Author version 2018. Open access in IEEE Access

Digital Object Identifier 10.1109/ACCESS.2018.2846201

Evaluating and Enhancing Information Dissemination in Urban Areas of Interest using Opportunistic Networks

LEONARDO CHANCAY-GARCÍA[1], ENRIQUE HERNÁNDEZ-ORALLO[1],(Member, IEEE), PIETRO MANZONI[1],(Senior Member, IEEE), CARLOS T. CALAFATE[1], AND JUAN-CARLOS CANO[1]

[1]Computer Engineering Department. Universitat Politècnica de València, Valencia, Spain

Corresponding author: Enrique Hernández-Orallo (e-mail: ehernandez@disca.upv.es).

This work was partially supported by the Ministerio de Economía y Competitividad, Spain, under Grant TEC2014-52690-R, and the Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación del Ecuador (SENESCYT), Ecuador.'

ABSTRACT Opportunistic Networks can provide an alternative way to support the diffusion of information in special locations within a city, particularly in crowded spaces where current wireless technologies can exhibit congestion issues. The efficiency of this diffusion relies mainly on user mobility. In fact, mobility creates the opportunities for contacts and, therefore, for data forwarding. This paper is therefore mainly focused on evaluating the dissemination of information in urban scenarios with different crowd densities and renewal rates. Through observation, we obtained real data from a local subway station and a plaza. These data were used, in combination with a pedestrian mobility simulator, to generate people mobility traces. We evaluated the diffusion of messages in these scenarios using the Direct and the Epidemic protocols. Experimental results show that content diffusion is mainly affected by two factors: degree of mobility and message size. Although it is well known that increasing the node density increases the diffusion rate, we show that, when keeping node density fixed, higher renewal rates cause the delivery ratio to drop. Moreover, we found that the relation between message size and contact duration is also a key factor, demonstrating that large messages can lead to a very low overall performance. Finally, with the aim of increasing the diffusion effectiveness of large messages, we propose an improvement over the Epidemic protocol, named *EpidemicX2*, based on the fragmentation of the data to be sent. The results show that the delivery ratio is increased, and the average delivery time is reduced, with no substantial increase in terms of overhead.

INDEX TERMS Mobile Computing, Opportunistic Networks, Network Performance Evaluation

I. INTRODUCTION

The objective of information dissemination is to circulate information and alerts through a communication infrastructure in a cost-effective and timely manner. Nevertheless, the utilisation of current wireless technologies (e.g., WiFi or 4G) in crowded spaces, such as pedestrian city squares, train stations, shopping malls, etc. can be seriously affected by propagation and congestion issues. A good solution for providing information diffusion in these areas could be the use of *Opportunistic Networks*. Opportunistic Networks (OppNets for short) [1], [2] are based on the possibility of exchanging messages between nearby devices when establishing some type of direct and localised communication link (e.g., through a Bluetooth or a WiFi direct channel).

This paper is mainly focused on evaluating the efficiency of information diffusion using OppNets in different realistic scenarios, and for different people renewal rates. More specifically, we consider a plaza and a subway station, which are typical crowded scenarios with a high degree of people renewal. This "renewal" process, as we will show, highly impacts the efficiency of message delivery but, to the best of our knowledge, most of the existing protocols were designed and optimised considering scenarios with a constant number of permanent users.

The diffusion of information using OppNets depends mainly on node mobility and the diffusion algorithm adopted. These factors affect how messages are exchanged between nodes when a contact occurs, being the main goal to optimise their diffusion. Clearly, the mobility of nodes is tightly coupled with human behaviour, as these mobile devices can communicate only when users come into contact. Thus, the performance evaluation of OppNets solutions in realistic scenarios must consider and combine both the technical aspects (related to protocol behaviour) and the human mobility characteristics.

The evaluation of these scenarios is a challenging problem due to the current methodologies used for evaluating OppNets [3]. Commonly, evaluating OppNets combines the use of a network simulation tool with realistic mobility traces. Despite the availability of a large collection of traces obtained from the observation of node mobility in real scenarios, the results that can be obtained are specific to those scenarios. This methodology, although useful, clearly poses some issues when trying to extend the results obtained to other "similar" scenarios. In addition, these traces have a fixed set of nodes, and the renewal (if it exists) is very reduced. In order to avoid these restrictions, synthetic mobility models can be used to provide an adjustable generation of traces, based on reproducing some statistical properties of human mobility. Nonetheless, these models cannot grasp the actual mobility pattern of people in specific spatial scenarios, nor the temporal variation of the number of persons in a place.

