Analytical Models of Floating Content in a Vehicular Urban Environment

Gaetano Manzo

University of Bern & HES-SO Valais, Switzerland
gaetano.manzo@hevs.ch

Marco Ajmone Marsan

IMDEA Networks Institute, Spain & Politecnico di Torino, Italy
marco.ajmone@imdea.org

Gianluca A. Rizzo

HES-SO Valais, Switzerland
gianluca.rizzo@hevs.ch

Abstract

Among the many proposed opportunistic content sharing schemes, Floating Content (FC) is of special interest for the vehicular environment, not only for cellular traffic offloading, but also as a natural communication paradigm for location-based context-aware vehicular applications.

Previously published results on the performance of vehicular FC have mostly focused on the conditions under which content persists over time in a given region of space, without addressing other important aspects of vehicular FC performance, such as the effectiveness with which content is replicated and made available, and the system conditions that enable good FC performance.

This work presents a first analytical model of FC performance in vehicular networks in urban settings. It is based on a new synthetic mobility model (called District Mobility Model - DMM) , and it does not require a detailed knowledge of the road grid geometry.

We validate our model extensively, by comparison against numerical simulations based on real-world traces, and we prove our model accuracy under a variety of mobility patterns and traffic conditions.

Our analytical and simulation results provide evidence of the effectiveness of the FC paradigm in realistic urban settings over a wide range of traffic conditions.

---

1. Introduction

After many years of mostly academic interest, vehicular communications and networks are now one of the top priorities for several industrial segments. On the side of telecommunication equipment manufacturers, the 5G PPP (5th Generation Public Private Partnership) has identified automotive as one of the most important vertical sectors for the application of 5G technologies [1], and the IEEE LAN/MAN Standard Committee (IEEE 802) has developed a specific standard for Wireless Access in Vehicular Environments (WAVE), labeled 802.11p [2]. On the side of car manufacturers, attention is high to both connected cars providing services based on the inclusion of the vehicle in a smart environment where other vehicles, as well as roadside units, are able to exchange data with the car, and to the extraction of value from the huge amount of data generated every day by each vehicle. This surge of industrial attention to vehicular communications is surely due to the fast progress in autonomous vehicles, along the 5 levels identified by the taxonomy and definition in the SAE (Society of Automobile Engineers) document J3016 [3], but also to the increasing volume of sensor data that can be collected within cars.

While telecommunication operators are trying to propose their networks and their technologies (and in particular 5G) as the solution for all vehicular networking applications, some of the planned vehicular communication services impose very demanding QoS (Quality of Service) targets. According to the classification of 3GPP TS 23.203 [4], V2X (vehicle to anything) messages fall into QCI (Quality Class Indicators) classes 3, 75 and 79, and can tolerate a delay of up to 50 ms with packet loss rate up to 10^{-3} or 10^{-2}. However, critical applications, such as remote driving, can impose much stricter requirements, with latencies as low as 10 ms (see 3GPP TS 22.261 [5]). This implies that the use of a telecommunications infrastructure to carry the huge volume of data generated by cars for all the applications of interest may not be the wisest choice, considering the difficulty in meeting the specified QoS targets. This becomes evident if

we consider the fact that most of the data of interest are generated at the network edge (inside the vehicle or close to it) and are directed to users (other vehicles or roadside units) in proximity of the vehicle. In such a context, offloading vehicular data transfers (at least those with the characteristics above) from the telecommunications infrastructure to a direct V2V (Vehicle-to-Vehicle) opportunistic link seems reasonable, both because it reduces unnecessary infrastructure overload, and because it allows much shorter latencies between data generation and delivery.

An example of opportunistic communication paradigm for the local dissemination of information goes under the name of Floating Content (FC) [6] or Hovering Information [7]. It enables probabilistic content storing in geographically constrained locations - denoted as *Anchor Zones* (AZ) - and over a limited amount of time. Given the infrastructure-less nature of the FC paradigm, and its reliance on opportunistic exchanges among mobile nodes, a significant portion of the studies of FC performance focused on the conditions under which content persists in the AZ (i.e., "floats" for a "large enough" amount of time). [6] introduces the *criticality condition*, a sufficient condition for the content to float indefinitely with very high probability, under various mobility models. [8] introduces an analytical model to estimate content persistence in the case of outdoor pedestrian mobility over large open spaces, such as a city square. However, for any practical application, content persistence over time within the AZ is only a necessary condition for FC viability. Indeed, when the content persists, it is essential to characterize how often the FC paradigm manages to deliver the content to the intended users. Despite its importance, few works consider this issue. For setups where nodes move according to a random direction mobility model, [9, 10] characterize analytically the *success probability* (i.e., the probability of delivering content to users in transit in the AZ) as a function of system parameters. [10] shows however that those results are heavily inaccurate when applied to realistic scenarios, and in particular to vehicular setups. Several works consider FC in the vehicular environment (see e.g., [11, 12]). However, these works still focus on content persistence within the AZ. Hence, how the features of vehicular mobility patterns impact FC performance, and how to engineer a vehicular application relying on FC so as to achieve a target performance, are issues that remain to date still open.

This work is motivated by the need to better characterize the FC performance in vehicular urban environments. We focus on urban scenarios (which in a European context means built environments with approximately 50-100 inhabitants per hectare), and we propose an analytical approach to performance evaluation of FC in vehicular scenarios. Our approach is based on mapping the patterns of vehicular mobility in an urban environment into a modified version of the random waypoint (RWP) mobility model. This allows the derivation of a performance model that is not based on any specific road grid geometry. The comparison of the analytical model results against results obtained with simulations based on measurement-based mobility traces shows that our approach accurately models vehicular FC performance over a wide range of values of system parameters. Through modeling and simulations, we perform a first characterization of the relation between land uses (office/commercial, industrial, or residential), their induced vehicular mobility patterns, and FC performance. As expected, we observe that in each scenario such features of mobility as: i) the local mean node density, ii) the mean node contact rate, iii) the degree of mixing between different traffic flows, play a crucial role in determining the optimal AZ dimensioning strategy, as well as the achievable success probability, and the type of applications which FC can support. Moreover, and contrary to what was previously shown in the literature, our results suggest that in realistic urban vehicular settings the criticality condition is not a good indicator of the probability for content to float over time.

The key contributions of this paper are the following:

- We propose a new synthetic mobility model (called District Mobility Model - DMM), which captures the main characteristics of urban mobility, without requiring as input the specific geometry of the road grid;

- We apply the DMM to the performance analysis of vehicular FC, deriving analytical expressions for the average content availability and the success probability in a vehicular FC system;

- We carefully validate our analytical predictions by comparing results against those of detailed simulations driven by realistic, measurement-based mobility

parameters. We show that our approach accurately models vehicular FC performance over a wide range of values of system parameters;

- We discuss the insight provided by numerical results on the performance of FC in vehicular urban scenarios; and

- We discuss the limitations of our modeling approach, pointing out the cases in which it can be inaccurate.

The rest of this paper is organized as follows. In Section 2, we present the system model, and in Section 3 we present the main analytical results on FC success probability. In Section 4, we numerically assess the accuracy of our analytic results, and we evaluate them by simulation on a realistic setup. In Section 5, we briefly discuss previous work, and finally Section 6 concludes the paper.

2. System Model

We consider a set of wireless nodes with transmission range r that move over a region of the plane called anchor zone (AZ). We assume two nodes come in contact when they are in range of each other, i.e. when their distance is not larger than r (Gilbert's model [13]). With $C(x)$ we denote the throughput between two nodes, when they are at a distance $x, x \leq r$ from each other. $C(x)$ can be modeled using Shannon's capacity law, or any other more realistic model.

In what follows, we focus on an urban environment. Coherently with typical features of vehicles mobility in an urban setting, we assume that mobility patterns of each node alternate between time intervals spent moving, and time intervals spent still at a waypoint, which could model a pause at a crossroad, or in a parking lot.

More specifically, we assume nodes move according to the *District Mobility Model* (DMM), which abstracts from the details of the geometry of the street grid. Its formal definition is as follows:

Definition 1 (District Mobility Model). *We assume that nodes arrive according to a Poisson process with intensity λ at a point chosen uniformly at random on the border of the AZ. Nodes, inside the AZ, move according to the RWP mobility model with pause,*

5

Table 1: Table of notation.

Parameter	Description	Units
r	Node transmission range	m
$C(x)$	Throughput between the two nodes at distance x	Mb/s
λ	Nodes arrival rate	s^{-1}
v	Node speed	m/s
p	Probability of the next waypoint with the AZ	
T_m	Moving time within an epoch	s
T_s	Stopping time within an epoch	s
T_t	Moving time (when crossing the AZ without pause)	s
R	AZ radius	m
ν	Frequency of contact between two nodes	s^{-1}
Q	Probability of successful content exchange	
$\overline{N}$	Mean number of nodes within the AZ	
$\overline{n}$	Mean number of nodes within the AZ with content	
$\overline{m}$	Mean number of nodes within the AZ without content	
P_{succ}	Probability to obtain the content	
P_e	Probability to obtain the content in an epoch	
P_m	Probability to obtain the content during moving time	
P_s	Probability to obtain the content during stopping time	
L	Content size	MB
B	Bandwidth	MHz
α	Coefficient of attenuation	cm^{-1}
N_0	Power of noise spectral density	W/Hz
P	Transmit power	W

with velocity v, constant and equal for all nodes (though our approach can be easily generalized to scenarios where velocity varies between waypoints according to a given distribution).

Upon reaching a waypoint, each node pauses for a random duration. Then, with probability p, the location of the next waypoint is selected uniformly at random within the AZ. Otherwise, with probability $1-p$ the location of the next waypoint is selected uniformly at random on the border of the AZ. Once the border is reached, a node disappears from the AZ.

As in RWP, in the DMM the sojourn time of every node within the region consists in a succession of *epochs*, where each epoch is the sum of the time spent moving between two consecutive waypoints, (the *moving time*), and of the time spent still at the destination waypoint (the *pause time*). The duration of moving time, T_m and pause

6

time, T_s, are assumed to be independent random variables. We denote their pdf with f_{T_m} and f_{T_s}, respectively. We assume such pdfs to be the same for all nodes. When a node does not stop into the region, its sojourn time is spent only moving from one point to the other of the border. We indicate such epoch as a *traversal epoch*. With T_t and f_{T_t} we denote, respectively, its duration and the associated pdf. Finally, given two nodes, of which at least one moving, the *trajectory distance* is the minimum value that the distance between the two nodes takes while they are in range of each other.

