of bins is reduced for the cases of tiers near the source as both the pmf and the histogram accumulate most of the probability in a shorter range (e.g. number of bins in Figure 6.3).

The results reported in Table 6.3 were sensible to small changes in the value of p. For x = 50, decreasing the value of p by less than 0.01 produced a better fit, yielding a pmf with longer tail. Conversely, for x = 30, x = 10, increasing p by adding values between 0.002 and 0.02, increased the P-value of the GoF tests. These small variations around p can be explained by considering the effects of packet collisions at the MAC layer (decreasing p far from the source) and the specific probabilistic protocol employed in the simulations (Table 6.2), in which probabilistic broadcast is avoided in the first 4 hops.

The difference between the model and the histograms in the first bins is more noticeable for x = 30 (Figure 6.4). Only in the case x = 30 with $p_f = 0.67$ the first bin caused the GoF test to fail. The P-value without the observations of the histogram at y = 30 was 0.29, showing an excellent fit of the tail. Overall, the model represents the behavior of probabilistic broadcast in square grids.



Figure 6.6: Topological pattern of the propagation of a broadcast packet over a hexagonal grid. a) Ideal pattern with all links going forward one tier. b) node **A** receives the the packet from tier i + 3 (path with two hops back), a path blocked by the sequence numbers of packets; node **B** receives the packet from tier i + 2 without blocking.

6.4.1.3 Hexagonal Grid

Figure 6.6 shows the topological tier-transition pattern of broadcast propagation for the case of a hexagonal grid (or honeycomb). Similar to the analysis of the square grid, Figure 6.6a shows the case in which all links occur from left to right. The pattern for nodes at the top of the figure is totally symmetric to the pattern at the bottom. These two equivalent patterns correspond to the propagation of broadcast packets towards different ends of the grid, which occurs as long as the broadcast source is not in one of the corners of the grid. When the source is in a corner, only one of the patterns would emerge. However, due to symmetry, the following analysis would be the same regardless of the possition of the source. In all cases, for any arbitrary tier, a node with one expected transmitter from its previous tier is followed by two expected receivers in the following tier (see the two nodes at the botom of Tier i + 1). This pattern is always followed by nodes with two expected transmitters and only one receiver (the two nodes at the botom of Tier i + 2). The two patterns always occur one after the other until the last reachable tier.

Figure 6.6b illustrates two cases that match the two patterns above. First, the two expected transmitters of node **A** decided not to forward the packet due to p_f . Since high reachability is assumed, node **A** may receive the message from a path that has hopped back two tiers, from Tier i + 3 to Tier i + 2, to node **A** in Tier i + 1. However, this path must end at node **A** because its three neighbors already received one copy of the message, so the sequence number of the packets will block the path completely.

In the second case of Figure 6.6b, the single expected transmitter of node **B** decided not to forward the broadcast packet due to p_f , and node **B** receives the packet from a node in the following tier (from Tier i + 2). Unlike node **A**, node **B** can easily continue to propagate the message forward since node **B** is the only one expected transmitter of the following node in Tier i + 2. In this case, the path only goes back and forth by one hop, as in the case of the square grid (Section 6.4.1.2).

The two cases in Figure 6.6b indicate that, for any given path, only in half the tiers can we expect packets to go back and forth *one hop*. In general, half the nodes (nodes with the local topology of node **B**) contribute to the formation of long paths in a hexagonal grid. Hence, the probability distribution of the path length for the hexagonal grid can be obtained from equation 6.2 with the following considerations; firts, the value of x

Hexagonal Grid pmf GoF								
p_f	Tier	64	Tier	32	Tier 16			
	P-value Bins		P-value	Bins	P-value	Bins		
0.85	0.46	13	0.03	7	0.12	6		
0.8	0.04	13	0.21	8	0.07	7		
0.75	0.12	8	0.57	8	0.19	6		

Table 6.4: Chi-square goodness of fit (GoF) test; pmf of path length in 50×50 hexagonal grid.

corresponds to half the number of tiers, so the other half of the tiers are included only by shifting the resulting pmf to the right; second, the value of p is now p_f since only one forwarder is involved in allowing paths to hop back.

Figures 6.7 – 6.9 show the histograms obtained from simulations of a 50×50 hexagonal grid compared to the pmf from equation 6.2 with the considerations above. As in the case of the square grid, the pmf succeeds in describing the behaviour of the path length for different values of p_f and tiers (or network sizes). As in the case of the square grid, the values of p_f were chosen around a critical value in which reachability can suddenly drop from over 90% to 20%. Consequently, the sample sizes (SmplSz) of histograms with $p_f \ge 0.8$ indicate that nodes were reached in almost all runs, whereas the sample sizes for $p_f = 0.75$ reveal that nodes were reached only 30% of the times, confirming that there is a critical value of p_f for reachability in grids [7].

GoF results for the hexagonal grid are shown in Table 6.4. Again, the tests fail to reject the proposed model at a 0.01 significance level.

Unlike the square grid, significant probability values in the hexagonal grid spread over a wider range of y (the path length), which allows more bins in GoF tests. Also, in terms of the sensitivity of parameter p, better fits can be obtained increasing p by 0.01–0.02. Only in the case of $p_f = 0.75$, Tier 64, better fits were obtained decreasing the value of p. This trend indicates that the model predicts longer paths for all tiers, i.e. the model is conservative for hexagonal grids. However, the lack of fit between the pmf and the histograms in the first bins never caused GoF tests to fail. Indeed, the P-values in Table 6.4 indicate that the model is more accurate for the hexagonal grid than for the square grid for nodes near the source (or smaller networks).

Again, the assumptions of the model are well supported by the histograms since only even values of y appear in histograms, following the fact that x is even (i.e. Tiers 16, 32, 64).

6.4.1.4 Triangular Grid

Figure 6.10 shows the topological tier-transition pattern of a triangular grid. As in the case of the hexagonal grid, the figure shows a section of the network where packets can propagate towards two different ends of the network area. Figure 6.10a shows how the nodes at the top and bottom share the same pattern, namely two expected transmitters from the previous tier, two expected receivers in the following tier and two expected receivers *in the same tier*. Figure 6.10b illustrates two events that may occur when the two expected transmitters of a node decide not to forward the broadcast packet due to p_f . First, node **A** receives the packet from a node in its own tier, describing a path that



Figure 6.7: Path length in hexagonal grid, pmf vs. histogram, Tier 16.



Figure 6.8: Path length in hexagonal grid, pmf vs. histogram, Tier 32.


Figure 6.9: Path length in hexagonal grid, pmf vs. histogram, Tier 64.



Figure 6.10: Topological pattern of the propagation of a broadcast packet over a triangular grid. a) Ideal pattern with links going forward one tier and also links between pairs of nodes in the same tier. b) Node **A** receives and sends the packet within the same tier, leading, a path that stays in the same tier two additional hops; the path through node **B** only stays in the same tier one hop.

remains in the same tier after an additional hop (see Tier i + 1). Then the path continues in the same tier when node **A** transmits the packet to its neighbor below, so that the path stays in the same tier after two hops.

In Figure 6.10b the situation of node **B** is slightly different. The path that goes through node **B** only stays in the same tier for one additional hop because the node above **B** received the packet from its other neighbor in tier i + 3 and not from **B**. To evaluate the model, we assume that the situation of node **A** (i.e. two hops in the same tier) is very unlikely because it implies that the node below **A** did not receive the message from at least other two neighbors (i.e. neighbors from below in Figure 6.10b). Then, the situation of node **B** should dominate the formation of paths in triangular grids.

According to the previous considerations, the pmf for the path length in a triangular grid should be calculated based on a random walk model in which $P(Z_i = 1) = p$ and $P(Z_i = 0) = 1 - p = q$. By including this modification, now $y - x_0 = j$, where j is the



Figure 6.11: pmf vs. histograms of path length in a triangular grid, Tier 10.

Triangular Grid pmf GoF									
p_f	Tier 50Tier 30Tier 10								
	P-value Bins P-value Bins P-value B								
0.72	0.33	7	0.41	10	0.24	4			
0.65	5 0.18 5 0.23 8 0.02								
0.6	0.20	7	0.97	5	0.11	3			

Table 6.5: Chi-square goodness of fit (GoF) test, path length in 50×50 triangular grid.

number of times the random walk stayed in the same state (i.e. the number of times $Z_i = 0$ occurred). Notice that now x and y can be even or odd independently, meaning that the histograms may show non-zero values (observations) in all possible values of y (even or odd) regardless of the value of x. As in the case of the square grid, $p = 1 - (1 - p_f)^2$ (i.e. two nodes must decide not to forward a packet to make the path stay in the same tier).

The pmf for a 50 × 50 triangular grid is compared to the histograms obtained from simulations in Figures 6.11–6.11. Once again, the values for p_f were chosen around a critical value of reachability [7], which can be confirmed by the sample sizes.

The corresponding GoF tests are shown in Table 6.5. The trend for a better fit of the model was to reduce the value of p by 0.06–0.12 in all cases. This can be explained by the intra-tier paths that may form independent from probabilistic broadcast and the increase in collision events caused by a larger number of neighbors (a denser grid).

As in the case of the square grid, for nodes near the source (i.e. the case of Tier 10 in Figure 6.11), histograms show a larger-than-expected number of observations exactly in y = 10, which can be attributed to the non-probabilistic forwarding decisions in the first four tiers.

Overall, equation 6.2 produces probability mass functions that fit the data from simulations for different values of p_f , different distances from the source (network sizes), and different grids.



Figure 6.12: pmf vs. histograms of path length in a triangular grid, Tier 30.



Figure 6.13: pmf vs. histograms of path length in a triangular grid, Tier 50.



Figure 6.14: A path from the two more distant nodes in a network compared to the Euclidean distance between the nodes.