In this paper, we propose a practical solution to this problem that is based on the use of a pedestrian mobility simulator for building realistic scenarios in order to generate people mobility traces. The utilisation of these mobility simulators was originally proposed in [4], for statistically studying the mobility along a street. A pedestrian simulator can model the microscopic (individual) and macroscopic (crowd) dynamics of pedestrian mobility, an option that is not supported by the synthetic mobility models. The combination of a pedestrian simulator with an OppNet simulator is an idea that was recently used in [5]. Following this idea, we have opted for the creation of external traces using the PedSim [6] pedestrian simulator, along with the use of the ONE simulator [7].

More specifically, we generated two scenarios in PedSim: a typical city plaza and a subway station, based on real measurements in order to replicate the real mobility behaviour of the pedestrians. The analysis of these generated scenarios, in both temporal and spatial dimensions, reflects their realism, and thus the feasibility of the approach. Using the generated traces, we evaluate the diffusion of information that is intended to be spread among the people inside the evaluated scenarios. For evaluating this diffusion, in the experiments we compare the *direct delivery* and the *epidemic* diffusion approaches.

The obtained results show that the performance of these protocols is clearly reduced as the message size increases, and that aspects like the number of nodes and their renewal rate drastically influence the dissemination of a message. Particularly, in scenarios where the number of nodes remains constant, increasing the people renewal rate reduces the diffusion of messages. This is especially relevant for large messages, whose distribution is very slow, leading to low delivery ratios. Based on these experiments, we propose an improvement over the standard *Epidemic* protocol (namely *EpidemicX2*), that consists of dividing large messages into smaller parts in order to increase their diffusion. The experiments show that *EpidemicX2* increases the delivery rate and reduces the average delivery delay. This improvement is evident in scenarios with high people renewal rates, and where we find that almost no messages are actually delivered when using the plain *Epidemic* protocol.

The rest of this paper is organised as follows: Section II presents some related work highlighting the relations with our own proposal, whereas section III describes the opportunistic information dissemination, the methodology for generating the scenarios, and the description of the scenarios used. In section IV we analyse the generated traces for the different scenarios, considering both temporal and spatial aspects. In order to improve the information diffusion, we describe in section V the proposed *EpidemicX2* protocol. Section VI presents the performance evaluation of the different scenarios, and, finally, section VII presents some conclusions and future research directions.

II. RELATED WORKS

Opportunistic networks are based on the opportunity of contacts between pairs of nodes as a way to propagate messages. The effectiveness of OppNets depends mainly on the number and duration of these contacts. The mobility and its impact on the performance of OppNets have been studied extensively, see for example [8]–[12]. The majority of these studies are solely focused on content diffusion among mobile devices. For example, Hang et al. [10] developed an analytic model to study epidemic content delivery, and the authors in [11] study the impact of node density on the data dissemination time using a synthetic model. Other proposals, such as [13], [14], study the message dissemination behaviour of the Epidemic protocol by focusing on the mobility patterns of the nodes, evaluating the relationship between factors such as mobility model, speed and node density, and locations of the nodes. In [15] the authors performed several experiments using the ONE simulator; in their results, we can observe how the message size and the routing protocol have an impact on network performance and communications.

The results of a real test-bed are described in [16]. The paper presents an experimental evaluation based on the mobility of visitors in an entertainment theme park in order to understand network requirements (minimum number and density of mobile devices, and supporting infrastructure nodes) for opportunistic communication. The results show that, in this scenario, the efficiency of the diffusion depends on user density and the distribution of contacts.

Meanwhile, some other papers are more focused on evaluating the impact of human behaviour in the opportunistic forwarding of messages [17], [18]. In general, the previous results show that information diffusion increases with the node density and the number of contacts, being mobility the main enabler of opportunistic data dissemination [19].