One of the distinctive features of vehicular mobility is being strongly influenced by the geometry of the road grid, and the related constraints on node speed and direction. As a consequence, most existing approaches to vehicular FC modeling focus on specific road geometries, such as highways, highway junctions, or Manhattan grids, deriving results which are hardly generalizable [11].

Different from the typical vehicle mobility over a street grid, the DMM is isotropic (i.e., at any point in space, all directions can be chosen with equal probability) and unconstrained (i.e., a vehicle can occupy any location within the area). This is the reason why we say that the DMM abstracts from the details of the geometry of the road grid. However, when a "sufficiently large" portion of the road grid is included inside an AZ (say, a few blocks), even if the possible instantaneous directions of movement are finite, for any two points chosen at random on the map (on the road grid), often there exists at least one path between them (albeit usually not on a straight line). Hence, both with DMM and in reality the movement between two points is possble. On a macroscopic scale, and for the purpose of modeling content replication and diffusion dynamics within the AZ, the approximations introduced by the DMM only impact the mean moving time between two waypoints.

2.1. Floating Content basic operation

We assume that at a time t_0, a node in the plane (the *seeder*) defines a circular area of radius R, the *Anchor Zone* (AZ), containing the node itself. At time t_0 the seeder (blue node in Figure 1a) generates a piece of content (e.g., a text message, a picture, etc.). For $t \geq t_0$, every time a node with the content comes in contact with a node without it within the AZ, the content is exchanged successfully with probability
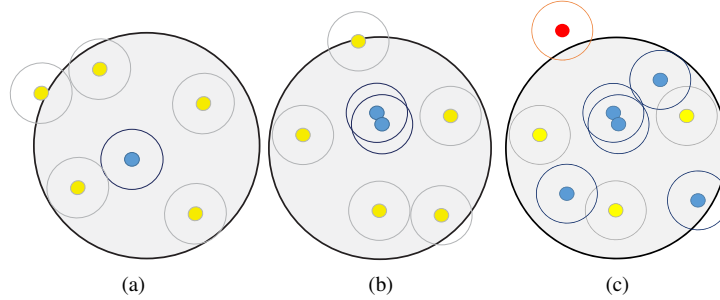
7

Figure 1: Basic operation of Floating Content. 1a) Seeder (blue) defines the AZ. 1b) Opportunistic message exchange between nodes. 1c) Nodes going out of the AZ (red) discard the content.

Q (Figure 1b). We assume that the time taken to replicate the content is negligible with respect to contact time. Nodes entering the AZ do not possess a copy of the content, and those exiting the AZ discard their copy. Hence, the typical behavior of a node consists in entering the AZ with no content, receiving the content from another node, distributing the content to nodes with no content met within the AZ, leaving the AZ and dropping the content. If the node re-enters the AZ, it is treated as a new node and therefore it receives the same content when a node with content is met. As a result of such opportunistic exchange, the content *floats* (i.e., it persists probabilistically in the AZ even after the seeder has left the AZ). In this way, the content is made available to nodes traversing the AZ (Figure 1c) for the whole duration of its floating lifetime.

A first performance parameter of FC is content *availability* at a given time, i.e. the ratio between the number of nodes with content over the total amount of nodes inside the AZ at that time. Indeed, a high value of availability is correlated with low likelihood of content disappearance from the AZ, but also with the probability of getting the content for a node entering the AZ. This last feature is captured by the *success probability*, the probability that a node entering the AZ will receive a copy of the content before leaving the AZ. This is our main performance indicator, and the one which is most directly related to the performance of applications and services relying on FC. Indeed, it is a direct indicator of the effectiveness with which the floating content is delivered and made available to nodes traversing the AZ.

Another relevant performance parameters is the mean time to get the content from

entering the AZ, which is specially important for safety applications requiring that the message reaches nodes as soon as possible.

3. An analytical model for success probability

In this section, we introduce the analytical results which relate the success probability to the main system parameters.

For our derivation, we consider a system in equilibrium, in which the mean number of nodes inside the AZ does not change over time. For nodes mobility within the AZ, we consider a district mobility model, with area coinciding with the floating content AZ. The sojourn time for a node in the AZ is given by the sum of the durations of the epochs spent within the AZ, plus the time to enter and exit the AZ, weighted with their probabilities:

$$T_{soj} = (1-p)\left[T_t + \sum_{k=1}^{\infty} p^k [k(E[T_m] + E[T_s]) + E[T_m]]\right] \tag{1}$$

and $\overline{N} = \lambda T_{soj}$ is the mean number of nodes in the AZ, by Little's law.

With q we indicate the mean fraction of the sojourn time in which a node moves:

$$q = \frac{(1-p)\left[T_t + \sum_{k=1}^{\infty} p^k (kE[T_m] + E[T_m])\right]}{T_{soj}}$$

Note that when p tends to 1, q is well approximated by $\frac{E[T_m]}{E[T_m]+E[T_s]}$.

In order to derive an analytical expression for success probability, we start with the following two results, which determine the frequency with which any two nodes come within the range of each other in the AZ, and the content availability (and hence the mean number of nodes with content in the AZ).

Lemma 1 (Frequency of contacts). *In the DMM, considering a region with area A, the mean frequency of contacts between any two nodes, with mean speed v and transmission range r, is*

$$\nu = \frac{2rqv[2(1-q) + 1.27q]}{A} \tag{2}$$

For the proof of Lemma 1, please refer to Appendix A.

In general, the number of nodes with content in the AZ at a given time $t \geq 0$, $n(t)$, and without content, $m(t)$, are random variables. In what follows, we indicate with $\bar{n}(t)$, and $\bar{m}(t)$ their stochastic averages, and we consider what happens to these quantities when the transient due to the initial spreading of the content within the AZ is exhausted. Let $\bar{n}$, and $\bar{m}$ denote the values taken by the previous quantities *in stationary regime*. We have the following result.

Lemma 2 (Content availability). *For $R >> r$, when the criticality condition*

$$\overline{N}T_{soj}\nu > 1 \tag{3}$$

is satisfied, the stationary mean number of nodes in the AZ with content, $\bar{n}$, is given by

$$\overline{n} = \bar{N} - \frac{1}{T_{soj}\nu Q}$$

The mean number of nodes without content, denoted with $\bar{m}$, is

$$\overline{m} = \frac{1}{T_{soj}\nu Q} \tag{4}$$

For the proof of Lemma 2 we refer the reader to Appendix B.

We now present our main result, which relates the success probability to the probability for a generic node to get the content during the epochs of its sojourn time inside the AZ.

Theorem 1 (Success Probability). *In the DMM, when $R >> r$ and the criticality condition (3) is satisfied, the probability for a node to get the content during its sojourn time within the AZ, in stationary regime is*

$$P_{succ} = (1 - p)P_t + \frac{p}{1 - p(1 - P_e)}[pP_e + P_m(1 - p)] \tag{5}$$

where P_e is the probability that a node gets the content during an epoch (other than the final one), given by

$$P_e = P_m + (1 - P_m)P_s \tag{6}$$

10

P_s is the probability of getting the content during the pause time, given by

$$P_s = \int_0^{+\infty} \left(1 - e^{-\nu \tau \bar{n} Q}\right) f_{T_s}(\tau) d\tau \qquad (7)$$

P_m is the probability of getting the content during the moving time, given by

$$P_m = \int_0^{\frac{2R}{v}} \left(1 - e^{-\nu \tau \bar{n} Q}\right) f_{T_m}(\tau) d\tau \quad . \qquad (8)$$

Where the pdf of T_m is given by

$$f_{T_m}(\tau) = \frac{4\tau v^2}{\pi R^2} \left(\arccos \frac{\tau v}{2R} - \frac{\tau v}{2R} \sqrt{1 - \left(\frac{\tau v}{2R}\right)^2} \right) \qquad (9)$$

P_t is the probability of getting the content during the transition epoch. Its expression is the same as (8), with f_{T_t} instead of f_{T_m}:

$$f_{T_t}(\tau) = \frac{\tau v^2}{2R\sqrt{4R^2 - \tau^2 v^2}} \qquad (10)$$

For the proof of Theorem 1 we refer the reader to Appendix C.

Note that the epoch in which the node exits the AZ coincides with the time spent moving towards the border of the AZ, as the node is assumed to disappear once it has reached the border. Hence for the final epoch $P_e = P_m$.

In realistic settings, every time two nodes come in range of each other, a substantial amount of time is required by each node to detect the presence of the other node, and to establish a connection. To such time overhead (which is in the range of a few seconds in Wi-Fi Direct, up to 10-12 s in Bluetooth [14], and sensibly inferior in DSRC [15]), it adds up the time for content transfer, which is a function of content size L and of the capacity between the two nodes during the contact time.

The following result provides a way to include such effects into the performance model of vehicular FC. Let τ_0 be the mean time taken by two nodes in contact to establish a connection and coordinate for content transfer.

11

Theorem 2. *When $r << R$, and T_s is exponentially distributed with mean $1/\mu$, the probability of successful content exchange Q is given by*

$$Q = 2(1-q)q\frac{h_0^s}{r} + q^2\frac{h_0^r}{r} + (1-q)^2 \exp\left[2\mu\left(\tau_0 + \frac{L}{C_0}\right)\right] \qquad (11)$$

where

- h_0^s *(respectively, h_0^r) is the maximum trajectory distance beyond which no successful content transfer can take place, when only one node is moving (respectively, when both nodes are moving). h_0^s is the solution of the following equation:*

$$L = \int_{\tau_0}^{\frac{2\sqrt{r^2-(h_0^s)^2}}{v}} C\left(\sqrt{(h_0^s)^2 + \left(\sqrt{r^2 - (h_0^s)^2} - vt\right)^2}\right) dt \qquad (12)$$

- h_0^r *is the expectation of the maximum trajectory distance h^r with respect to the distribution of the relative speed v_r between two moving nodes:*

$$h_0^r = \int_0^{2v} h^r(v_r) f_r(v_r) dv_r$$

where $f_r(v_r)$ is the pdf of the relative speed between two moving nodes:

$$f_r(v_r) = \frac{v_r}{v^2\pi\sqrt{1-(v_r/v-1)^2}}$$

and where $h^r(v_r)$ is the solution of the following equation:

$$L = \int_{\tau_0}^{\frac{2\sqrt{r^2-(h^r(v_r))^2}}{v_r}} C\left(\sqrt{(h^r(v_r))^2 + \left(\sqrt{r^2 - (h^r(v_r))^2} - v_r t\right)^2}\right) dt$$
$$(13)$$

- $C(x)$ *is the channel capacity between the two nodes, as a function of their distance* x.