6.4.2 Random Geometries

Similar results to those in Section 6.4.1 can be obtained for networks with nodes placed at random. Specifically, this section focuses on the path length pmf for nodes arranged according to a homogeneous Poisson process (also called Poisson Point Process). The results in this section can be used for the case of N nodes independently and uniformly distributed in the same region due to its theoretical equivalence with the homogeneous Poisson process [138].

The pmf of path length is derived from two ideas. First, the authors in [139] proved that the graph distance (i.e. distance in hops) for nodes in a Poisson process is only a constant factor larger than their Euclidean distance, as long as nodes are suficiently distant. Second, for the same Poisson process, the results gathered in [138], show that the Euclidean distance to the *n*-th neighbor follows an Erlang distribution. Since the graph distance is a linear function of the Euclidean distance [139], the pmf of the path length should belong to the same family of probability density functions (pdf) that describe the Euclidean distance between to nodes (i.e. and Erlang-based pmf), only modified a constant scale factor.

6.4.2.1 Estimating the Constant

A first step to finding a constant that relates the Euclidean distance to the graph distance is to estimate the number of hops that cover such a distance. Given a node, the positions of its neighbors in a Poisson Point Process are equivalent to having neighbors Uniformly distributed [138]; then, assuming that the transmission area is a circle of radius r, neighbors can be found at a distance x from the transmitting node, in any arbitrary direction, with the following pdf [8]:

$$f(x) = \frac{2\pi x}{\pi r^2}, \qquad 0 < x < r \tag{6.3}$$

According to equation 6.3, the expected value of the distance between the transmitter and any of its neighbors (i.e. only nodes within r) is E[X] = 2r/3. This means that, if a series of nodes were arraged in a straight line of length l, the average number of hops (h_{av}) from a source, placed at the beginning of the line, to the farthest node would



Figure 6.15: Expected angular area (sector), for a single neighbor, regardless of the distance to the center. a) single node appearing, on average, at half the arc of the expected sector. b) two contiguous sectors (dashed lines) with their nodes at their average angular positions. Either above or below the horizontal line neighbors will appear, on average, at $\theta/2$.

be $h_{av} = 3l/2r$ (which coincides with the number of transmitters in a one-dimensional network using the counter based scheme in [98]).

Although the assumption of a pseudo-straight path closely following the line of the Euclidean distance (between suficiently distant nodes in an *m*-dimensional random network) is not intuitive, the proof of Theorem 3 in [139] shows that the detours from an ideal path only increase the graph distance by a constant factor. Figure 6.14 shows a graphical representation of this idea in two dimensions. The figure shows one realization of a Poisson process in the square $[0, 1]^2$ where the average number of nodes is 2500. The most distant nodes are near the corners of the square, so the length of the black line joining them is nearly $\sqrt{2}$. A gray thick line shows the path (graph distance) between the two nodes. As predicted, the detours of the straight line are moderate. Additionally, the path deviates to both sides of the straight line equally, either crossing the line or getting very close to it repeatedly.

With this idea in mind, the constant can be obtained by projecting the Euclidean distances associated to each hop over the ideal straight line. Here, the projection angles will depend on node density; therefore, different constants can be found as a function of the parameter p_f of probabilistic broadcast schemes (i.e. the lower p_f the lower the node density that produces the path).

A more intuitive explanation is shown in Figure 6.15. Suppose the average number of nodes in a circle of radius r is n_r . Then, around a node in the center of the circle, on average, one neighbor should appear every $\theta = 2\pi/(n_r - 1)$ radians. This means that, within a given sector of the circle, with arc-length θ , the average angular position of a neighbor is $\theta/2$. The position of node A in Figure 6.15a corresponds to this average angular position within a shaded sector. Figure 6.15b shows how a deviation of node \mathbf{A} , in more than $\theta/2$ from the horizontal line, implies that another node (node **B**) is very likely to appear at around $2\pi - \theta/2$. Notice that both nodes, A and B, continue appearing at their own average angular position for two contiguous sectors (sectors shown with dashed lines). If node A deviates by more than $\theta/2$, the expected location of node B would be closer to the horizontal line and vice versa. Therefore, a multi-hop path that reaches the farthest node of a network will fluctuate around the ideal trajectory with an average angle deviation of $\theta/2$. Therefore, the role of probabilistic broadcast is to reduce the average number of neighbors from $n_r - 1$ to $n_r p_f - 1$; then, as p_f decreases, the arc-length of sectors increases, resulting in larger average angular deviations of the path from the straight line, i.e. $\theta = 2\pi/(n_r p_f - 1)$.



Figure 6.16: Projection of the distance in hops over the Euclidean distance (horizontal line). After four hops, the arrows indicate the progress over the Euclidean distance, which corresponds to the distance between nodes multiplied by $cos\theta$.

In summary, the average number of hops to reach the farthest nodes of a network, using probabilistic broadcast, can be calculated from the Euclidean distance to the source (l) as:

$$h_{av} = \frac{3l}{2r\cos(\theta/2)},\tag{6.4}$$

where the $\cos \theta/2$ factor represents the projections of the path over the Euclidean distance, and depends on p_f . Figure 6.16 shows a magnified segment of the path shown in Figure 6.14. The arrows indicate the projection of each hop over the Euclidean distance between the two more distant nodes, while solid lines represent the actual path traversed by a broadcast packet after four hops.

6.4.2.2 Finding the pmf

The previous analysis is used now to find an Erlang-based distribution that describes the pmf of the path length of broadcast packets when nodes are placed according to a homogeneous Poisson process. From equation 6.4, the minimum value for the average number of hops (h_{min}) , given a distance (l) is obtained with $\theta = 0$ (i.e. a path with no deviation from a straight line). Then, for different values of p_f an average expected value h_{av} is calculated. Since the Erlang probability density function is of the form:

$$f(x;k,\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, \qquad x,\lambda \ge 0$$
(6.5)

with expected value $\frac{k}{\lambda}$, where k must be an integer, a pmf can be fitted from h_{min} and h_{av} by inspection on integer values of k. Specifically, the Erlang-based pmf to describe the path length is obtained from the differences between integer values of the Erlang's cumulative distribution function (cdf), that is:

$$f(n;k,\lambda) = \sum_{m=0}^{k-1} \frac{1}{m!} \left([\lambda n]^m e^{-\lambda n} - [\lambda(n+1)]^m e^{-\lambda(n+1)} \right)$$
(6.6)

where $n = h - h_{min}$, and $\lambda = k/(h_{av} - h_{min} + 1)$, with $h \ge h_{min}$ taking integer values.

Figures 6.17 – 6.19 show the pmf obtained using Equations 6.4, 6.6 for different values of p_f and the histograms obtained from the network simulator (QualNet($\mathbb{R}v.7.3$)). The simulated scenario had 2500 on average in an area of $7000 \times 7000m^2$ with transmission



Figure 6.17: pmf and histogram of path length in a random network, nodes following a Poisson Point Process; maximum Euclidean distance between the source and the farthest node.

range of 250*m*. All other simulation parameters correspond to those presented in Table 6.2. In each of the 120 replications a different realization of the Poisson process was used so as to eliminate any possible geometrical bias.

The results in Figures 6.17 - 6.19 illustrate the behavior of path length as a function of the Euclidean distance and p_f . As in the case of grids, the values of p_f were chosen around the critical value that yields acceptable reachability for nodes placed at random [1,96].

Figures 6.17 – 6.19 indicate that path length distributions concentrate more probability at the beginning of the range of h for nodes near the source. As the Euclidean distance increases, probability is distributed over a longer range. A similar effect is caused by reducing the value of p_f . Therefore, the effect of p_f reducing node density and leading to larger deviations from a straight-line path is confirmed.

For the results in Figure 6.17, the maximum Euclidean distance was taken from the average distance between the two more distant nodes in every replication (i.e. nodes in opposite corners of the network area). However, since Equation 6.4 is linear in l, an upperbound on the distribution of path length can be obtained by using $l = 7000\sqrt{2}$. The same is true for Figures 6.18 and 6.19, since 0.8 of the maximum distance was obtained between a broadcast source near the point (0, 3500) and the farthest reachable nodes in the opposite corners of the square, and 0.5 of the maximum distance represents the case with the broadcast source near the center of the square region.

To test the validity of the results, the P-values from the Chi-square goodness of fit (GoF) test are shown in Table 6.6. Similar results were obtained for the case of 2500 nodes Uniformly distributed in the same region (see Table 6.7). All P-values reported were obtained with k = 4 in Equation 6.6. Higher P-values were obtained when testing



Figure 6.18: pmf and histogram of path length, nodes following a Poisson Point Process; 0.8 maximum Euclidean distance.



Figure 6.19: pmf and histogram of path length, nodes following a Poisson Point Process, 0.5 maximum Euclidean distance.

pmf GoF Poisson-distributed nodes									
p_f	max d	ist.	0.8max dist.		0.5max dist.				
	P-value	lue Bins P-value Bins P-value B							
0.75	0.02	12	0.04	11	0.19	8			
0.65	0.17 13		0.60	12	0.61	8			
0.6	_	9	0.89	10	0.05	9			

Table 6.6: Chi-square goodness of fit (GoF), pmf of path length; Poisson Point Process; P-values reported with k = 4 in Equation 6.5.

pmf GoF Uniformly-distributed nodes									
p_f	max d	ist.	0.8max dist.		0.5max dist.				
	P-value	P-value Bins P-value Bins H							
0.75	0.07	12	_	12	0.34	8			
0.65	0.11	12	0.97	12	0.03	9			
0.6	0.01 10 0.49 9 0.45								

Table 6.7: Chi-square goodness of fit (GoF) test for the pmf of path length, nodes uniformly distributed; P-values reported with k = 4 in Equation 6.5.

the maximum Euclidean distance with k = 5; also, for 0.5 of the maximum distance, higher P-values were obtained with k = 3. This observation coincides with the analysis in [138], in which higher values of k are associated to neighbors that are farher away from the reference node.