Various papers studied how to offload the wireless infrastructure through opportunistic communications. Hui et al. [20] evaluate how these hybrid networks can improve message delivery ratios. Specifically, they conclude that opportunistic communication improves the system capacity and delay, even with infrastructure networks with a high access point density. In order to improve connectivity in mobile networks, several papers have proposed adding fixed infrastructures such as base stations, relays or meshes, known as hybrid networks [21]. Another approach, such as placing a sparse set of well-connected base stations in an ad-hoc wireless network, has been studied in [22]. In [23], the authors analysed whether the use of autonomous agents can improve the performance of sparse mobile networks.

A related approach is the so-called Floating Content [24] paradigm. Floating content is a contact sharing approach where a message in a certain area is tagged with the geographical coordinates of that location, which is referred to as the anchor-zone of the message. This approach is a kind of best effort service, in which messages are locally generated, their availability is geographically limited, and their lifetime and diffusion depend on the mobility and resources of mobile nodes. Recently, this approach has been analytically evaluated in open city squares [25] using a custom mobility model, spatial analysis and Markov chains, and assuming that nodes may enter and leave the city square. Since the analytic model assumes too many simplifications, it is not clear whether the results are realistic.

Since mobility is a key factor for evaluating OppNets, several models have been devised in order to represent it. From the basic models, such as *Random Walk* and *Random Waypoint* [26], to more realistic models that consider some social aspects of human movements like working days and meal hours. Examples of such models are: SWIM (Small Worlds In Motion) [27], which is based on the assumption that users either select a location close to their home or a very popular location, SLAW (Self-similar Least Action Walk) [28], where the movement of people is expressed using gaps among fractal waypoints, and WDM (Working Day Movement) [29], which models the everyday life of average people that go to work in the morning, spend their day at work, and commute back to their homes in the evening. Nevertheless, these synthetic models can only capture some specific characteristics of human mobility. So, in order to have more realistic simulations, the best approach is to use real mobility traces [30], [31] combined with an OppNet simulator.

Given the limitation and availability of real traffic traces, some recent papers use pedestrian simulators for evaluating the mobility of pedestrians. In [4], the authors introduce a complex model for streets based on queues, along with contact and duration probabilities. This model is compared with simulation results using the commercial pedestrian simulator LEGION. Using this pedestrian simulator, the authors in [32] study the impact of mobility and the scenario used on opportunistic communication (inter-contact time and contact duration). The results show that, as expected, the type of scenario is the most important aspect to consider, and so a general model cannot be derived. Finally, [5] proposes a model for crowd-counting based on an application that receives messages from an AP (access point), and also by the detection of contacts between nodes. They propose a model based on SDE (Stochastic Differential Equations), and evaluate them using some mobility traces generated with LEGION and other mobility models.

Besides the mobility of the nodes, the performance of OppNets also depends on two important aspects: how messages are forwarded, and how they are locally managed (in the buffer nodes). The first aspect depends on the routing algorithm adopted. Regarding the buffer management, it has been recently shown [33] that it is important to implement certain mechanisms to improve buffer management, prioritising the forwarding and discarding of messages. Social aspects can also be considered in the management of local buffers and in the message forwarding strategy. In this context, the authors in [18] used theoretical analysis applied to social networks to classify and study some diffusion schemes based on the homophily (social networks phenomenon) by combining node relationships and their interests in the data. The authors in [34] propose a technique to decrease the use of resources. They use an algorithm called FSF (Friendship and Selfishness Forwarding Algorithm) that performs a validation taking into account aspects like whether the users work together, live nearby, or concur in some places, before forwarding a message.

To conclude, there are a lot of interesting research works about opportunistic networks, where the authors have evaluated their performance from different perspectives. In our case, using a novel tool called PedSim [6], we focused on generating traces with a behaviour much closer to the pedestrian mobility, defining realistic people mobility traces that allow the evaluation of different degrees of users densities and renewal rates.

III. SCENARIOS FOR EVALUATING OPPORTUNISTIC INFORMATION-DISSEMINATION

The scenarios considered in this paper are bounded places where people can enter, stay for a while, and eventually leave. We can think about a lot of real scenarios of this type, such as shopping malls, public buildings, stations, touristic places, and so on. In these delimited scenarios, the availability and performance of the communications technology used, such a 4G, can be very limited and, when the place is crowded, can be seriously affected by congestion and propagation issues.