- C_0 *is the mean throughput between two static nodes:*

$$C_0 = \frac{2}{r^2} \int_0^r x C(x) dx$$

For the proof of Theorem 2, please refer to Appendix D.

Note that Theorem 2 assumes that on a contact between two nodes, the time for the content to be successfully transferred is not affected by other already ongoing content transfers involving the two nodes. This is a reasonable assumption for setups with low node densities, i.e., setups where it is unlikely that a node is simultaneously in contact with several other nodes, and with $r << R$, while bringing to optimistic performance estimations in all other cases. Finally, note that results in this section have been derived under the *non-spatial approximation*, by which, for the sake of analytical tractability, nodes have been assumed to be uniformly distributed within the AZ. In Section 4, we numerically assess the impact of such assumptions on the model accuracy.

4. Simulation and validation

In this section, we illustrate the approach we followed for the validation of our model, and for the numerical assessment of the performance of FC in a vehicular environment, based on two scenarios. In a first scenario, we consider nodes moving according to the DMM, and we compare simulation results with the analytical predictions of our model. Our goal is to characterize FC performance as a function of the main system parameters, and to validate Theorem 1 by assessing the impact of those effects, such as clustering of nodes with content and border effects, which are not included in our model.

In a second scenario, in order to perform a more realistic assessment, we consider a setup where mobile users model vehicles on the streets of two European cities, moving according to a set of measurement-based mobility traces.

In both scenarios, simulations are performed using the VEINS framework [16], based

13

on OMNET++ [17] for network simulation, and on SUMO [18] for road traffic simulation.

Unless otherwise stated, we assume that content is transferred (with probability $Q = 1$) every time two nodes come in the range of each other, regardless of the amount of time spent in range. This corresponds to the case in which the content size is small enough for content transfer time to be much smaller than the typical mean contact time in urban scenarios. This is the case, for instance, of such applications as traffic warning, accident warning, or textual advertisements. We assume T_s to be exponentially distributed, with mean $1/\mu$. In what follows, we model $C(x)$ using Shannon's capacity law (though any other, more realistic model can be used):

$$C(x) = B \log_2 \left(1 + \frac{P x^{-\alpha}}{N_0} \right)$$

where α is the attenuation coefficient, N_0 the power spectral density of the additive white Gaussian noise, and B is the available bandwidth. We do not include the effects of fading and interference because transmission ranges are typically very short. Note however that our approach can be easily adapted to incorporate such effects.

Finally, with respect to the expected validity time of the floated content, we group vehicular services relying on FC into two broad categories. The "near real-time" category typically floats messages with a very short validity time (a few minutes). Examples are situated introductions, or infrastructure-less ridesharing, where passing cars inform neighboring pedestrians of their availability and their planned trip. Services in the "delay tolerant" category, instead, are associated with events with slower dynamics, and hence they require longer floating lifetimes of content inside the AZ (of the order of one-two hours).

4.1. Baseline Scenario

In order to assess the accuracy of our model, we performed a steady-state analysis of the system. Each simulation run started with an empty AZ. Then, after waiting long enough for the transient on node population to be exhausted, we assumed that the closest node to the center of the AZ generates a message, and it starts replicating
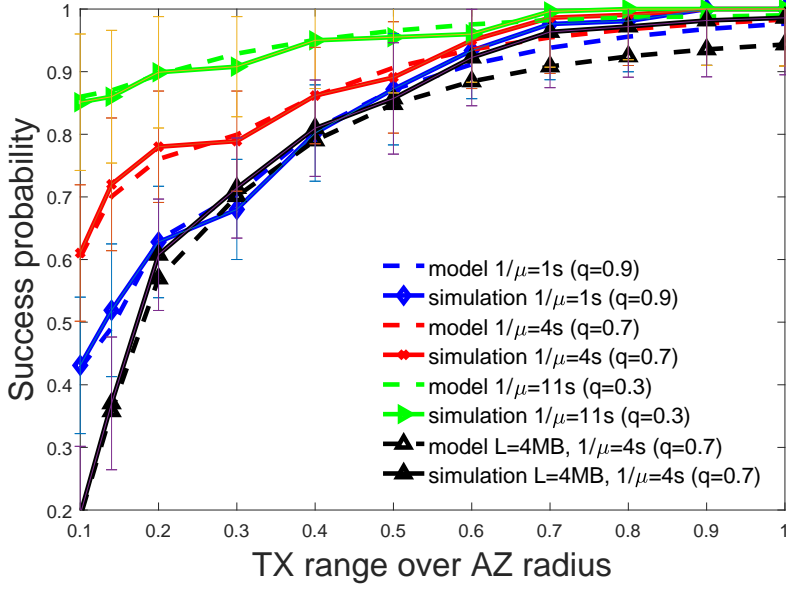
Figure 2: Success probability over transmission radius, with 95% confidence intervals.

it opportunistically. Finally, once the transient on content population in the AZ ends, we measure content availability within the AZ, averaged over two hours. Being this a steady-state analysis, we only retained data for those contents which float. Moreover, for the welfare of homogeneity in measured data, we retained only simulation results for which the content floats for the whole simulation time. Note that, since the criticality condition was satisfied in all simulation setups, the vast majority of the content fluctuated for the whole simulation time.

In addition, for each simulation, we measured the average *success rate* over the simulation time. The average success rate is defined as the fraction of nodes which leave the AZ with a copy of the content during the simulation time, and it is an estimator of the success probability for the given content.

Unless otherwise stated, we considered an AZ radius $R = 50$ m, arrival rate $\lambda = 0.1 \ s^{-1}$, and a mean moving time 9 s. The probability p for a node to remain in the AZ after a pause has been set to 0.9. In Figure 2 and 3, respectively, we plot the values of success probability and availability derived from simulations, as functions of the ratio between the transmission range r and the AZ radius R, for different values of mean
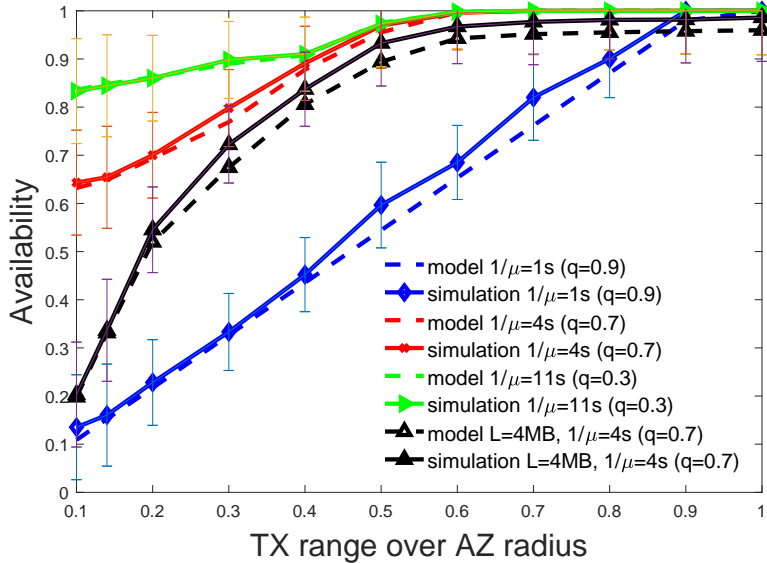
Figure 3: Availability over transmission radius, with 95% confidence interval.

pause time $1/\mu$, and hence of the fraction of moving nodes q.

In both plots, we compare simulation results with analytical predictions.
As the plots show, both performance parameters increase monotonically as a function
of r/R. Indeed, both are directly related to contact rate, and from Lemma 1 we know
that contact rate is directly proportional to r/R.

The range of values of r/R was chosen as follows. Values of r/R below 0.1 do not
allow the content to float for more than a few minutes (the criticality condition is not
satisfied). When $r/R > 1$ instead, even at very low node densities, nodes within the
AZ form a strongly connected component, bringing success probability very close to
one.

Indeed, when r/R gets closer to one, node clusters start to emerge, lowering the
estimates of success rate from our model with respect to values from simulation. How-
ever, this effect is counterbalanced by border effects. Indeed, in our derivation, we
discarded the fact that users close to the border of the AZ have part of their coverage
area lying outside of the AZ. Ignoring this effect brings to overestimate the replication
opportunities of a content with the AZ. Such effect, typically negligible for low values

of r/R, starts to have some impact on the accuracy of our model for values of r/R close to one. Moreover, in the derivation of contact rate, we assumed that node directions are all equally likely. For the DMM, this is a good approximation when nodes are close to the center of the AZ, while bringing to an overestimate of contact rate (and hence of content exchange opportunities), when nodes are close to the border of the AZ. Figures 2 and 3 show a good accuracy of our model in predicting FC performance, across various system configurations. Indeed, the overall difference between analysis and simulation is always less than 5%. Note that, for the same value of r/R, and keeping constant all other system parameters, an increase of $1/\mu$ has two contrasting effects on success probability. On the one side, it increases the mean sojourn time, increasing the mean number of nodes in the AZ and therefore (at least in principle) the content replication opportunities. On the other, it also decreases the fraction q of nodes which at any time instant are moving within the AZ, thus decreasing the overall contact rate in the AZ. As we can see from the plots, the net result is an increase of success probability and success rate.

We also see that the marginal benefit of increasing r/R is higher for high values of q, indicating that the more mobile nodes there are, the higher are the advantages of a large transmission range.

Figure 3 shows that availability has a similar dependency with respect to r/R and q as success probability. The figure also shows that there is no direct proportionality between availability and success probability, and that typically availability is lower than success probability (indeed, success probability is the availability of nodes leaving the AZ, which typically have more chances of possessing the content than the average node in the AZ). Hence, as the plots suggest, and despite being often chosen as the main FC performance parameter, availability is, in general, a poor indicator of success probability. We can also see that, increasing the mean amount of time spent moving by a node allows decreasing the availability required for achieving a given value of success probability.