For the case of $p_f = 0.6$ with maximum Euclidean distance, in Table 6.6, the GoF test failed with a P-value of 0.0046. However, as observed in Figure 6.17, the lack of fit was caused by a reduced number of obsevations in the tail of the distribution. This means that the tail predicted by the model is conservative for long Euclidean distances.

Regarding the lack of fit in Table 6.7 ($p_f = 0.75$ and 0.8 of max. Euclidean distance), a P-value of 0.13 was obtained by increasing h_{av} in one hop. This suggests good stability of the expected values obtained from equation 6.4 since only in one case (out of 18) was it necessary to modify the value calculated for h_{av} . It is worth mentioning that h_{min} is also a stable average. Indeed, there are only occasional observations below h_{min} , especially for 0.5max Euclidean distance, e.g. the bins before h = 29 in Figure 6.19 correspond to only one observation from the samples.

Overall, the Erlang-based pmf obtained from h_{min} and h_{av} , with k = 4, can be confidently used to estimate the distribution of the path length of broadcast packets in ad hoc networks with nodes placed according to a homogeneous Poisson Point Process or according to the Uniform distribution. More accurate results can be obtained with k = 3 when nodes are closer to the source, and k = 5 can be used for longer Euclidean distances.

6.5 pdf of Random Sums

Having the probability mass function (pmf) of the path length, broadcast latency will be computed from the sum of random delays. Specifically, for a sum of independent and identically distributed continuous random variables X_i , when the number of these variables is itself a random variable with discrete probability distribution f(n) = P[N = n], the pdf of the sum $Y = \sum_{i=1}^{N} X_i$ has Laplace transform [140]:

$$Y^*(s) = \sum_{n=0}^{\infty} P[N=n] [X^*(s)]^n,$$
(6.7)

where $X^*(s)$ is the Laplace transform of the common pdf of the set of continuous variables X_i . This model of *random sums* suits probabilistic broadcast schemes since the random delays of protocols are independent and identically distributed.

In this section the sequence P[N = n] corresponds to the pmf of path length, obtained from (6.2) for grids or from (6.6) for random networks, and $X^*(s)$ is the Laplace transform of the pdf of random delays.

6.5.1 Latency pdf for uniform random delays

A continuous uniform random variable defined in [0, T] has Laplace transform:

$$X^*(s) = \frac{1 - e^{-Ts}}{Ts}$$
(6.8)

Then, the Laplace transform of the random sum of uniform random delays corresponds to:

$$Y^{*}(s) = \sum_{n=0}^{\infty} \frac{P[N=n]}{T^{n}} \left(\frac{1-e^{-Ts}}{s}\right)^{n}$$
(6.9)

Let us denote $Y_n^*(s)$ as one of the terms of the sum in (6.9). Then, by using the binomial identity, the inverse transform of each $Y_n^*(s)$ follows from:

$$Y_{n}^{*}(s) = \frac{P[N=n]}{T^{n}} \left(\frac{1-e^{-Ts}}{s}\right)^{n}$$

= $\frac{P[N=n]}{T^{n}} \sum_{k=0}^{n} {n \choose k} \frac{(-e^{-Ts})^{k}}{s^{n}}$
 $y_{n}(t) = a_{n} \sum_{k=0}^{n} {n \choose k} \frac{(-1)^{k}}{(n-1)!} \left(\frac{t-kT}{T}\right)^{n-1} u_{kT},$ (6.10)

where $a_n = P[N = n]/T$, u_{kT} represents the unit step function shifted by kT, and $y_n(t)$ is the inverse Laplace transform of each term $Y_n^*(s)$. Threfore, the pmf of latency corresponds to:

$$\mathcal{L}^{-1}\{Y^*(s)\} = \sum_{n=0}^{\infty} y_n(t)$$
(6.11)

Before closing this section, we derive an equivalent expression of (6.10) for the case of truncated exponential random delays. This allows a more comprehensive evaluation of the models by using different distributions of random delays. The truncated exponential distribution was chosen since it involves no additional computational burden to be implemented in protocols (i.e. no additional random numbers using direct inverse method) and it always produces a smaller mean compared to a uniform distribution defined in the same range. Therefore, the model can be evaluated with a different distribution and a reduced mean of random delays at the same time.

Following the same procedure shown in (6.8) and (6.9), the inverse transform of each $Y_n^*(s)$ when using truncated exponential random delays follows from:

$$Y_{n}^{*}(s) = P[N = n] \left[\frac{\lambda(1 - e^{-(s+\lambda)T})}{(1 - e^{-\lambda T})(s+\lambda)} \right]^{n}$$

= $\frac{\lambda^{n} P[N = n]}{(1 - e^{-\lambda T})^{n}} \sum_{k=0}^{n} \binom{n}{k} \frac{(-e^{-(s+\lambda)T})^{k}}{(s+\lambda)^{n}}$
 $y_{n}(t) = b_{n} \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^{k} e^{-\lambda t}}{(n-1)!} \left(\frac{\lambda(t - kT)}{1 - e^{-\lambda T}} \right)^{n-1} u_{kT},$ (6.12)

where $b_n = \lambda P[N = n]/(1 - e^{-\lambda T})$.

6.6 Validation of Latency Models

To avoid biased validation, the simulation scenarios used in Section 6.4 were modified. In this section, the simulation area is a rectangle with height-width ratio of 2:5 and broadcast sources are placed on one of the shorter sides of the area (i.e. broadcast packets propagate along a narrow corridor). These scenarios correspond to 20×50 grids and a rectagle of $7500 \times 3000m^2$ for random scenarios with 1000 nodes on average. Also, the model was tested for different values of p_f , different distances from the source and different types of random delays. All other simulation parameters remain as in Table 6.2.

Since the pdf of latency can be used to estimate the service level that a broadcast scheme provides for dissemination services such as route discovery, the accuracy of the models is tested by comparing simulation results to the percentiles 95 and 99 computed using equation 6.11.

Results are shown for: i) two values of the forwarding probability p_f , one value around the minimum value of p_f reported for reachability above 90% [6,7], and a higher value to see the effects of redundancy. ii) four types of random delays, two instances of uniform random delay and two of truncated-exponentially distributed random delays [129]. One of the uniform instances uses T = 10ms since it is the most common value used in different studies [10,127], and the other uses T = 100ms to consider the studies in which delays are long enough to resemble a scenario with no packet collisions [39,93]. The exponentially distributed instances correspond to truncated-exponential distributions defined with T =20ms with mean 5ms and 7ms. This permits comparing the results with uniform random delays for the same mean (5ms), and observing the accuracy of the model with a variety of random delay distributions. iii) *different distances from the source*. For grids, the distances from the source are associated to the Tiers presented in Section 6.4.1, whereas Euclidean distances are associated with nodes placed at random. Different distances can be interpreted as networks of different sizes since broadcast messages are forwarded regardless of the direction of propagation of packets and so the error introduced by packets hopping back to the tier that represents the end of the network is small (recall the probability of having more than one hop back in Section 6.4.1.2).

6.6.1 Validation of Latency Model for Grids

Tables 6.8–6.13 show the average percentage of broadcast packets with latency higher than the percentiles 95 and 99 obtained from the pdf of latency for grids. The pdf was computed using 40 probability mass points in (6.2), i.e. 40 terms in (6.11).

The average values in Tables 6.8-6.13 were obtained considering all the nodes in the same tier and are presented with 95% confidence intervals.

6.6.1.1 Validation for the Square Grid

In Tables 6.8 and 6.9, the parameter p in (6.1) was adjusted in -0.05 to consider collisions increasing the length of the paths only in some instances of $p_f = 0.72$ (as explained in Section 6.4.1.2). The results obtained with $p_f = 0.8$ have no adjustment. This indicates that the redundant packets caused by using $p_f = 0.8$ compensate the effects of packet loss due to collisions.

		rn.dl.Ur	n.T = 10	$\mathrm{rn.dl.Un.}T = 100$		
p_f	Tier	p95	p99	p95	p99	
	20	$9.0{\pm}1.3$	$3.8{\pm}1.2$	5.6 ± 0.9	$3.1{\pm}0.5$	
0.72	35	$9.6{\pm}2.8$	$3.9{\pm}1.3$	$2.6{\pm}0.7$	0.3 ± 0.2	
	50	12 ± 2.5	$1.3{\pm}0.7$	$2.1{\pm}0.7$	0.1 ± 0.2	
	20	11 ± 5.6	4.1±2.3	$2.0{\pm}1.0$	0.8±0.6	
0.8	35	$20{\pm}7.2$	6.7 ± 2.3	$0.9{\pm}0.7$	0	
	50	19 ± 7.9	5.1 ± 2.4	0.5 ± 0.4	.04±.09	

Table 6.8: Square grid, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Uniform random delays.

In Table 6.8, the largest deviation from the prediction of the model occurred for uniform random delays with T = 10ms. In this case, larger errors always exhibit larger confidence intervals, and this pattern is more noticeable with $p_f = 0.8$. This indicates that more contention (higher values of p_f and reduced T) prompt the broadcast wave into nonuniform propagation patterns in which nodes in certain areas receive broadcast packets considerably later than the other nodes. Indeed, in this set of results, nodes located towards the borders followed the predictions of the model a lot more closer than nodes located towards the center of the rectangular area. For example, at Tier 20 considering 14 nodes (7 nodes closest to each side) the percentage of packets with latency higher than p95 is 4.3 ± 2.0 and for p99 the result is 1.3 ± 0.7 . This pattern was observed at all tiers, but it was less evident for $p_f = 0.72$. In Table 6.8, the case of T = 100 shows a slight over estimation of the model. In this case, no adjustments to the value of p in (6.1) were used at all, confirming that large values of T resemble an ideal broadcast event with no collisions. For studies that use large values of T to avoid collisions completely, the model offers a tight upperbound.

		dl.Ex $T =$	$=20 \ \mu = 5$	dl. Ex $T\!=\!20~\mu\!=\!7$	
p_f	Tier	p95	p99	p95	p99
	20	8.5 ± 1.7	$3.1{\pm}1.2$	6.9±1.0	$1.9{\pm}0.7$
0.72	35	8.3 ± 2.5	$2.7{\pm}0.7$	5.1 ± 1.9	$0.9{\pm}0.5$
	50	6.6 ± 2.2	$1.4{\pm}0.5$	$3.4{\pm}0.9$	$0.2{\pm}0.2$
	20	$5.1{\pm}1.9$	1.4±0.4	5.1±1.3	$1.9{\pm}0.8$
0.8	35	$1.8{\pm}0.9$	$0.2{\pm}0.2$	$3.0{\pm}1.4$	$0.6{\pm}0.4$
	50	0.8 ± 0.4	.04±.09	0.4±0.4	0

Table 6.9: Square grid, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Truncated Exponential random delays.