In this paper we used two real scenarios located in the city of Valencia, Spain (see figure 1): a city square or *plaza*, namely "*Plaza de la Virgen*" that is a typical touristic pedestrian square in the downtown area of Valencia, Spain, and a subway station, known as "*Estación de Alameda*", that is a centrally located stop for the subway line having four tracks and three platforms.

In this class of scenarios, our goal is to provide a valid

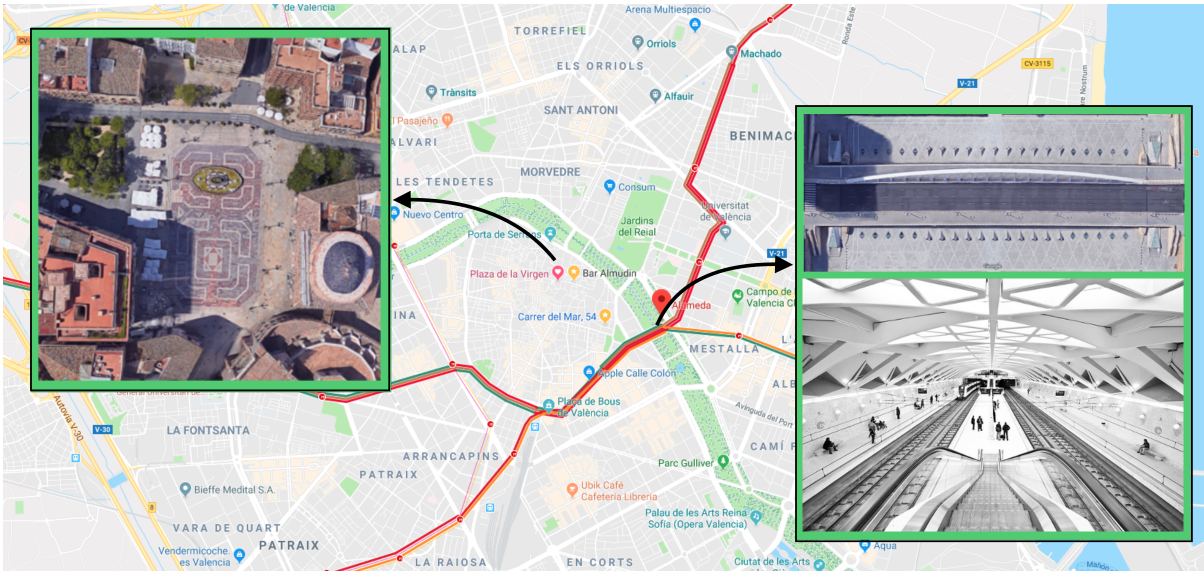


FIGURE 1: Location of the scenarios evaluated in the city map of Valencia, Spain. On the left, the "Plaza de la Virgen", a typical touristic pedestrian square. The dimensions of this plaza are roughly 120×120 meters. The pedestrian space is surrounded by several buildings with a fountain in the centre, and has seven entry points. On the right, the subway station, known as "Estación de Alameda". The station is a centrally located stop for trains which connect the various suburbs of Valencia; it has four tracks, three platforms, and four main entrances.

alternative for information diffusion using Opportunistic Networks. Specifically, we are considering the use of contact-based messaging applications that are based on establishing a direct short-range communication link between devices. We suppose that mobile devices have a messaging application that notifies and shows the user the received messages. We consider that new messages with information are generated by fixed nodes located in these places. This information can be generated periodically, or when relevant information is required to be sent. Note that no messages are sent or stored in servers; instead, all information is stored on the mobile devices in a given area.

Two main approaches can be considered for the diffusion of this information between mobile devices. The *direct delivery* protocol [8] can be considered the simplest way to spread a message. The fixed nodes deliver the message only to the nearby mobile nodes. Thus, the efficiency of the message diffusion will depend on the opportunity of having a direct contact between the fixed nodes and the mobile nodes. On the other hand, *flooding* protocols spread a message over the network. Mainly known as *Epidemic* diffusion protocol [35], it makes a copy of the message for all contacted nodes, in order to increase the possibility of spreading the message. Each node has a limited buffer where the messages in transit can be stored. The information is initially disseminated through *direct delivery* when users are in contact with the fixed nodes. Moreover, using *Epidemic* diffusion, two nodes establishing a pair-wise connection will exchange the messages they have in their buffers, and check whether some of the newly received messages are suitable for notification to the user. *Epidemic* diffusion achieves a good delivery ratio of the message at the expense of an increased usage of local buffers and an increased number of transmissions.