Finally, we analyzed the impact of finite content size and capacity between two nodes on content diffusion and on FC performance. We assumed an attenuation $\alpha = 3$ (typical of urban scenarios), a transmitter SINR of 5 dB (which models setups with
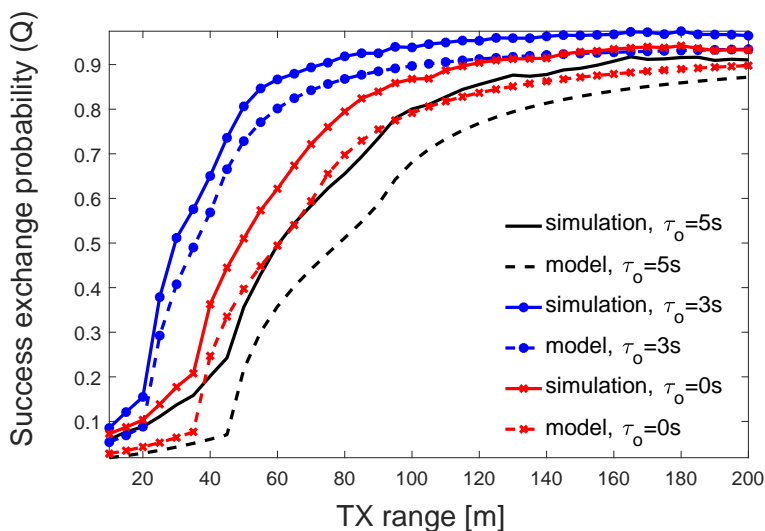
Figure 4: Probability of successful content exchange Q vs transmission range, for three values of initial delay τ_0. Simulations are with a 95% confidence interval of 4%.

high interference, common in crowded city centers, e.g. at the 2.4GHz frequency), a content of size $L = 5$ MB, to emulate the exchange of multimedia content such as pictures and short videos, and a bandwidth $B = 1$ MHz (as in Bluetooth). Note however that, as results are invariant with respect to the ratio between content size and bandwidth, they hold also for longer range WiFi transmissions of 50 MB files over a 10 MHz bandwidth. As for mobility, we have assumed a speed of 10 m/s (typical of city centers out of congestion), and mean pause time of 10 s (which models typical delays at crossroads and junctions). The initial delay before content exchange τ_0 was varied between zero and 5 s, to take into account connection setup delays, such as those experienced on Bluetooth and WiFi. In Figure 4, we plot the values of the probability of successful content exchange Q obtained from simulation and those derived from Theorem 2. As we can see, the model maintains high levels of accuracy even for values of transmission range comparable to the radius of the AZ. We see that, in all configurations, the model is slightly pessimistic with respect to simulations. This is due to the fact that the model neither considers those contacts during which a change in mobility (from static to moving, or vice versa) takes place, nor it includes border

(a) Luxembourg area [19]
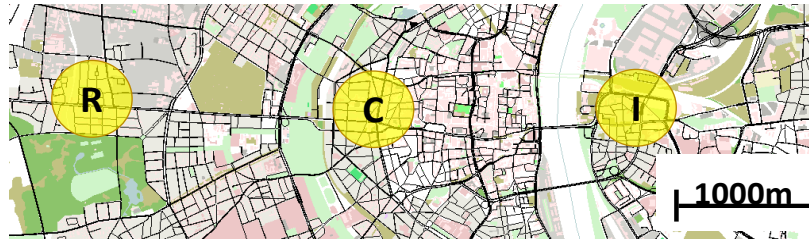
(b) Cologne area [20]

Figure 5: The areas of Luxembourg City and Cologne considered in our evaluations. In each area, the yellow circles correspond to three AZ with $R = 260m$, located in the city center (C), in the industrial district (I), and in a residential district (R).

effects. Indeed, the loss of accuracy increases with increasing τ_0 because those effects are more likely to impact those contacts which last longer.

The effects of finite content size and limited channel capacity on FC performance are also visible in Figures 2 and 3, where we plot success probability and availability for a content of size $4MB$. As expected, the impact of these factors on performance is relevant for values of mean contact time less than or equal to the mean time taken by a contact transfer, while slowly decreasing with increasing transmission range.

4.2. Assessment in urban scenarios

In order to test our model in a realistic context, and to characterize those conditions in which FC is feasible in a vehicular environment, we considered two European city scenarios, in which nodes move on an urban road grid, according to measurement-based vehicular mobility traces.

Specifically, we consider a section of the area of Luxembourg City, and of Cologne,

19

Table 2: Main mobility parameters in the Luxembourg and Cologne scenarios, for the three AZs in Figure 5 for low-density $(2.30AM - 5.30AM)$ and high-density $(7AM - 9AM)$ traffic periods. AZ radius $R = 260$ m, transmission range $r = 20$ m.

| City Scenario			Sojourn Time T_{soj}	Contact Rate $\nu(N-1)$	N. of nodes in the AZ N	Arrival rate λ
Luxembourg	C	low traffic	335.7	0.04	29.5	0.087
		peak traffic	368.4	1.16	489.2	1.327
	I	low traffic	412.2	0.05	38.7	0.0938
		peak traffic	440.8	0.72	512.1	1.161
	R	low traffic	189.6	0.0063	1.9	0.010
		peak traffic	192.4	0.043	12.9	0.067
Cologne	C	low traffic	192.2	0.037	20.7	0.107
		peak traffic	245.8	0.75	294.9	1.257
	I	low traffic	258.3	0.029	32.9	0.128
		peak traffic	311.8	0.81	326.5	1.051
	R	low traffic	124.3	0.0011	2.6	0.021
		peak traffic	136.1	0.018	10.7	0.079

Germany. The Luxembourg scenario (Figure 5a) consists of an surface of $155.95\ km^2$, which includes the city center as well as part of its surroundings. The street grid and the measurement-based mobility traces for this scenario for a 24-hour period are derived from [19]. The Cologne scenario extends over an area of $400\ km^2$, (Figure 5b) in which the road grid is derived from [21], while mobility traces are derived from the TAPASCologne project [20]. Given the heterogeneity of the urban environment, in order to capture the effects on FC performance of a specific road structure and of the resulting mobility patterns, in both settings we chose three different locations as AZ centers. One is set in the city center (downtown, area "C"), another in a residential area (area "R"), and a third in the industrial district (area "I"). In order to take into account temporal fluctuations in the mobility patterns of these cities, we considered two different time intervals over the course of the 24 hours. A first one, from 7AM to 9AM, corresponds to a period of peak traffic, due to people commuting to work. A second time interval was chosen early in the morning, from 2.30AM to 5.30AM. Being this typically a period of very low car density, it usually represents a worst-case scenario for FC performance. For both time intervals, we chose a duration that is of the same order of magnitude of the typical validity of messages in such services as traffic congestion

warnings, or accident warnings.

In Table 2 we list some of the main parameters of mobility for the three AZs, for an AZ radius of 260 m. These values show that, within each city, the three districts differ substantially in terms of mean node density, mean sojourn time, and mean contact rates. More specifically, in both cities, the industrial district and the city center have
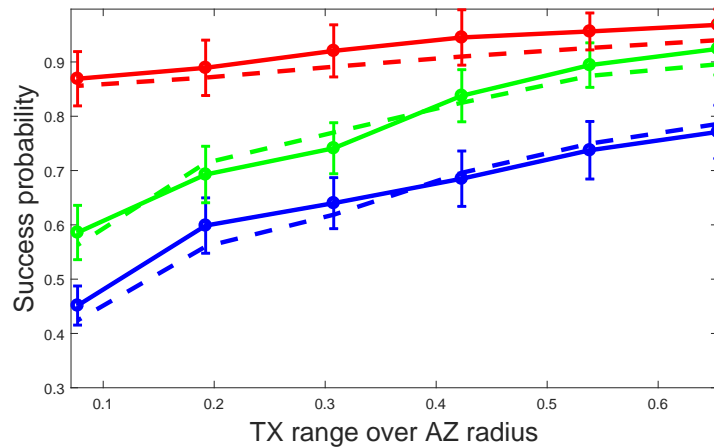
(a) Luxembourg
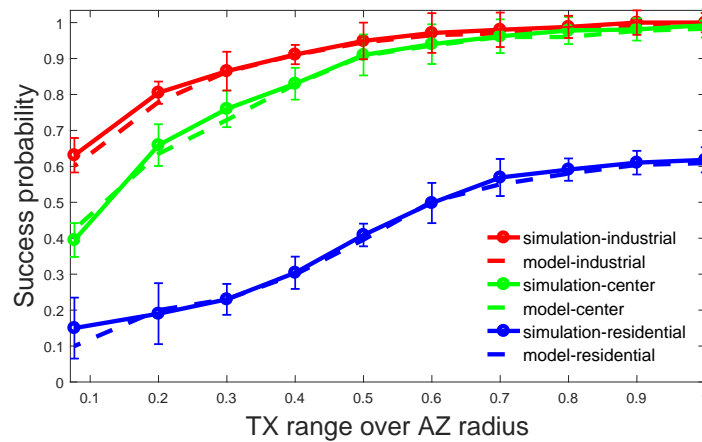
(b) Cologne

Figure 6: Success probability over transmission radius, in the three AZs, for low traffic periods in the Luxembourg and Cologne scenarios. 95% confidence interval.

.

a sensibly higher node density and contact rate than the residential district. Indeed, as

the maps in Figure 5 show, typically the residential zone is mainly a transit zone for vehicles, with little local traffic, and with few main roads carrying the majority of the traffic of the area. This implies, on average, shorter sojourn times and hence fewer opportunities for content replication with respect to other areas of a city. Despite these differences, however, we observed that mean pause time, mean moving time, and mean speed are essentially the same for the three districts in Luxembourg City (and equal to about 15 s, 25 s, and $v = 14$ m/s, respectively), and they do not vary significantly in the course of the 24 hours. The Cologne scenario is characterized, in each AZ, by a similar arrival rate as the corresponding AZs in Luxembourg city. However, due to differences in road grid geometry (and in particular to an average block size larger than in Luxembourg City), it is also characterized by a lower average stopping time, a higher average speed (16 m/s), a longer moving time (30 s), and a lower mean contact rate. The latter feature is probably related to the fact that the blocks, being larger on average, make it more difficult for nodes on opposite sides of a block to come in range of each other. Note that, for the computation of the mean pause time, we assumed that cars which are parked do not participate in content exchange.

In each simulation run, the adopted seeding strategy is as follows. Among all nodes passing in the AZ in the first 5 minutes of the given time interval, we choose at random a node as seeder. Then, from the moment in which the seeder enters the AZ, we simulate the process of content replication and floating for the whole duration of the time interval. Such strategy models those FC application contexts in which a single vehicle is the producer of the content. This is the case, for instance, of infrastructure-less car sharing, in which each car offering a lift advertises its availability (plus, possibly, its approximate location and trip plan), or of parking management applications, in which a car leaving a parking location advertises the location of the available parking spot.