Table 6.9 shows that the use of truncated-exponential random delays reduces the error of the model. Specifically for $p_f = 0.8$, which has no adjustment to consider collisions, the predictions of the model improved significantly. Considering that uniform random delays with T = 10 have also a mean of $\mu = 5$ ms, the results in Table 6.9 suggest that exponentially distributed random delays avoid the effects of excessive contention better than uniform delays. The results in Table 6.9 show an excellent fit of the model, with slight overestimations only for Tier 50.

6.6.1.2 Validation for the Hexagonal Grid

Given the results in Section 6.4.1.3 (no influence of collisions on path length), the results in Tables 6.10 and 6.11 were obtained with no adjustment on the value of p in (6.1).

		rn.dl.Ur	n.T = 10	rn.dl.Un.T = 100	
p_f	Tier	p95	p99	p95	p99
	30	7.1 ± 2.4	$2.0{\pm}2.0$	1.3±1.2	0
0.8	60	1.7 ± 2.6	0	$0.2{\pm}0.4$	0
	90	0	0	0	0
	30	$9.0{\pm}2.0$	$2.8{\pm}1.1$	0.8 ± 0.5	$0.2{\pm}0.3$
0.9	60	7.7 ± 1.4	2.3 ± 0.8	0	0
	90	11 ± 2.3	$3.5{\pm}1.1$	0	0

Table 6.10: Hexagonal grid (Honeycomb), percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Uniform random delays.

As in the case of the square grid, the maximum error occured when using uniform random delays with T = 10 and $p_f = 0.9$. This confirms that the combined effect of reducing T and increasing p_f causes non-homogeneous propagation of packets, increasing the error and the width of confidence intervals (as explained for square grids).

		dl.Ex05 μ T = 20		dl.Ex07 μ T = 20	
p_f	Tier	p95	p99	p95	p99
	30	5.6 ± 2.7	3.1±1.9	6.7 ± 3.6	1.0±1.2
0.8	60	$2.6{\pm}2.0$	$1.6{\pm}1.8$	8.1 ± 5.8	3.8 ± 3.9
	90	0	0	0	0
	30	6.6 ± 2.0	$1.9{\pm}1.3$	$3.6{\pm}1.5$	$0.7{\pm}0.4$
0.9	60	$2.4{\pm}1.1$	$0.1{\pm}0.2$	0.3 ± 0.3	0
	90	0	0	0	0

Table 6.11: Hexagonal grid (Honeycomb), percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Truncated Exponential random delays.

For the Hexagonal grid, the model overestimates latency for distant nodes more quickly than in the case of the square grid. In ideal conditions (i.e. T = 100), Table 6.10 shows that the model is an upper-bound even for nodes in Tier30. Moreover, despite the reduced error of the model with exponentially distributed delays (Table 6.11), no packets were registered arriving at Tier90 after p95.

6.6.1.3 Validation for the Triangular Grid

Unlike the square and hexagonal grids, the triangular grid shows stable results for reduced T and increased p_f (Table 6.12, uniform delay with T = 10 and $p_f = 0.75$). This indicates that, despite the increased contention, a higher number of neighbors produces more homogeneous broadcast propagation, leading to more precise predictions of the model for uniform delays and T = 10.

		rn.dl.Ur	n.T = 10	rn.dl.Un.T = 100	
p_f	Tier	p95	p99	p95	p99
	20	7.2 ± 1.4	2.8 ± 0.9	4.6 ± 1.3	2.2±1.1
0.65	35	$8.0{\pm}2.1$	$3.1{\pm}1.3$	$0.8{\pm}0.5$	0
	50	$3.8{\pm}1.4$	0.5 ± 0.5	0	0
	20	8.9±2.3	$3.4{\pm}1.0$	0.3 ± 0.3	.04±.09
0.75	35	7.8 ± 2.2	$1.9{\pm}1.0$	0	0
	50	2.5 ± 1.4	$0.4{\pm}0.2$	0	0

Table 6.12: Triangular grid, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Uniform random delays.

For T = 100 in Table 6.12, the model shows the same trends of the predictions for the hexagonal grid. This indicates that the model can be improved to make more accurate predictions for distant nodes.

Regarding the use of exponentially distributed random delays, the results in Table 6.13 are closer to the predictions of the model. However, the trend of overestimation for more distant nodes remains.

		$dl.Ex05\mu$	$\iota T = 20$	dl.Ex07 μ T = 20	
p_f	Tier	p95	p99	p95	p99
	20	$3.9{\pm}1.1$	$1.9{\pm}0.7$	2.7±0.8	1.1 ± 0.4
0.65	35	$2.8{\pm}0.7$	$0.4{\pm}0.3$	2.1 ± 0.7	$0.9{\pm}0.3$
	50	$0.4{\pm}0.3$	0	0.3±0.2	0
	20	1.8 ± 0.7	0.8 ± 0.6	1.1±0.8	$0.2{\pm}0.2$
0.75	35	$0.6{\pm}0.5$	$0.2{\pm}0.2$	0.4 ± 0.3	0
	50	$.05{\pm}0.1$	0	0	0

Table 6.13: Triangular grid, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Truncated Exponential random delays.

6.6.2 Trends of the Model for Grids

In all grids, two trends are identified: i) Higher values of p_f produce shorter latency; ii) the model predicts longer latency for more distant nodes.

These two trends can be explained from the assumptions about path formation in Sections 6.4.1.2 6.4.1.3 and 6.4.1.4. In the assumption, paths that go back and forth, propagating in parallel, reach nodes in specific tiers with the same probability. However, when simulating a real network, parallel paths influence each other. As distance increases, paths with less back-hopping will reach the farthest nodes first.

6.6.3 Validation of Latency Model for Random Networks

Since broadcast latency in random networks was modelled according to Euclidean distances, instead of tiers, the average number of packets exceeding the latency from the model (p95 and p99) is computed from the traces of 20 nodes with euclidean distance closest to the target distance. In all cases the deviation from the target distance was smaller than 150m (60% of the transmission range of nodes).

Interestingly, for random networks following a Poisson Point Process, the model shows accurate results without any adjustment to consider the effects of collisions. In particular, for uniform random delays using T = 10 (Table 6.14), results show stability regardless of the value of p_f . In the case of T = 100, the model certainly becomes an upper-bound of latency.

		rn.dl.Ur	n.T = 10	rn.dl.Un.T = 100	
p_f	Km	p95	p99	p95	p99
	2.3	7.4±1.3	4.0±0.8	2.1 ± 0.6	$1.2{\pm}0.4$
0.7	4.6	5.0 ± 0.6	$1.6{\pm}0.4$	0.3 ± 0.2	0
	7	0.5 ± 0.1	$0.1{\pm}0.2$	0	0
	2.3	7.4±0.7	$3.6{\pm}0.6$	0.2 ± 0.2	$0.1{\pm}0.1$
0.8	4.6	1.7 ± 0.6	$0.9{\pm}0.3$	0	0
	7	.04±.09	0	0	0

Table 6.14: Poisson Point Process, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Uniform random delays.

When using exponentially distributed delays, the model precisely predicts latency at 2.3Km from the source. However, at 7Km from the source no packets exceeded the values of latency at p95.

		dl.Ex05µ	$\mu T = 10$	dl.Ex07 μ T = 20	
p_f	Km	p95	p99	p95	p99
	2.3	5.4±1.0	3.3 ± 0.7	4.5 ± 0.9	2.2 ± 0.7
0.7	4.6	1.5 ± 0.5	$0.3{\pm}0.2$	1.3 ± 0.6	0.3 ± 0.2
	7	0	0	0	0
	2.3	3.1 ± 0.6	1.5 ± 0.3	1.8±0.4	0.7 ± 0.3
0.8	4.6	0.8 ± 0.3	$.04 \pm .09$	0	0
	7	0	0	0	0

Table 6.15: Poisson Point Process, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Truncated Exponential random delays.

Equivalent results for nodes placed uniformly at random are shown in Tables 6.16 and 6.17. These tables show a slight, but consistent, increase in latency when the assumptions of the Poisson Point Process are relaxed by having a fixed number of nodes. However, this new scenario causes the model to be a tighter upper-bound for the most distant nodes.

		rn.dl.Ur	n.T = 10	rn.dl.Un.T = 100		
p_f	Km	p95	p95 p99		p99	
	2.3	9.6±1.0	$7.2{\pm}1.0$	3.5 ± 0.8	1.8 ± 0.4	
0.7	4.6	6.3 ± 1.2	$3.5{\pm}0.8$	0.9 ± 0.4	$0.1{\pm}0.2$	
	7	.05±0.1	0	0	0	
	2.3	6.9±1.0	$4.1{\pm}0.7$	0.6 ± 0.2	$0.4{\pm}0.2$	
0.8	4.6	$3.1{\pm}0.7$	$1.3{\pm}0.5$	0	0	
	7	0.3 ± 0.2	0	0	0	

Table 6.16: Random Uniform layout, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Uniform random delays.

The largest error of the model in random scenarios occurs for uniform random delays with T = 10 and 2.3Km, regardless of the value of p_f (Tables 6.16 and 6.14). Two factors may introduce such error. First, the model is not adusted to consider collisions, and the reduced value of T increases the probability of collision events. Second, in [139], the relation between graph distance (path-length) and Euclidean distance is maintained for sufficiently distant nodes, indicating that for closer nodes, the same relation may not hold, producing longer paths.