In order to evaluate the diffusion of these messages, we first propose a novel methodology to generate realistic scenarios based on the use of a pedestrian mobility simulator. Moreover, we specifically focus on the evaluation of crowded spaces with people renewal, i.e., where users can either enter or leave the evaluated scenario, a problem that still remains mostly unexplored.

A. SCENARIO GENERATION AND EVALUATION PROCESS

This subsection describes how the selected scenarios were generated using the PedSim and the ONE simulator. We first detail the simulation setup, describing the process employed to generate and evaluate the different scenarios. The performance evaluation process, including tools and methods used in the experiments, is depicted in Figure 2. First, we use PedSim for generating a mobility trace for each of the evaluated scenarios. Then, the generated mobility trace is imported into the ONE, where we can evaluate the behaviour of the different diffusion protocols, generating different metrics.

PedSim [6] is an open source microscopic pedestrian crowd simulator. Using this mobility simulator, we created our selected scenarios, and defined the number of pedestrians, their movement type, and their destination. PedSim comprises two tools: a C++ library, that allows us to use pedestrian dynamics in our software, and a real-time visual-

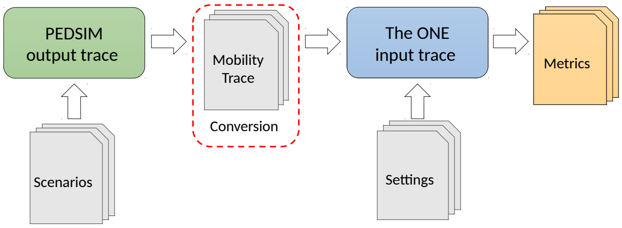


FIGURE 2: Scheme for the evaluation of mobility using real scenarios.

ization tool. To generate the scenario we must create a script defining the different physical elements in the environment, and the pedestrian behaviour. Several options can be configured regarding the mobility of pedestrians, the number of pedestrians in a particular area, the speed at which they move, the time when a group of pedestrians enters or leaves, etc. Scenarios can comprise walls, as well as fixed and mobile obstacles, which allows defining very realistic places.

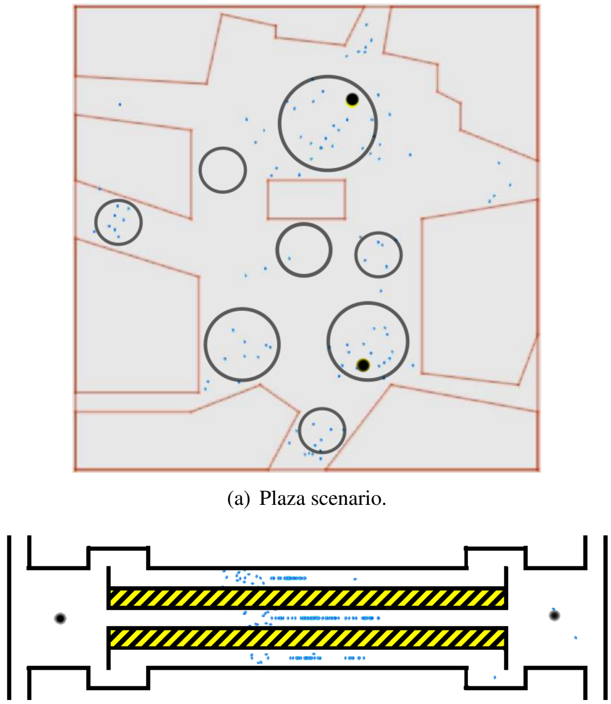
When the scenario is defined, PedSim simulates the movement of the pedestrian based on a generic coupled differential equation model, known as the *social force model*, developed by Helbing et al. [36]. This social force model is usually used in this kind of pedestrian simulators such as Legion and SUMO. The simulation generates the movement of the pedestrians in the defined allowed areas while avoiding obstacles, and it can be visualised using the real-time visualiser made available by the tool.