For each of the two cities, Figure 6 shows success probability for the three districts, for an AZ radius of 260 m, and for the low traffic time interval, as a function of the ratio between transmission range and AZ radius. These plots show that, even in scenarios with realistic mobility patterns, simulation results are in good agreement with analytical values of success probability computed with the model in Section 3. The fact that our results maintain a high accuracy for very different road grid scenarios, and mobility

patterns suggests that the district mobility model is effective in capturing those aspects of vehicular mobility, which have a significant impact on FC performance in urban settings.

In the industrial area, the node density and mobility are such that even with a small transmission range (26 m) success probability is very high (above 0.8 in Luxembourg, and slightly lower in Cologne). As for the city center AZ, despite having only marginally lower node density, it exhibits a markedly smaller success probability at low values of r/R. Similarly to what we already saw in the baseline scenario, in scenarios with smaller sojourn times, increasing the ratio between transmission range and AZ radius brings to larger marginal increases in success probability. Finally, in both cities the plots indicate that the residential area is the most critical for FC performance, requiring a large ratio r/R to achieve high success probabilities. Note that in practical settings, values of r/R close to one defeat the purpose of having an AZ, as the vast majority of users who should get the content are within transmission range.

We also note that the higher traffic "fluidity" in Cologne, with less frequent stops, and with a slightly less dense road grid, brings in each district to generally inferior performance in terms of success probability and availability. The higher traffic regularity brings however also to a higher level of accuracy between model and simulations, for a same number of simulations, as averages of system parameters and of mobility parameters are more representative of the actual mobility features in the considered observation window.

The strategy for content seeding can have a significant impact on the performance of FC. In order to show the effect of seeding, we look at the AZ in the Luxembourg city center in the high traffic period (from $7AM$ to $9AM$). We consider an AZ with radius $R = 260$ m and a transmission radius $r = 160$ m, with 94 nodes in the AZ at time $t = 0$. In Figure 7, we report the evolution of the success probability versus time under four different seeding strategies. Each point in the graph is an ensemble average of 30 simulation runs, with a further average over 30 s. In the first case, all nodes in the AZ receive the content at $t = 0$, for example with a broadcast message issued by the cellular infrastructure. In the other cases, the fractions of nodes that receive the content are 75%, 50%, 25% or just one node. We clearly see that the seeding strategy only
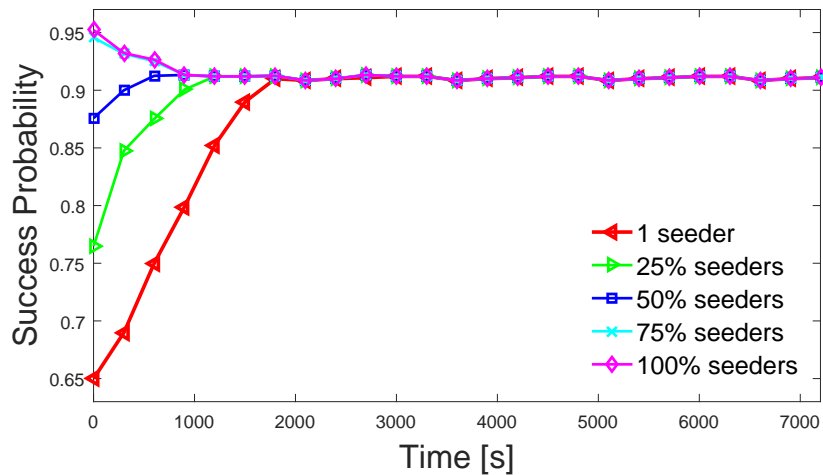
Figure 7: Success Probability over simulation time. Several seeding strategies in the city center of Luxembourg City at high-density traffic ($7AM - 9AM$). Anchor Zone radius 260 m, transmission radius 160 m. At $t = 0$ there are 94 nodes within the AZ. Mean over sample population of 20 (each point) every 300 s.

affects the duration of the transient before the success probability stabilizes to slightly more than 0.9. A similar, but more noisy behavior is observed in Figure 8, where we report the evolution of the availability versus time in the same conditions.

The fact that the availability plot is much noisier than the success probability is due to the fact that short-term fluctuations in the overall population of nodes as well as in the population of nodes with content within the AZ have an impact only for a fraction of the time that a node spends in the AZ. While this is enough to generate fluctuations in availability, the probability of getting out of the AZ with the content depends on the compound effect of many such fluctuations along the whole sojourn time of a user in the AZ. As it can be seen from Figure 7, such compound effect brings to smoothing out fluctuations in success probability over time.

One of the main features of mobility which are not accounted for by our model is node clustering, which typically brings to a higher rate of contacts, and hence to a larger amount of opportunities for content exchange with respect to scenarios which are homogeneous in space. In particular, given that our approach is based on mapping mobility features to a synthetic mobility model, such as the DMM, which tends to distribute nodes evenly in the AZ, we expect our approach to yield substantially pessimistic es-
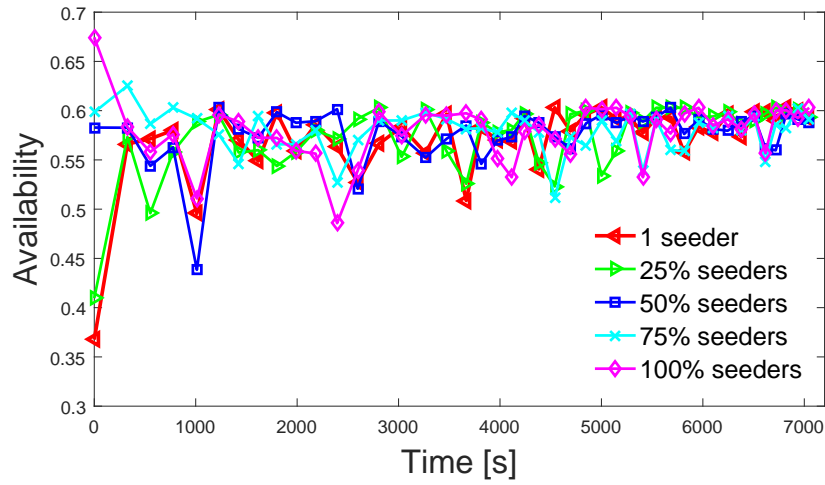
24

Figure 8: Availability over simulation time. Several seeding strategies in the city center of Luxembourg City at high-density traffic ($7AM - 9AM$). Anchor Zone radius 260 m, transmission radius 160 m. At $t = 0$ there are 94 nodes within the AZ. Mean over sample population of 20 (each point) every 300 m.
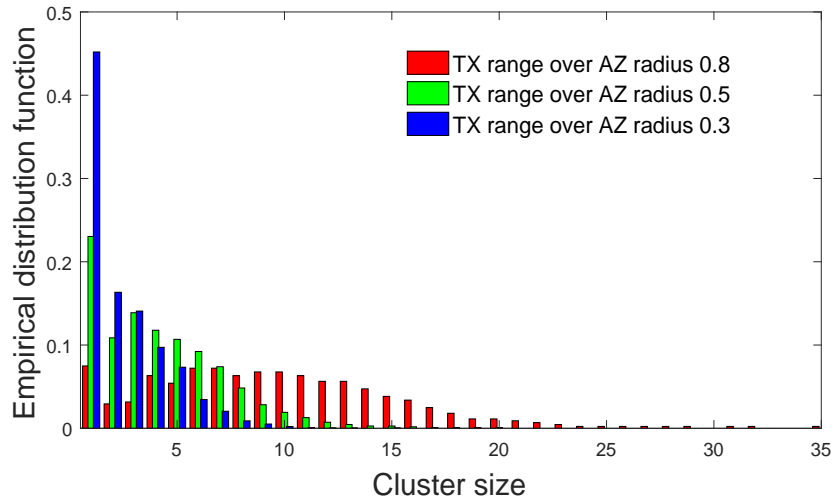

Figure 9: Empirical distribution function of cluster size versus r/R for the industrial AZ in the Luxembourg scenario, for the low-density traffic period.

timations of FC performance in scenarios with a significant amount of clustering. To clarify this aspect, we have considered, for a specific AZ (the industrial AZ in the Luxembourg scenario), the empirical distribution of cluster size. In order to be conser-

vative, we have considered the low-density traffic time interval $(2.30AM - 5.30AM)$, when clustering phenomena should be at their daily minimum for the considered area. Figure 9 shows the empirical distribution of cluster size for three values of the ratio r/R. We see that the amount of clustering is significant even at small values of transmission range, and that the cluster size (i.e., the number of vehicles in the cluster), as well as the relative number of clusters, increase with transmission range, as expected. This observation, together with the high model accuracy at every value of transmission range observed in Figure 6, suggests that our approach is robust against very high levels of node clustering.

In order to get insights on these performance features of FC, we analyzed the evolution over time of the mean content availability for each of the three AZs in the Luxembourg scenario. In Figure 10 we plot (in red) the mean content availability over time for 30 simulation runs.

Content availability, being a property of the system in steady state, is not defined when content does not float. In experimental scenarios, the content is ultimately subject to disappearance, possibly due to stochastic fluctuations in the populations of nodes, or to specific mobility patterns, which bring out of the AZ those nodes with content, or which prevent content from being exchanged. In Figure 10, we plot (in gray) the fraction of nodes with content over simulation time, for each of the 30 simulation runs. As we can see, in some configurations (e.g., in the city center AZ, for $R = 260$ m in the low traffic time interval), content does not float for the whole duration of the interval. For this reason, we also plotted (in black in Figure 10) the mean fraction of nodes possessing a copy of the content at a given time from content generation. This last quantity, taking into account content which disappears, is a better measure of the effective likelihood for a node inside the AZ to possess a copy of the content at a given time from content generation.