Again, the use of exponentially distributed random delays reduced the error, especially for nodes closer to the broadcast source. Overall, the results with nodes uniformly distributed show the same trends as the results obtained with nodes following a Poisson Point Process. Then the model can be used for comparison purposes with studies that consider a fixed number of nodes uniformly distributed in the network area.

		dl.Ex05µ	$\iota T = 20$	dl.Ex07 μ T = 20		
p_f	Km	p95	p99	p95	p99	
	2.3	7.6±1.4	$3.4{\pm}0.7$	5.7 ± 0.7	3.5 ± 0.7	
0.7	4.6	2.2 ± 0.4	$1.3{\pm}0.5$	$1.6{\pm}0.5$	$0.7{\pm}0.4$	
	7	0.5 ± 0.3	0	0	0	
	2.3	$3.2{\pm}0.8$	$1.9{\pm}0.5$	$1.7{\pm}0.5$	0.7 ± 0.3	
0.8	4.6	0.3 ± 0.3	$.05{\pm}0.1$	$0.2{\pm}0.2$	0	
	7	0	0	0	0	

Table 6.17: Random Uniform layout, percentage of packets with delay above percentiles 95 and 99, broadcast protocol using Truncated Exponential random delays.

6.7 Chapter Summary and Future Work

A way to compute the probability density function (pdf) of broadcast latency in ad hoc networks using probabilistic broadcast protocols was proposed. Broadcast latency was modelled in two stages. First, two models for the probability mass function (pmf) of the path length traversed by broadcast packets were presented; one model for grids and the other for random networks. Second, the distribution of random sums was used to combine the pmf of the path length with the parameters of random delays employed in probabilistic broadcast protocols.

All models were validated using simulation. The resulting pdf of latency was compared to simulation results using different values of forwarding probability (p_f) , different distances from the broadcast source and different types of random delays.

Results show that the model lacks precision for uniform random delays with a mean of 5ms. However, for the same mean, precision improves when introducing exponentially distributed random delays. Precision increases as the mean of random delay increases. The model becomes an upper-bound of broadcast latency for distant nodes.

The model allows quantifying the reductions in latency caused by reducing the interval in which random delays are defined or by increasing the value of p_f . However, when the interval of uniform random delays is equal or smaller than T = 10, there is a counter-productive effect (excessive contention leads packets to take longer paths, increasing latency).

The model and the validation reveal that allowing some redundancy in probabilistic broadcast schemes (p_f larger than the minimum value required for a target reachability) produces faster broadcast, especially as network size increases (i.e. for the farthest nodes in a large network).

Future work is aimed at improving the precision and the simplicity of the model for distant nodes. An approximation of the formulas for random sums should allow a closed expression for the pdf of broadcast latency, facilitating adjustments to consider a decrease in latency estimation for more distant nodes. Moreover, theoretical background on other random distributions different from the Poisson Point Porcess can be explored to estimate latency in other types of random scenarios.

6.8 Table with Related Work

Table 6.18 compares the type of variables for which analytical models have been proposed in the literature. The variables of interest when assessing probabilistic broadcastig schemes as a packet dissemination technique include the forwarding probability (p_f) , reachability (Re), the amount of Saved Rebroadcast (SBR) and the end-to-end latency (Late.) of packets propagating from the source to the last reachable node.

The columns that refer to the expression provided by authors use the convention E[X] to refer to the expected value (or average value) of the variable of interest, whereas the convention f(X) suggests that authors computed (or simulated) the probability distribution function of the variables.

	N	letric of	Expression			
	p_f	Re	SBR	Late.	E[X]	f(X)
Williams [96]	\checkmark					
Zhang [97, 98]		\checkmark	\checkmark	\checkmark	\checkmark	
Viswanath [99]		\checkmark			\checkmark	
Lysiuk [1], Haas [6]		\checkmark				\checkmark
Drabkin [12]		\checkmark			\checkmark	
This Chapter				\checkmark		\checkmark

Table 6.18: A comparison of studies that provide analytical models to compute probabilistic broadcasting variables of interest.

Chapter
SEVEN

Conclusions and Recommendations

The problem of disseminating packets in ad hoc networks has been addressed in this thesis from a probabilistic perspective. A comprehensive survey of the literature has been provided with special emphasis on the fundamental ideas that support most proposals. The two main categories of probabilistic broadcasting schemes are based on the estimation of node density by using information from neighbor nodes (neighbor-based schemes) or by computing the additional area that should be covered when forwarding a packet (areabased schemes). Schemes that fall in these two categories have a distinguishing feature: the use of periodic Hello packets. Common tradeoffs to all schemes were identified at the end of the survey.

This thesis was oriented towards studying factors that can impact the reachability of probabilistic broadcasting schemes in general. As a result, the studies herein were aimed at analyzing the impact of variables that affect the reachability of probabilistic broadcasting regardless of the scheme (e.g. the impact of node layout and the calibration of random delays).

The first study (Chapter 4) presented an analysis that made use of the Signal to Interference ratio of nodes when receiving broadcast packets. Such an analysis evinced the impact of the geometry of node distribution on the successful propagation of broadcast packets away from the source. By exploring a variety of geometrical node layouts the study revealed that networks with randomly-placed nodes exhibit a *near-optimum* (near collision-free) behavior, while grid layouts are extremely sensitive to the impact of node interference. The study was successful in incorporating realistic assumptions (i.e. S/I model) in the analysis of the causes of the limitations in the reachability of packets in probabilistic broadcasting schemes.

The second study (Chapter 5) focused on the random delays that have been widely used in ad hoc broadcasting schemes. Specifically, a change in the distribution of random delays was proposed to reduce the end-to-end latency of probabilistic schemes. Changing the distribution of random delays, from uniform to truncated-exponential, reduced latency without compromising reachability. Reducing latency is essential to delay-sensitive applications running over ad hoc networks. Fast broadcasting of information is also important to compensate the uncertainty resulting from topological changes in the analysis of Vehicular Ad Hoc Networks (VANETs). The most important aspect of this study is that the change in the distribution of random delays can be applied to any existing scheme. The last study (Chapter 6) employed the topologies studied in Chapter 4 and the parameters of random delays to provide an analytical model for the latency in probabilistic broadcasting schemes. Analytical expressions to compute the probability density function (pdf) of broadcast latency were provided and validated using simulation. The resulting models show the way broadcast latency distributions change as a function of the forwarding probability (p_f) , the network geometry and the different types of random delays.

Future work about the impact of topology and node layouts in the performance of probabilistic broadcasting schemes should focus on evaluating probabilistic broadcasting schemes with static node distributions different from the random uniform distribution. Likewise, for mobile networks, evaluation studies should be conducted using mobility models different from the random way point model. All these future studies should consider in their analysis the effects of the wireless channel so that the results reported are related to realistic conditions.

Further work is necessary on the proper use of random delays. Finding the optimal length of the interval in which random delays are defined to increase reachability with minimum latency is still an open issue, and the suggested analysis should consider the way protocols ocupy the wireless channel (adding Hello packets or longer headers to broadcast packets). Moreover, additional studies should be conducted to compare the impact of truncated exponential random delays on the reachability of different schemes, especially counter-based schemes, which rely entirely on random delays.

As pointed out in [10], analytical models are scarce in the literature. Given that probabilistic broadcasting schemes have been proposed for several protocols, theoretical analysis in the field will continue to be an area for future development.

Recommendations:

The present work has shown that theoretically expected results (e.g. predicted by percolation models) can be achieved by observing the conditions of an ad hoc network that prompts a near-ideal behavior of probabilistic broadcasting schemes (e.g. geometry and random delays). The community is encouraged to continue looking for network variables and conditions that affect the performance of protocols, providing explanatory analysis.

Broadcast should be lightweight, avoiding the dependence on previously collected state information. Neighbor-based schemes require the use of a Hello protocol, using additional energy and bandwidth, which has been a problem in test-bed studies. It is recommended that the proposals in this work be used towards the success of copy-counting schemes.

References

- I. S. Lysiuk and Z. J. Haas. Controlled gossiping in ad hoc networks. In 2010 IEEE Wireless Communication and Networking Conference, pages 1–6, April 2010.
- [2] M. Penrose. Random Geometric Graphs. Oxford studies in probability. Oxford University Press, 2003.
- [3] P. Santi. Topology control in wireless ad hoc and sensor networks. Wiley, 2005.
- [4] G. Grimmett. Percolation. Springer-Verlag, 1989.
- [5] D. Stauffer. Introduction to Percolation. Taylor & Francis, 1994.
- [6] Zygmunt J. Haas, Joseph Y. Halpern, and Li Li. Gossip-based ad hoc routing. IEEE/ACM Trans. Netw., 14(3):479–491, jun 2006.
- [7] Chien-Chung Shen, Zhuochuan Huang, and Chaiporn Jaikaeo. Directional broadcast for mobile ad hoc networks with percolation theory. *IEEE Transactions on Mobile Computing*, 5(4):317–332, April 2006.
- [8] Sze-Yao Ni, Yu-Chee Tseng, Yuh-Shyan Chen, and Jang-Ping Sheu. The broadcast storm problem in a mobile ad hoc network. In *Proceedings of the 5th Annual* ACM/IEEE International Conference on Mobile Computing and Networking, Mobi-Com '99, pages 151–162, New York, NY, USA, 1999. ACM.
- [9] LAN MAN Standards Committee of the IEEE Computer Society. Ieee standard 802.11-2012. Technical Report Part 11, LAN/MAN Standards Committee of the IEEE Computer Society, IEEE 3 Park Avenue New York, NY 10016-5997 USA, March 2012.
- [10] D.G. Reina, S.L. Toral, P. Johnson, and F. Barrero. A survey on probabilistic broadcast schemes for wireless ad hoc networks. Ad Hoc Networks, 25, Part A:263 – 292, 2015.
- [11] V. Drabkin, R. Friedman, G. Kliot, and M. Segal. Rapid: Reliable probabilistic dissemination in wireless ad-hoc networks. In *Reliable Distributed Systems*, 2007. SRDS 2007. 26th IEEE International Symposium on, pages 13–22, Oct 2007.