In all configurations, the criticality condition (3) is satisfied. However, in the low traffic period with $R = 260$ m, only in the industrial district the content floats (in most of the cases) for the whole duration of the interval (Figure 10a). In the city center AZ, instead, (Figure 10g), despite achieving rapidly a high availability (0.7 within the first 10 minutes), on average content floats for 30 minutes, while after an hour content

26

(a) Zone I, $R = 260$ m, low traffic

(b) Zone I, $R = 1\ km$, low traffic

(c) Zone I, $R = 260$ m, peak traffic

(d) Zone R, $R = 260$ m, peak traffic

(e) Zone R, $R = 1\ km$, low traffic

(g) Zone C, $R = 260$ m, low traffic

(h) Zone C, $R = 1\ km$, low traffic
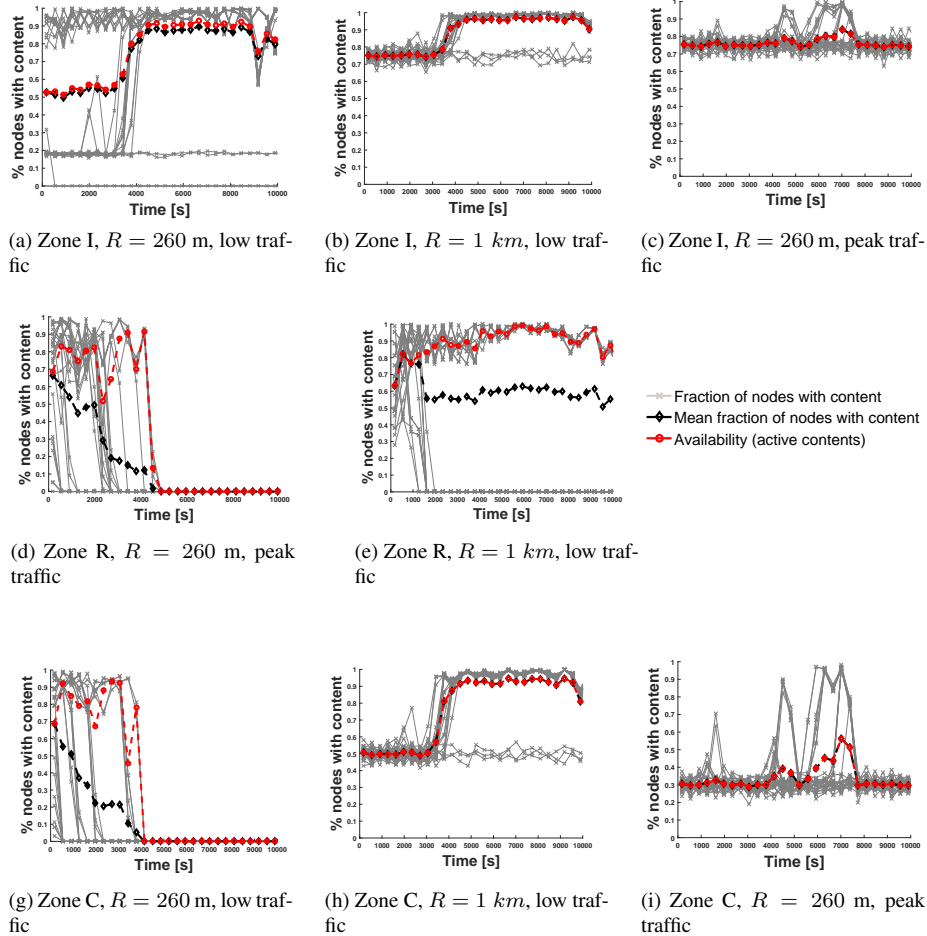
(i) Zone C, $R = 260$ m, peak traffic

Figure 10: Percentage of nodes with content, mean fraction of nodes with content, and mean availability over floating time for the three AZs areas in the Luxembourg scenario. In each run a different seeder node has been chosen. The low traffic time interval is $2.30AM - 5.30AM$, while the peak traffic time interval is $7AM - 9AM$. Transmission radius is 100 m.

has almost invariably disappeared. Similar behavior is observed in the residential area, where the content floats for at most 10 minutes (not shown).

Summarizing, in low traffic periods, even with relatively large ratios r/R, content floats for a few minutes at most, making FC suitable only for "near real time" applications, with short lifetime. In the residential zone, at all times of the day, a large AZ radius is required to float the content with high probability for more than 30 minutes. In low
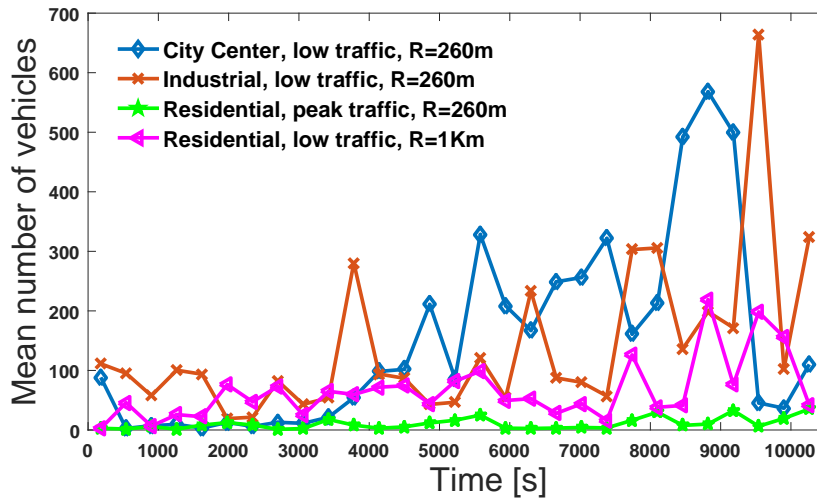
Figure 11: Mean number of nodes in the three locations of the Luxembourg Scenario over simulation time.

traffic, extending the AZ radius while keeping constant the transmission radius has an overall positive impact on availability, and on floating lifetime, in all parts of the city (Figures 10b, 10h and 10e).

When content floats for the whole two hours, we note a strong correlation between the evolution of availability of different content over time (Figure 10a, 10b, 10c and 10h). In addition, we see availability taking only a finite set of values over time. This is an indication of the existence of few clusters of vehicles, with little inter-cluster exchanges. Only the increase of the number of vehicles arriving in the AZ (at around 4000 s from the beginning of floating time, in the low traffic time interval, see Figure 11) increases the exchanges between these clusters (sharp increase in Figure 10a, 10b and 10h).

Finally, from these results, it clearly emerges how the criticality condition is a poor indicator of the actual feasibility of FC in realistic scenarios. However, they also show that, at least in the considered urban settings, by appropriately setting the AZ radius (within reasonable limits, i.e. not extending it to the whole city) it is always possible to set content to float for a consistent amount of time.

Figure 12: Residential scenario in Luxembourg, including a portion of a highway.

4.3. Limitations of our approach

When applying our approach in a realistic scenario, an important aspect is the choice of the AZ radius. A first element to consider in such a choice is the desired FC performance, usually in terms of success ratio. A reasonable choice for the AZ radius is typically the minimum value of AZ radius that ensures a success ratio at least equal to the target value. However, when such radius is small enough (say, of the same order of magnitude as the typical block size), then the spatial inhomogeneities in user distribution and differences in mobility patterns between various parts of the AZ may strongly affect the accuracy of the model, as the mobility data which parametrize our model, and which are computed as averages over the AZ area do not describe accurately anymore the actual features of mobility within the AZ. As an example, we consider the setup in Figure 12. It consists of an area of the city of Luxembourg, in the time interval (7 AM, 9 AM). As one can see, it is located in a residential area, but it also includes a small portion of a highway. The considered transmission radius is $r = 26$ m, and hence it is such that the highway vehicles have little or no chances of exchanging content with the residential nodes. In addition, highway vehicles spend a very small amount of time in the AZ. As a result, they hardly get the content, and when they get it, they often go out before being able to exchange it. Despite their modest contribution to content diffusion, highway vehicles, being numerous, give an important contribution to

29

the mobility parameters on which our model is based. The result is a high discrepancy between the success ratio predicted by our model (0.741) and the value arising from simulations (0.13).

5. Related Work

A scheme related to FC in VANETS is Geocasting (e.g., see [22] and references therein). Similarly to FC, it aims at delivering a message to a target region through opportunistic replications. However, typically the message source resides out of the target region. Hence Geocasting focuses more on getting content to that region, rather than (as done in FC) proposing strategies to make sure that, once in the region, it persists in the region over time, and it is delivered to the intended recipients with a given target probability.

The work in [12] considers a setup in which AZ size and location are based exclusively on application requirements (such as the region within which the content is of interest to users), and it proposes an algorithm which dynamically tunes content replication within the AZ, in order to ensure a target availability while trying to minimize the number of replications. Being tied to a fixed size AZ, such an approach can easily lead to infeasibility or to content disappearance from the AZ, as it is only able to decrease the number of content replications when they are more than those required to achieve the performance target.

In [9], the authors consider settings where users move according to the Random Direction mobility model (RDMM) without pauses. The analytical model they propose allows them to derive accurate estimates of availability and of success probability for various configurations of the AZ, as well as of the probability for a node to get the content as a function of the time spent in the AZ. Such approach for FC performance modeling is tightly dependent on the assumption of uniformity in space of user distribution, and on the geometrical properties of the AZ and of the mobility patterns. As such, it does not lend itself to generalizations to heterogeneous settings.

[23, 24] consider a campus setup, and propose an analytical model of FC based on a Poisson Jumps mobility model, which captures exchanges within user clusters, and

between clusters. The proposed model is based on the assumption that on-the-fly exchanges between nodes on the move are negligible, in a way that results are not easily generalizable to other contexts, such as vehicular ones, where such assumption does not hold. For the estimation of FC performance in vehicular environments, [11] develops an analytical model for the mean floating lifetime on a two-lanes highway. [25] considers a setup where a Software Defined Networking (SDN) controller collects mobility information, such as vehicle speed and position. It proposes a general strategy for optimizing the AZ radius for vehicular applications in urban environments. However, so far such strategy could not rely on accurate models of FC performance in vehicular settings.

An earlier version of some of the results in the present paper has been proposed in [26]. With respect to that paper, in the present work, we introduced the following variations. First, some of the approximations on which the derivation of a formula for success probability were based in [26] have been removed, by deriving a new formula for success ratio. Such approximations greatly reduced the accuracy of the model in scenarios with a small average number of stops during a sojourn time. Second, the model has been enhanced to include the effects of a non-negligible content transfer time, of limited contact time, of finite channel capacity, and of non-negligible connection setup time. Indeed, the latter has been shown experimentally to be key for the practical feasibility of a given wireless technology for FC. These new results allow more realistic and accurate estimations of FC performance in urban vehicular settings. All these new results have been validated by simulation. In order to appreciate the impact of the enhancements presented in this work with respect to the state of the art, we have considered the three areas of the Luxembourg scenario during the peak traffic period (7AM-9AM). In these setups, we have compared the success rate arising from simulations with the success probability derived from the approach presented in this paper. In addition, we have considered the success probability obtained from the only two other results which could be applied to a realistic, non-homogeneous setup, that is the approach in [26] and the one in [9]. As the latter has been derived under the assumption of nodes moving according to the Random Direction (RD) mobility model, we parametrized it by substituting the PDF of the number of contacts during the sojourn

31

Table 3: Accuracy of the approach presented in this paper and of the two main existing analytical approaches, in terms of % difference in success probability with respect to simulation values, for three different scenarios in Luxembourg City in a high-density traffic period ($7AM - 9AM$). Content size: $4\ MB$, connection time $\tau = 5\ s$, bandwidth $4\ Mhz$, coefficient attenuation 3, SNR 5 dB. Anchor zone radius 260 m, transmission radius 160 m.

Approach	Mobility	Success probability, % difference wrt simulation		
		City Center	Industrial	Residential
This paper	DMM	-2.18	-1.45	-2.34
[26]	DMM	18.63	18.04	18.53
[9]	Random Direction	29.69	20.22	28.10

time with a Poisson Distribution with mean equal to the experimental mean value. As it can be seen from Table 3, in all of the three scenarios the accuracy of the approach presented in this paper is substantially better than any other existing result.

6. Conclusions

In this work, we presented and evaluated an analytical model for the prediction of floating content performance in vehicular urban environments. The model does not require a detailed description of the road grid geometry, but only a few key parameters of mobility, still proving to be quite accurate. Indeed, the validation of analytical predictions against simulations, based both on synthetic mobility patterns and on real-world vehicular traces, showed very good matches. Hence, our work provides the first accurate tool to engineer a vehicular application based on FC in a realistic setting.

In addition, through simulation and analysis, we have evaluated the feasibility of the FC paradigm in realistic urban setups under a variety of system parameter values, traffic conditions, and mobility patterns. Our results suggest that, in European city contexts, in any city district and at any time of the day it is always possible to find a reasonable size of the AZ such that the content floats for the whole target duration with very high probability. As for application performance, our results indicate that in industrial and city center districts the FC paradigm is able to satisfy tight application-level performance requirements at any time of the day. Conversely, in residential districts and in areas with low density of inhabitants, FC is best suited for the diffusion of information with relatively short time of validity, or for those applications (such as proactive

mechanisms for traffic congestion avoidance) for which reaching less than 50% of the total amount of nodes within the area of interest does not compromise application-level performance.

7. Acknowledgments

References

[1] W. Mohr, 5G empowering vertical industries, 5G Infrastructure Association white paper, 2016.

[2] J. Gozálvez, M. Sepulcre, R. Bauza, IEEE 802.11 p vehicle to infrastructure communications in urban environments, IEEE Communications Magazine, 2012.

[3] Standard, J3016, SAE international taxonomy and definitions for terms related to on-road motor vehicle automated driving systems, 2014.

[4] TS 123 203 - V12.6.0 - Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Policy and charging control architecture (3GPP TS 23.203 version 12.6.0 Release 12), 2014.

[5] 3GPP, TS # 22.261, Service requirements for next generation new services and markets, 3GPP Technical Report, 2016.

[6] E. Hyytiä, J. Virtamo, P. Lassila, J. Kangasharju, J. Ott, When does content float? Characterizing availability of anchored information in opportunistic content sharing, IEEE Conference on Computer Communication (INFOCOM), 2011.

[7] A. A. V. Castro, G. D. M. Serugendo, D. Konstantas, Hovering information–self-organizing information that finds its own storage, Springer Autonomic Communication, 2009.

[8] M. Desta, E. Hyytia, J. Ott, J. Kangasharju, Characterizing content sharing properties for mobile users in open city squares, 10th Annual Conference on Wireless On-demand Network Systems and Services (WONS), 2013.

[9] S. Ali, G. Rizzo, B. Rengarajan, M. A. Marsan, A simple approximate analysis of floating content for context-aware applications, Proceedings of the fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC), 2013.

[10] S. Ali, G. Rizzo, M. A. Marsan, V. Mancuso, Impact of mobility on the performance of context-aware applications using floating content, Lecture Notes of the Institute for Computer Sciences, Context-Aware Systems and Applications, Springer International Publishing, 2014.

[11] K. Nakano, K. Miyakita, Information floating on a road with different traffic volumes between opposite lanes, Journal of Advanced Simulation in Science and Engineering, Japan Society for Simulation Technology, 2016.

[12] E. Koukoumidis, L.-S. Peh, M. Martonosi, Regres: Adaptively maintaining a target density of regional services in opportunistic vehicular networks, Pervasive Computing and Communications (PerCom), IEEE International Conference, 2011.

[13] E. N. Gilbert, Random Plane Networks, Journal of the Society for Industrial and Applied Mathematics, 1961.

[14] S. Ali, G. Rizzo, V. Mancuso, V. Cozzolino, M. Ajmone Marsan, Experimenting with floating content in an office setting, IEEE Communications Magazine, 2014.

[15] A. Miloslavov, M. Veeraraghavan, An integrated vehicular-wireless evaluation of wave/dsrc connected vehicle probe data service, IEEE Computing, Communications and Applications Conference (ComComAp), 2012.

[16] C. Sommer, R. German, F. Dressler, Bidirectionally coupled network and road traffic simulation for improved IVC analysis, IEEE Transactions on Mobile Computing, 2011.

[17] A. Varga, R. Hornig, An overview of the OMNeT++ simulation environment, ICST Simutools, 2008.

[18] D. Krajzewicz, G. Hertkorn, C. Rössel, P. Wagner, SUMO (Simulation of Urban MObility), an open-source traffic simulation, Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM2002), 2002.

[19] L. Codeca, R. Frank, T. Engel, Luxembourg SUMO traffic (LuST) scenario: 24 hours of mobility for vehicular networking research, IEEE Vehicular Networking Conference (VNC), 2015.

[20] TAPAS Cologne Project, "http://sumo.dlr.de/wiki/Data/Scenarios/TAPASCologne" (2017).

[21] OpenStreetMap, https://www.openstreetmap.org (2017).

[22] R. Jiang, Y. Zhu, T. He, Y. Liu, L. M. Ni, Exploiting trajectory-based coverage for geocast in vehicular networks, IEEE Transactions on Parallel and Distributed Systems 25 (12) (2014) 3177–3189.

[23] S. Ali, G. Rizzo, V. Mancuso, M. A. Marsan, Persistence and availability of floating content in a campus environment, IEEE Conference on Computer Communication (INFOCOM), 2015.

[24] G. A. Rizzo, V. Mancuso, S. Ali, M. A. Marsan, Stop and forward: Opportunistic local information sharing under walking mobility, Ad Hoc Networks, Elsevier, 2018.

[25] A. Di Maio, R. Soua, M. R. Palattella, T. Engel, G. A. Rizzo, A centralized approach for setting floating content parameters in vanets, Consumer Communications & Networking Conference (CCNC), 14th IEEE Annual, 2017.

[26] G. Manzo, M. A. Marsan, G. Rizzo, Performance modeling of vehicular floating content in urban settings, IEEE Teletraffic Congress (ITC 29), 2017 29th International, 2017.

[27] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Performance analysis of mobility-assisted routing, MobiHoc, 2006.

[28] G. F. Newell, The M/G/∞ queue, SIAM Journal on Applied Mathematics, Society for Industrial and Applied Mathematics, 1966.

[29] B. Christian, H. Hannes, P. C. Xavier, Stochastic properties of the random waypoint mobility model: Epoch time, direction, distribution and cell change rate, ACM/Kluwer Wireless Networks, Special Issue on Modeling and Analysis of Mobile Networks, 2003.

Appendix A. Frequency of contacts

Proof. Let us consider two nodes a and b within the region, and the case in which $p = 1$, i.e. nodes never exit the region. Let us assume that node a moves according to the district mobility model, while node b is static. We consider the location of b as well as the initial location of a to be random within the region.

Let us divide a time of an epoch of node a into intervals of same duration τ, with τ much smaller than the expected value of T_m and of T_s, and small enough to assume that, in each interval, a is either moving or static. The mean fraction of intervals in which a moves is hence q. q is also the probability that a is moving during an interval of duration τ. Let us compute now the mean number of contacts which node a experiments in a time interval $[0, \tau]$ during which a is moving. Note that since a moves along a straight line during a moving time, it cannot meet b more than once. Hence, the expected number of contacts between the two nodes is equal to 1 which multiplies the probability of finding b in the surface swept by a during the given time interval. As the location of b is uniformly distributed within the region, such probability is the ratio between the area of the surface swept by a, and the total area A.

Note that we assume an event of contact occurs at the first time instant in which two nodes are in range of each other. Hence in the computation of the contacts occurred in $[0, \tau]$, if node b is already in range of a at $t = 0$ we do not consider it as a contact occurred in the given time interval. As a consequence, for the computation of the above

probability we have to subtract from the total area swept by a in the time interval, the area covered by a at $t = 0$. The resulting expression of the probability p_c that the two nodes come in contact during $[0, \tau]$, when a moves, is

$$p_c = \frac{2rv\tau}{A} \tag{A.1}$$

In every interval, a moves with probability q. Hence, the mean number of contacts during $[0, \tau]$ is $p_c q$, and the contact rate is $p_c q / \tau$. Finally, the mean time between contacts is the inverse of the mean contact rate.

Let's now consider the case in which a and b both move without ever exiting the considered area. In each interval τ, the two nodes both move with probability q^2. For computing the mean time between contacts, we consider the equivalent setup in which a moves at a relative speed v_r while b is still. Due to the uniform choice of waypoint at every epoch, the expected value of v_r is given by

$$E_{v_r} = \frac{v}{2\pi} \int_0^{2\pi} \sqrt{(1 + \cos\theta)^2 + \sin^2(\theta)} d\theta \tag{A.2}$$

which is equal to $1.27v$. With probability $2(1-q)q$ only one of the two nodes is moving (with speed v), and both do not move with probability $(1 - q)^2$. Therefore, the mean node speed during a generic time interval is $1.27vq^2 + 2(1-q)qv$, and the mean contact time between the two nodes is

$$\frac{A}{2rqv(1.27q + 2(1 - q))}$$

The inverse of this quantity is the frequency of contacts between any pair of nodes in our system, and we denote it with ν. If there are N nodes in the area, the overall contact rate is hence $\nu N(N - 1)/2$.