- [12] V. Drabkin, R. Friedman, G. Kliot, and M. Segal. On reliable dissemination in wireless ad hoc networks. *IEEE Transactions on Dependable and Secure Computing*, 8(6):866–882, Nov 2011.
- [13] J. Cartigny and D. Simplot. Border node retransmission based probabilistic broadcast protocols in ad-hoc networks. In System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on, pages 10 pp.-, Jan 2003.
- [14] Abdalla M. Hanashi, Aamir Siddique, Irfan Awan, and Mike Woodward. Performance evaluation of dynamic probabilistic broadcasting for flooding in mobile ad hoc networks. *Simulation Modelling Practice and Theory*, 17(2):364 – 375, 2009.
- [15] Abdelmajid Khelil, Pedro Jose Marron, Christian Becker, and Kurt Rothermel. Hypergossiping: A generalized broadcast strategy for mobile ad hoc networks. Ad Hoc Networks, 5(5):531 – 546, 2007.
- [16] Z. J. Haas, J. Y. Halpern, and Li Li. Gossip-based ad hoc routing. In INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, volume 3, pages 1707–1716 vol.3, jul 2002.
- [17] M. Bani Yassein, M. Ould Khaoua, L. M. Mackenzie, S. Papanastasiou, and A. Jamal. Improving route discovery in on-demand routing protocols using local topology information in manets. In *Proceedings of the ACM International Workshop on Performance Monitoring, Measurement, and Evaluation of Heterogeneous Wireless and Wired Networks*, PM2HW2N '06, pages 95–99, New York, NY, USA, 2006. ACM.
- [18] Jamal-Deen Abdulai, Mohamed Ould-Khaoua, and Lewis M. Mackenzie. Adjusted probabilistic route discovery in mobile ad hoc networks. *Computers and Electrical Engineering*, 35(1):168 – 182, 2009.
- [19] J. d. Abdulai, A. Mohammed, K. S. Nokoe, and E. Oyetunji. Route discovery in wireless mobile ad hoc networks with adjusted probabilistic flooding. In 2009 2nd International Conference on Adaptive Science Technology (ICAST), pages 99–109, Jan 2009.
- [20] Qi Zhang and D. P. Agrawal. Analysis of leveled probabilistic routing in mobile ad hoc networks. In *Communications*, 2004 IEEE International Conference on, volume 7, pages 3896–3900, 2004.
- [21] X. M. Zhang, E. B. Wang, J. J. Xia, and D. K. Sung. A neighbor coverage-based probabilistic rebroadcast for reducing routing overhead in mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 12(3):424–433, March 2013.
- [22] Jung-Pil Ryu, Min-Su Kim, Sung-Ho Hwang, and Ki-Jun Han. An Adaptive Probabilistic Broadcast Scheme for Ad-Hoc Networks, pages 646–654. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [23] D. G. Reina, S. L. Toral, P. Jonhson, and F. Barrero. Hybrid flooding scheme for mobile ad hoc networks. *IEEE Communications Letters*, 17(3):592–595, March 2013.

- [24] H. Alshaer and E. Horlait. An optimized adaptive broadcast scheme for inter-vehicle communication. In 2005 IEEE 61st Vehicular Technology Conference, volume 5, pages 2840–2844, May 2005.
- [25] J. D. Abdulai, M. Ould-Khaoua, L. M. Mackenzie, and A. Mohammed. Neighbour coverage: A dynamic probabilistic route discovery for mobile ad hoc networks. In *Performance Evaluation of Computer and Telecommunication Systems*, 2008. SPECTS 2008. International Symposium on, pages 165–172, June 2008.
- [26] H. Ling, D. Mosse, and T. Znati. Coverage-based probabilistic forwarding in ad hoc routing. In Proceedings. 14th International Conference on Computer Communications and Networks, 2005. ICCCN 2005., pages 13–18, Oct 2005.
- [27] Hyocheol Jeong, Jeonghyun Kim, and Younghwan Yoo. Adaptive broadcasting method using neighbor type information in wireless sensor networks. *Sensors*, 11(6):5952, 2011.
- [28] P. Kyasanur, R. R. Choudhury, and I. Gupta. Smart gossip: An adaptive gossipbased broadcasting service for sensor networks. In 2006 IEEE International Conference on Mobile Ad Hoc and Sensor Systems, pages 91–100, Oct 2006.
- [29] Boto Bako, Igor Rikanovic, Frank Kargl, and Elmar Schoch. Adaptive Topology Based Gossiping in VANETs Using Position Information, pages 66–78. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- [30] B. Bako, E. Schoch, F. Kargl, and M. Weber. Optimized position based gossiping in vanets. In Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th, pages 1–5, Sept 2008.
- [31] Donald J. Scott and Alec Yasinsac. Dynamic probabilistic retransmission in ad hoc networks. In *International Conference on Wireless Networks*, *ICWN '04*, pages 158–164, 2004. http://www.odysci.com/article/1010112986293953.
- [32] D. G. Reina, S. L. Toral, P. Johnson, and F. Barrero. Improving discovery phase of reactive ad hoc routing protocols using jaccard distance. *The Journal of Supercomputing*, 67(1):131–152, 2014.
- [33] I. A. Khan, A. Javaid, and Hua Lin Qian. Distance-based dynamically adjusted probabilistic forwarding for wireless mobile ad hoc networks. In 2008 5th IFIP International Conference on Wireless and Optical Communications Networks (WOCN '08), pages 1–6, May 2008.
- [34] Qian Qi Qing-wen Wang, Hao-shan Shi. A dynamic probabilistic broadcasting scheme based on cross-layer design for manets. *International Journ. of Modern Education and Computer Science (IJMECS)*, 2(1):40 – 47, 2010.
- [35] N. Wisitpongphan, O. K. Tonguz, J. S. Parikh, P. Mudalige, F. Bai, and V. Sadekar. Broadcast storm mitigation techniques in vehicular ad hoc networks. *IEEE Wireless Communications*, 14(6):84–94, December 2007.

- [36] Xiang-Yang Li, Kousha Moaveninejad, and Ophir Frieder. Regional gossip routing for wireless ad hoc networks. *Mobile Networks and Applications*, 10(1):61–77, 2005.
- [37] R. Beraldi. A directional gossip protocol for path discovery in manets. In 26th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW'06), pages 82–82, July 2006.
- [38] Roberto Beraldi. The polarized gossip protocol for path discovery in {MANETs}. Ad Hoc Networks, 6(1):79 – 91, 2008.
- [39] R. Beraldi, L. Querzoni, and R. Baldoni. A hint-based probabilistic protocol for unicast communications in {MANETs}. Ad Hoc Networks, 4(5):547 – 566, 2006.
- [40] Aminu Mohammed, Mohamed Ould-Khaoua, Lewis M. Mackenzie, Colin Perkins, and Jamal-Deen Abdulai. Probabilistic counter-based route discovery for mobile ad hoc networks. In Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly, IWCMC '09, pages 1335–1339, New York, NY, USA, 2009. ACM.
- [41] Mohammad Naserian and Kemal Tepe. Game theoretic approach in routing protocol for wireless ad hoc networks. Ad Hoc Networks, 7(3):569 578, 2009.
- [42] Wahabou Abdou, Adrien Henriet, Christelle Bloch, Dominique Dhoutaut, Damien Charlet, and Francois Spies. Using an evolutionary algorithm to optimize the broadcasting methods in mobile ad hoc networks. Journal of Network and Computer Applications, 34(6):1794 – 1804, 2011. Control and Optimization over Wireless Networks.
- [43] D. G. Reina, J. M. León-Coca, S. L. Toral, E. Asimakopoulou, F. Barrero, P. Norrington, and N. Bessis. Multi-objective performance optimization of a probabilistic similarity/dissimilarity-based broadcasting scheme for mobile ad hoc networks in disaster response scenarios. *Soft Computing*, 18(9):1745–1756, 2014.
- [44] Max do Val Machado, Raquel A.F. Mini, Antonio A. F. Loureiro, Daniel L. Guidoni, and Pedro O. S. V. de Melo. Gossiping using the energy map in wireless sensor networks. In *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, MSWiM '07, pages 368–372, New York, NY, USA, 2007. ACM.
- [45] B. Blywis, P. Reinecke, M. Günes, and K. Wolter. Gossip routing, percolation, and restart in wireless multi-hop networks. In 2012 IEEE Wireless Communications and Networking Conference (WCNC), pages 3019–3023, April 2012.
- [46] A. Senthil Kumar and K. S. Ramaswami. A new density based probabilistic counter based scheme for solving broadcast storm in mobile ad hoc networks. *Journal of Theoretical and Applied Information Technology*, 59(2):451–459, 01 2014.
- [47] D. G. Reina, S. L. Toral, E. Asimakopoulou, N. Bessis, and F. Barrero. A dissemination analysis in mobile wireless ad hoc networks using probabilistic broadcast. In 2014 Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pages 110–114, July 2014.