When $p < 1$, in equilibrium the mean number of nodes does not vary. That means, when our system is in equilibrium state, ν is also the mean contact rate per couple of nodes. $\square$

Appendix B. Node density

Proof. We write the balance equations for the mean number of nodes with (and without) content. The mean number of nodes with content is the result of nodes with content going out from AZ, and of nodes without content coming in contact with nodes with content, and successfully exchanging it. If T_{soj} is the mean sojourn time in the AZ, and if we assume that nodes with content are uniformly distributed within the AZ, $n(t)/T_{soj}$ is the rate at which, at time t, nodes with content decide to get out of the AZ. The rate at which nodes acquire content is determined by the frequency with which a node with content comes in contact with a node without it, within the AZ. The mean rate of contacts within the AZ is given by ν (the mean rate at which a given pair of nodes comes in contact) multiplied by the number of node pairs within the AZ at t, $\frac{N(t)(N(t)-1)}{2}$. Of these contacts, those which increase $n(t)$ are those in which only one node of the two has the content, and for which the content is transferred successfully. As $p_n(t) = \frac{n(t)}{N(t)}$ the fraction of nodes with content in the AZ, and as Q is the probability that a content transfer is successful, the mean rate at which $n(t)$ increases at t is given by

$$2p_n(t)(1 - p_n(t))Q\frac{N(t)(N(t)-1)}{2}\nu \tag{B.1}$$

Since $N(t) \gg 1$, we can approximate (B.1) with $\nu n(t)m(t)Q$. Hence we have

$$\nu \bar{n}\bar{m}Q - \frac{\bar{n}}{T_{soj}} = 0 \tag{B.2}$$

Over time, m increases due to arrivals of new nodes into the AZ, and it decreases because of nodes without content leaving the AZ, and as nodes without content come in range of nodes with content. Hence

$$\lambda - \nu \bar{n}\bar{m}Q - \frac{\bar{m}}{T_{soj}} = 0 \tag{B.3}$$

By Little's law and balance equations we get (4). □

38

Appendix C. Success probability

Proof. We start by analyzing the probability for one or more successful content exchanges during a single epoch in the sojourn time of a node in the AZ. In the stationary regime, for the given mobility model, we assume that content is uniformly distributed among the population of nodes. This implies that at a given time instant the probability for a node to have the content is the same for all nodes, it is independent of node position within the AZ, and of the relative position of nodes. Such an assumption holds when clusters of more than two nodes are rare, which is typically the case when node densities are not high and $R \gg r$ [27].

We consider an epoch in the sojourn time of a node, and we evaluate the probability for a node to acquire the content during the moving time of that epoch. At time t, let $N(t)$ be the number of nodes in the AZ. Hence, there are $N(t) - 1$ node pairs of which a given node is part. The mean contact rate at t is $\nu(N(t) - 1) \approx \nu N(t)$. It can be easily seen that the system can be modeled as a $M/G/\infty$ queue, whose distribution of number of customers in the queue is Poisson [28]. Hence, in stationary state, $N(t)$ is Poisson with intensity $\overline{N}$. Therefore the expected contact rate for a single node at t (where the expectation is with respect to the contact rate between a pair of nodes) is also Poisson distributed, with intensity $\nu\overline{N}$. Note that this holds because the two processes (contact rate for a pair of nodes, and number of nodes in the AZ) are independent.

Thanks to the uniform content distribution, the probability for a node to have the content is $\frac{\overline{n}}{N}$. The probability for a node of having an unsuccessful contact (i.e., to get in contact with a node without content, or to fail in transferring a piece of content) is therefore $\frac{\overline{m}}{N} + \frac{\overline{n}}{N}(1 - Q) = 1 - \frac{\overline{n}}{N}Q$. The probability for a moving node to get the content during T_m, conditioned to having j contacts is the complement of the probability that j contacts bring to no transfer of content. That is,

$$\mathbb{P}[succ, m \mid T_m, j] = 1 - \left(1 - \frac{\overline{n}}{N}Q\right)^j \tag{C.1}$$

By the law of total probability,

$$\mathbb{P}[succ, m \mid T_m] = \sum_{j=1}^{\infty} \mathbb{P}[succ \mid T_m, j] Poiss_{\overline{N}\nu T_m}(j) = 1 - \exp\left(-\overline{N}\nu T_m Q \frac{\bar{n}}{\overline{N}}\right)$$

For the derivation of the pdf of the moving time f_{T_m}, note that T_m has a different distribution when it refers to the first moving time of a node in an AZ, when it refers to the last moving time in the AZ, and when the moving time is the first and the last of the node path inside the AZ (i.e., the node does not stop inside the AZ), and in all the other cases. However, when the mean number of waypoints on the path of a node is large enough, we can assume all these distribution to be the same as in those cases in which the node is neither entering nor exiting the AZ. The distribution of the transition length L is given by [29]

$$f_L(l) = \frac{4l}{\pi R^2}\left(\arccos\frac{l}{2R} - \frac{l}{2R}\sqrt{1 - \left(\frac{l}{2R}\right)^2}\right)$$

Then, given that $T_m = \frac{L}{v}$, $f_{T_{m,d}}(\tau) = v f_L(\tau v)$. By the law of total probability, the probability of success during a moving time is hence

$$P_m = \int_0^{\frac{2R}{v}} \left(1 - e^{-\nu\tau\bar{n}Q}\right) f_{T_m}(\tau)d\tau$$

The probability of success during a transit time P_t is derived in the same way as P_m. The pdf of the transit time T_t is derived from the pdf of the chord length c for a circle of radius R. As this last pdf is given by

$$\frac{c}{2R\sqrt{4R^2 - c^2}}$$

and given that $T_t = \frac{c}{v}$, we get

$$f_{T_t}(\tau) = \frac{\tau v^2}{2R\sqrt{4R^2 - \tau^2 v^2}}$$

In a similar manner, the probability for a static node to get the content during T_s is

$$P_s = \int_{\tau=0}^{+\infty} \left(1 - e^{-\nu\tau\bar{n}Q}\right) f_{T_s}(\tau)d\tau \quad . \tag{C.2}$$

Finally,

$$P_e = P_m + (1 - P_m)P_s \tag{C.3}$$

The probability of getting the content during the sojourn time in the AZ, conditioned to no stops in the AZ is P_t. The probability of no stops is $1 - p$. For $k > 0$ stops in the AZ, the conditional probability is $[1 - (1 - P_e)^k(1 - P_m)]$, and the probability of such event is $p^k(1 - p)$.

The probability of getting the content during the sojourn time in the AZ is hence given by

$$P_{succ} = (1 - p)P_t + \sum_{k=1}^{\infty} p^k[1 - (1 - P_m)(1 - P_e)^k] \tag{C.4}$$

which gives (5). □

Appendix D. Probability of successful content transfer

Proof. When $r << R$ and when two nodes are in range, a change of state (i.e. from still to moving, and vice versa) is relatively rare. Hence we consider only the following three cases:

- both nodes move for the whole duration of contact time;

- one node is still, and the other moves for the whole duration of contact time;

- both nodes are still for the whole duration of contact time. This is equivalent to not counting as part of the contact time the time intervals when both nodes are in range, but one is moving (before and/or after stopping).

Let us consider first the case in which one node is still and the other moves. Let $t = 0$ be the time at which the moving node comes in range of the still node. The moving node will move along a secant line of a circle of radius r, centered at the

41

location of the still node, and h is the distance of the chord from the center of the circle.

At any instant t during which the moving node is in range of the still node, let $vt \leq 2\sqrt{r^2 - h^2}$ be the length of the path within the circle until time t, where $2\sqrt{r^2 - h^2}$ is the chord length. Then the distance of the moving node at time t from the still node can be derived from standard trigonometric results:

$$d(t, h) = \sqrt{h_0^2 + \left(\sqrt{r^2 - h_0^2} - vt\right)^2}$$

The total amount of (useful, i.e. not control) bits which can be transferred during the contact time, is the integral of the capacity of the link between the two nodes, over the contact time. If τ_0 is the time taken by two nodes to sense their proximity and to get ready for content exchange, such total amount of bits which can be transferred during a contact with trajectory distance h is

$$B(h) = \int_{\tau_0}^{\frac{2\sqrt{r^2 - h^2}}{v}} C\left(d\left(t, \frac{h}{r}\right)\right) dt \qquad \text{(D.1)}$$

A successful content transfer implies that $B(h) \geq L$. One can easily see that $B(h)$ is monotonic decreasing with h. Hence the maximum trajectory distance h_0 is the one satisfying (12). As for the statistics of h in the RMM, when nodes are not too close to the border of the AZ, the assumption that all directions are equally likely is a good approximation, and h (the trajectory distance of a moving node with respect to a still node, conditioned to the event of the two nodes coming in contact of each other) can be modeled as a uniform random variable in the interval $[0, r]$. Hence, the probability of successful content transfer conditioned to having one node moving and the other still, is equal to the CDF of h evaluated in $h_0(v)$.

The derivation for the case with two moving nodes is similar to the case with only one moving node, considering one node still and the other moving at the relative speed given by the composition of the two speeds. Let f_r be the pdf of the relative speed between the two nodes. In our case, in which all nodes have the same speed, f_r can be

obtained with simple geometrical arguments, and it is given by

$$f_r(v_r) = \frac{v_r}{v^2 \pi \sqrt{1 - (\frac{v_r}{v} - 1)^2}}$$

Now, we have

$$B(h) = \int_0^{2v} f_r(v_r) B(h|v_r) dv_r$$

Where $B(h|v_r)$ is given by (D.1) with $v = v_r$.

Finally, let us consider the case in which both nodes are still. Given the properties of the DMM, all positions of a node within the transmission range of the other are equally likely (excluding again effects due to the border of the AZ). In this case, the mean throughput between the two nodes is

$$C_0 = \int_0^r C(x) \frac{2\pi x dx}{\pi r^2}$$

If T_s is distributed as an exponential, thanks to the memoryless property of the exponential distribution, the time spent still by the home node, counted from the arrival of the second node, is still exponential with mean $1/\mu$. Hence the duration of the contact time is the minimum of the time spent still by the nodes. By the properties of the exponential distribution, such contact time is distributed as an exponential random variable with mean $\frac{1}{2\mu}$. If $1 - \exp(-\frac{1}{2\mu}\tau)$ is the CDF of such distribution, the probability of successful content transfer is $\exp -\frac{1}{2\mu}(\tau_0 + \frac{L}{C_0})$. Putting it all together, we have (11). $\qquad\square$