- [48] D.G. Reina, S.L. Toral, E. Asimakopoulou, F. Barrero, and N. Bessis. The role of congestion in probabilistic broadcasting for ubiquitous wireless multi-hop networks through mediation analysis. *Pervasive and Mobile Computing*, 24:16 – 29, 2015. Special Issue on Secure Ubiquitous Computing.
- [49] H. Y. Adarbah, S. Ahmad, B. Arafeh, and A. Duffy. Efficient broadcasting for route discovery in mobile ad-hoc networks. In 2015 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), pages 1–7, July 2015.
- [50] H. Y. Adarbah and S. Ahmad. Efficient route discovery using channel adaptive probabilistic broadcasting in zigbee wireless sensor networks. In 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), pages 1–5, Sep. 2017.
- [51] Haitham Y. Adarbah, Shakeel Ahmad, and Alistair Duffy. Impact of noise and interference on probabilistic broadcast schemes in mobile ad-hoc networks. *Computer Networks*, 88:178 – 186, 2015.
- [52] A. Tahmasbi-Sarvestani, Y. P. Fallah, and V. Kulathumani. Network-aware doublelayer distance-dependent broadcast protocol for vanets. *IEEE Transactions on Vehicular Technology*, 64(12):5536–5546, Dec 2015.
- [53] Felipe Forero R., Néstor M. Peña T., and Nelson L.S. da Fonseca. Geometric aspects of probabilistic broadcasting in ad hoc networks. Ad Hoc Networks, 87:146 – 156, 2019.
- [54] X. Zhang, K. Chen, Y. Zhang, and D. K. Sung. A probabilistic broadcast algorithm based on the connectivity information of predictable rendezvous nodes in mobile ad hoc networks. In 2014 23rd International Conference on Computer Communication and Networks (ICCCN), pages 1–6, Aug 2014.
- [55] G. Kaur and R. Garg. Ncpr-pso neighborhood coverage probabilistic rebroadcast with particle swarm optimization. In 2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence), pages 298–303, Sep. 2014.
- [56] S. A. Jain and V. T. Raisinghani. Load equilibrium neighbor aware routing in mobile ad hoc network. In 2014 Annual IEEE India Conference (INDICON), pages 1–6, Dec 2014.
- [57] P. Shete and R. N. Awale. Channel quality based adaptive gossip flooding mechanism for aodv. In 2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), pages 553–559, May 2015.
- [58] Francesco Palmieri. A wave propagation-based adaptive probabilistic broadcast containment strategy for reactive manet routing protocols. *Pervasive and Mobile Computing*, 40:628 – 638, 2017.

- [59] M. M. Agarwal, M. C. Govil, and A. K. Jhankal. A probabilistic method to optimize energy consumption in mobile ad-hoc networks. In *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, pages 1–5, May 2014.
- [60] R. Ragul Ravi and V. Jayanthi. Energy efficient neighbor coverage protocol for reducing rebroadcast in manet. *Procedia Computer Science*, 47:417–423, 2015. Graph Algorithms, High Performance Implementations and Its Applications (ICGHIA 2014).
- [61] K. Alsmearat, M. Al-Ayyoub, and M. B. Yasseinz. A new broadcast scheme for sensor networks. In 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), pages 824–828, Nov 2014.
- [62] S. Mehfuz and S. Kumar. Energy aware probabilistic broadcasting for mobile adhoc network. In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pages 1028–1033, March 2015.
- [63] Sumit Kumar and Shabana Mehfuz. Energy efficient probabilistic broadcasting for mobile ad-hoc network. Journal of The Institution of Engineers (India): Series B, 98, 08 2016.
- [64] Sumit Kumar and Shabana Mehfuz. Efficient fuzzy logic based probabilistic broadcasting for mobile ad hoc network. *International Journal of Computational Intelli*gence Systems, 9:666–675, 07 2016.
- [65] A Kayalvizhi and R Karthik. Energy saving dsr and probabilistic rebroadcast mechanism are used to increase routing performance in manet. International Journal of Applied Engineering Research, 10:17761–17766, 01 2015.
- [66] H. Liu D. Luo L. Zhou, G. Cui and Z. Wu. Nppb: A broadcast scheme in dense vanets. *Information Technology Journal*, 9(2):247 – 256, 2010.
- [67] S. Panichpapiboon. Irresponsible forwarding under general inter-vehicle spacing distributions. In *Electrical Engineering/Electronics, Computer, Telecommunications* and Information Technology (ECTI-CON), 2011 8th International Conference on, pages 357–360, May 2011.
- [68] S. Panichpapiboon and L. Cheng. Irresponsible forwarding under real intervehicle spacing distributions. *IEEE Transactions on Vehicular Technology*, 62(5):2264–2272, Jun 2013.
- [69] S. Panichpapiboon and G. Ferrari. Irresponsible forwarding. In ITS Telecommunications, 2008. ITST 2008. 8th International Conference on, pages 311–316, Oct 2008.
- [70] M. Chaqfeh and A. Lakas. Speed adaptive probabilistic broadcast for scalable data dissemination in vehicular ad hoc networks. In 2014 International Wireless Communications and Mobile Computing Conference (IWCMC), pages 207–212, Aug 2014.

- [71] Yiannos Mylonas, Marios Lestas, and Andreas Pitsillides. Speed adaptive probabilistic flooding in cooperative emergency warning. In *Proceedings of the 4th Annual International Conference on Wireless Internet*, WICON '08, pages 81:1–81:7, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [72] Y. Mylonas, M. Lestas, A. Pitsillides, P. Ioannou, and V. Papadopoulou. Speed adaptive probabilistic flooding for vehicular ad hoc networks. *IEEE Transactions* on Vehicular Technology, 64(5):1973–1990, May 2015.
- [73] A. Y. Al-Dubai, M. B. Khalaf, W. Gharibi, and J. Ouenniche. A new adaptive probabilistic broadcast protocol for vehicular networks. In 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), pages 1–5, May 2015.
- [74] J. M. Garcia-Campos, J. Sanchez-Garcia, D. G. Reina, S. L. Toral, and F. Barrero. Evaluation of dissimilarity-based probabilistic broadcasting algorithms in vanets. In 2015 International Conference on Developments of E-Systems Engineering (DeSE), pages 29–34, Dec 2015.
- [75] Daniel Gutiérrez-Reina, Vishal Sharma, Ilsun You, and Sergio Toral. Dissimilarity metric based on local neighboring information and genetic programming for data dissemination in vehicular ad hoc networks (vanets). Sensors, 18(7):2320, 2018.
- [76] Andrea Gorrieri and Gianluigi Ferrari. Irresponsible adv routing. Vehicular Communications, 2(1):47 – 57, 2015.
- [77] Lei Liu, Chen Chen, Tie Qiu, Mengyuan Zhang, Siyu Li, and Bin Zhou. A data dissemination scheme based on clustering and probabilistic broadcasting in vanets. *Vehicular Communications*, 13:78 – 88, 2018.
- [78] X. Zeng, M. Yu, and D. Wang. A new probabilistic multi-hop broadcast protocol for vehicular networks. *IEEE Transactions on Vehicular Technology*, 67(12):12165– 12176, Dec 2018.
- [79] Xi Hu and Tao Wu. An adaptive game model for broadcasting in vanets. In Ying Tan, Yuhui Shi, and Qirong Tang, editors, *Advances in Swarm Intelligence*, pages 58–67, Cham, 2018. Springer International Publishing.
- [80] F. T. Zohra and A. Rahman. Mathematical analysis of self-pruning and a new dynamic probabilistic broadcast for manets. In 2015 International Conference on Networking Systems and Security (NSysS), pages 1–9, Jan 2015.
- [81] Raqeebir Rab, Ashikur Rahman, and Fatema Tuz Zohra. Analytical modeling of selfpruning and an improved probabilistic broadcast for wireless multihop networks. *Ad Hoc Networks*, 52:106 – 116, 2016. Modeling and Performance Evaluation of Wireless Ad Hoc Networks.
- [82] B. R. Vinay Kumar, Roshan Antony, and Navin Kashyap. The effect of introducing redundancy in a probabilistic forwarding protocol. CoRR, abs/1901.02033, 2019.

- [83] M. Chekhar, K. Zine-Dine, M. Bakhouya, and A. Aaroud. A dynamic thresholdbased probabilistic scheme for broadcasting in ad hoc networks. In 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA), pages 511–516, Dec 2015.
- [84] A. Gaba, S. Voulgaris, K. Iwanicki, and M. van Steen. Revisiting gossip-based adhoc routing. In 2012 21st International Conference on Computer Communications and Networks (ICCCN), pages 1–6, July 2012.
- [85] Kannan Shanmugam, Karthik Subburathinam, and Arunachalam Velayuthampalayam Palanisamy. A dynamic probabilistic based broadcasting scheme for manets. *TheScientificWorldJournal*, 2016:1832026, 02 2016.
- [86] Ali Ejmaa, Shamala Subramaniam, Zuriati Zukarnain, and Mohd Hanapi Zurina. A scalable neighbor-based routing protocol for mobile ad hoc networks. *International Journal of Distributed Sensor Networks*, 12, 09 2016.
- [87] D.G. Reina, M. Günes, and S.L. Toral. Real experimentation of probabilistic broadcasting algorithms based on dissimilarity metrics for multi-hop ad hoc networks. Ad Hoc Networks, 47:1 – 15, 2016.
- [88] M. Koseoglu, A. Bereketli, I. Yazgi, and B. Yeni. Probabilistic broadcast for dense auv networks. In OCEANS 2016 MTS/IEEE Monterey, pages 1–5, Sep. 2016.
- [89] Alper Bereketli, Muharrem Tümçakır, and Burcu Yeni. P-auv: Position aware routing and medium access for ad hoc auv networks. *Journal of Network and Computer Applications*, 125:146 – 154, 2019.
- [90] Brad Williams and Tracy Camp. Comparison of broadcasting techniques for mobile ad hoc networks. In Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Amp; Computing, MobiHoc '02, pages 194–205, New York, NY, USA, 2002. ACM.
- [91] Fei Dai and Jie Wu. Performance analysis of broadcast protocols in ad hoc networks based on self-pruning. In 2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04 TH8733), volume 2, pages 802–807 Vol.2, March 2004.
- [92] S. Basagni, M. Mastrogiovanni, and C. Petrioli. A performance comparison of protocols for clustering and backbone formation in large scale ad hoc networks. In 2004 IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE Cat. No.04EX975), pages 70–79, Oct 2004.
- [93] B. Blywis, M. Gunes, F. Juraschek, and S. Hofmann. Gossip routing in wireless mesh networks. In 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pages 1572–1577, Sept 2010.
- [94] Bastian Blywis, Mesut Güneş, Sebastian Hofmann, and Felix Juraschek. A study of adaptive gossip routing in wireless mesh networks. In Ad Hoc Networks - Second International Conference, ADHOCNETS 2010, Victoria, BC, Canada, August 18-20, 2010, Revised Selected Papers, pages 98–113, 2010.

- [95] B. Blywis, M. Guenes, F. Juraschek, and O. Hahm. Challenges and limits of flooding and gossip routing based route discovery schemes. In *Local Computer Networks* (*LCN*), 2011 IEEE 36th Conference on, pages 283–286, Oct 2011.
- [96] B. Williams, D. P. Mehta, T. Camp, and W. Navidi. Predictive models to rebroadcast in mobile ad hoc networks. *IEEE Transactions on Mobile Computing*, 3(3):295–303, July 2004.
- [97] Hao Zhang and Zhong-Ping Jiang. Performance analysis of broadcasting schemes in mobile ad hoc networks. *IEEE Communications Letters*, 8(12):718–720, Dec 2004.
- [98] Hao Zhang and Zhong-Ping Jiang. Modeling and performance analysis of ad hoc broadcasting schemes. *Performance Evaluation*, 63(12):1196 1215, 2006.
- [99] Kumar Viswanath and Katia Obraczka. Modeling the performance of flooding in wireless multi-hop ad hoc networks. *Computer Communications*, 29(8):949 – 956, 2006. Performance Evaluation of Wireless Networks and Communications.
- [100] Theodore S. Rappaport. Wireless Communications Principles and Practice (2nd Edition)Hardcover. Prentice Hall PTR, 2 edition edition, 2001.
- [101] Heitor S. Ramos, Azzedine Boukerche, Alyson L.C. Oliveira, Alejandro C. Frery, Eduardo M.R. Oliveira, and Antonio A.F. Loureiro. On the deployment of large-scale wireless sensor networks considering the energy hole problem. *Computer Networks*, 110:154 – 167, 2016.
- [102] S. Savazzi, L. Goratti, D. Fontanella, M. Nicoli, and U. Spagnolini. Pervasive uwb sensor networks for oil exploration. In Ultra-Wideband (ICUWB), 2011 IEEE International Conference on, pages 225–229, Sept 2011.
- [103] S. Savazzi and U. Spagnolin. Synchronous ultra-wide band wireless sensors networks for oil and gas exploration. In *Computers and Communications*, 2009. ISCC 2009. IEEE Symposium on, pages 907–912, July 2009.
- [104] F. Zhixin and Y. Yue. Condition health monitoring of offshore wind turbine based on wireless sensor network. In 2012 10th International Power Energy Conference (IPEC), pages 649–654, Dec 2012.
- [105] C. Bajracharya, R. Grodi, and D. B. Rawat. Performance analysis of wireless sensor networks for wind turbine monitoring systems. In *SoutheastCon 2015*, pages 1–4, April 2015.
- [106] Gyula Simon, Miklós Maróti, Ákos Lédeczi, György Balogh, Branislav Kusy, András Nádas, Gábor Pap, János Sallai, and Ken Frampton. Sensor network-based countersniper system. In Proceedings of the 2Nd International Conference on Embedded Networked Sensor Systems, SenSys '04, pages 1–12, New York, NY, USA, 2004. ACM.
- [107] Lakshman Krishnamurthy, Robert Adler, Phil Buonadonna, Jasmeet Chhabra, Mick Flanigan, Nandakishore Kushalnagar, Lama Nachman, and Mark Yarvis. Design and

deployment of industrial sensor networks: Experiences from a semiconductor plant and the north sea. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*, SenSys '05, pages 64–75, New York, NY, USA, 2005. ACM.

- [108] Kok-Kiong Yap, Vikram Srinivasan, and Mehul Motani. Max: Human-centric search of the physical world. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*, SenSys '05, pages 166–179, New York, NY, USA, 2005. ACM.
- [109] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal. Wireless sensor network survey. *Computer Networks*, 52(12):2292 – 2330, 2008.
- [110] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393 – 422, 2002.
- [111] Y. Sasson, D. Cavin, and A. Schiper. Probabilistic broadcast for flooding in wireless mobile ad hoc networks. In Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE, volume 2, pages 1124–1130 vol.2, March 2003.
- [112] L. Nduwayo, R. Lindebaum, and N. Chetty. Suppressed bond-site percolation. Computer Physics Communications, 180(4):503 – 508, 2009. Special issue based on the Conference on Computational Physics 2008CCP 2008.
- [113] D.G. Reina, M. Günes, and S.L. Toral. Real experimentation of probabilistic broadcasting algorithms based on dissimilarity metrics for multi-hop ad hoc networks. Ad Hoc Networks, 47:1 – 15, 2016.
- [114] Michele Nitti and Luigi Atzori. Modeling of network connectivity in multi-homed hybrid ad hoc networks. In Luigi Atzori, Jaime Delgado, and Daniele Giusto, editors, *Mobile Multimedia Communications*, pages 307–320, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [115] P. Santi. Mobility Models for Next Generation Wireless Networks: Ad Hoc, Vehicular and Mesh Networks. Wiley, 2012.
- [116] Giuseppe Colistra and Luigi Atzori. Estimation of physical layer performance in wsns exploiting the method of indirect observations. Journal of Sensor and Actuator Networks, 1(3):272 – 298, 2012.
- [117] S. Basagni, D. Bruschi, and I. Chlamtac. A mobility-transparent deterministic broadcast mechanism for ad hoc networks. *IEEE/ACM Transactions on Networking*, 7(6):799–807, Dec 1999.
- [118] Anand Kashyap, Samrat Ganguly, and Samir R. Das. Measurement-based approaches for accurate simulation of 802.11-based wireless networks. In Proceedings of the 11th International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM '08, pages 54–59, New York, NY, USA, 2008. ACM.
- [119] Qualnet simulator. http://web.scalable-networks.com/qualnet-network-simulatorsoftware. last accessed: 2019-03-28.

- [120] Mathew D. Penrose. Extremes for the minimal spanning tree on normally distributed points. Advances in Applied Probability, 30:628–639, 9 1998.
- [121] M.P. Salamanca, N. Peña, and N.L.S. da Fonseca. Impact of the routing protocol choice on the envelope-based admission control scheme for ad hoc networks. Ad Hoc Networks, 31:20 – 33, 2015.
- [122] J. Alonso-Zárate, C. Crespo, Ch. Skianis, L. Alonso, and Ch. Verikoukis. Distributed point coordination function for ieee 802.11 wireless ad hoc networks. Ad Hoc Networks, 10(3):536 – 551, 2012.
- [123] Horst Tempelmeier and Malte Bürger. Performance evaluation of unbalanced flow lines with general distributed processing times, failures and imperfect production. *IIE Transactions*, 33(4):293–302, Apr 2001.
- [124] Stefano Busanelli, Gianluigi Ferrari, and Roberto Gruppini. Recursive analytical performance evaluation of broadcast protocols with silencing: application to vanets. EURASIP Journal on Wireless Communications and Networking, 2012(1):1– 21, 2012.
- [125] F. Nourazar and M. Sabaei. Dapf: An efficient flooding algorithm for mobile ad-hoc networks. In 2009 International Conference on Signal Processing Systems, pages 594–598, May 2009.
- [126] Ahmad Mostafa, Anna Maria Vegni, and Dharma P. Agrawal. A probabilistic routing by using multi-hop retransmission forecast with packet collision-aware constraints in vehicular networks. Ad Hoc Networks, 14:118 – 129, 2014.
- [127] 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 the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking, MobiCom '98, pages 85–97, New York, NY, USA, 1998. ACM.
- [128] Michael Camarri and Jim Pitman. Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.*, 5:18 pp., 2000.
- [129] F. Forero, N. M. Peña, and N. L. S. da Fonseca. Latency reduction in probabilistic broadcast protocols for ad hoc networks. *IEEE Wireless Communications Letters*, 8(4):1268–1271, Aug. 2019.
- [130] O. Liang, Y. A. Sekercioglu, and N. Mani. A survey of multipoint relay based broadcast schemes in wireless ad hoc networks. *IEEE Communications Surveys Tutorials*, 8(4):30–46, April 2006.
- [131] O. Younis and S. Fahmy. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on Mobile Computing*, 3(4):366–379, Oct 2004.

- [132] Zeng Yuanyuan, Xiaohua Jia, and He Yanxiang. Energy efficient distributed connected dominating sets construction in wireless sensor networks. In Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing, IWCMC '06, pages 797–802, New York, NY, USA, 2006. ACM.
- [133] L. Hanzo and R. Tafazolli. A survey of qos routing solutions for mobile ad hoc networks. *IEEE Communications Surveys Tutorials*, 9(2):50–70, Second 2007.
- [134] Ronald Meester and Rahul Roy. Continuum Percolation. Cambridge Tracts in Mathematics. Cambridge University Press, 1996.
- [135] Hwei Hsu. Probability, Random Variables, and Random Processes. Schaum's Outline Series. McGraw-Hill, 1996.
- [136] Athanasios Papoulis and S. Unnikrishna Pillai. Probability, Random Variables and Stochastic Processes. McGraw-Hill Europe, 2002.
- [137] S. L. Myers R. E. Walpole, R. H. Myers and K. Ye. Probability and Statistics for Engineers and Scientists. Pearson, 2016.
- [138] D. Moltchanov. Distance distributions in random networks. Ad Hoc Networks, 10(6):1146 - 1166, 2012.
- [139] Tobias Friedrich, Thomas Sauerwald, and Alexandre Stauffer. Diameter and broadcast time of random geometric graphs in arbitrary dimensions. *Algorithmica*, 67(1):65–88, Sep 2013.
- [140] Leonard Kleinrock. Queueing Systems. Interscience. Wiley, 1975.