The output generated by PedSim can be used as an input to the ONE simulator after performing some re-formatting. The main problem is that the ONE requires, during the whole simulation, for all nodes to be placed in the scenario, thus not allowing nodes to enter or leave the scenario. To tackle this problem, all nodes are included in the trace from the beginning, but only those ones that are in the target area at a particular time are marked as *active* nodes, being allowed to communicate. After this conversion, the resulting traces can be processed by the ONE.

The ONE (Opportunistic Network Environment) [7] is a simulation tool specifically designed for evaluating OppNets. This simulator has been proved to be a trustworthy tool for evaluating OppNets [3], and it offers a wide variety of mobility models and routing protocols. The ONE allows generating node movements using different models, to reproduce message traffic, routing, and cache handling, and to visualise both mobility and message passing through its graphical user interface. It includes the main diffusion protocols such as *Epidemic*, PRoPHET, *Spray & Wait*, etc. It can also produce a variety of reports about node movements and message passing, as well as general statistics.

B. PLAZA SCENARIOS

We based our first set of scenarios in a real city square, that corresponds to the "*Plaza de la Virgen*" (Plaza for short) in the City of Valencia, as shown in figure 1. This is a typical touristic place that usually has a high degree of people



(a) Plaza scenario.

(b) Station scenerio.

FIGURE 3: PedSim generated scenarios. On the right figure we can see the PedSim generated scenario with several pedestrians entering and leaving the place (small blue points); the two big black dots are the fixed nodes. For the Plaza scenario we can also see the points of interest with circles, and the two big black dots are the access points.

renewal. The selected area has a dimension of 120 $\times$ 120 meters. From this real place, we defined in PedSim a physical area with the obstacles and open spaces where pedestrians are allowed to enter and leave, as shown in figure 3(a).

Based on this place we define four scenarios. All scenarios were generated for a simulation time of 1 hour, and the number of pedestrians remained the same: 100 persons. Nevertheless, the different scenarios differ in the renewal rate of pedestrians, from "no renewal" to "high renewal", namely:

1) **No Renewal,** the pedestrians remain within the target area for the entire simulation time. This scenario is created for evaluating the behaviour of the diffusion protocol when there is no renewal, so that we can compare the results with the following scenarios.
2) **Low Renewal,** where 50 pedestrians are replaced every 15 minutes (that is, 50 pedestrians leave the target area, and 50 new ones enter that area), so the final number of pedestrians that have been in the plaza will be 250.
3) **Medium Renewal,** where the renewal rate is increased to 50 pedestrians every 5 minutes, summing up a total of 650 pedestrians.
4) **High Renewal,** contemplates an extreme situation, where every minute 50 pedestrians are renewed; at the end of the simulation a total of 3050 different pedestri-

TABLE 1: Main parameters of the Plaza scenarios, where A_T is in seconds

Scenario	N_0	A_T	P_R	N
No Renewal	100	-	0	100
Low Renewal	100	900	50	250
Medium Renewal	100	300	50	650
High Renewal	100	60	50	3050

ans have visited the plaza.

The main parameters are summarised in table 1, where A_R is the average time between renewals, P_R the number of people that enter/exit in each renewal, and N the number of nodes (people) generated.

The renewal and movement of pedestrians were implemented in PedSim as follows: the Plaza scenario has seven entry/exit points, so that pedestrians will be randomly placed in any of the entries. The movement of pedestrians within the plaza follows the social force model, with an average speed in the range of 0.3-1.5 m/s, moving through the defined main points of interest, as shown in figure 3(a). These points are defined as the locations where the pedestrians go and stay (such as monuments, restaurants, etc.). Regarding the renovation of the pedestrians, at each renewal interval a given number of nodes are randomly selected from the pedestrians that are in the plaza. These nodes are then notified to leave the square using one of the exits. Then, 50 new nodes are created and randomly located in one of the entrances of the plaza. Once the four mobility traces are generated, they must be modified to be valid for the ONE simulator, as explained in the previous subsection.

C. STATION SCENARIOS

The second location selected for our experiments is a subway station. We chose it since it clearly represents a typical crowded place with a high degree of people renewal. We used a real station, namely "*Estación de Alameda*" (Station for short), in the city of Valencia (Spain) depicted in Figure 1. This station has four tracks and three platforms (two side platforms and one centre island platform), comprising an area of about 150 × 50 meters. The platforms can be reached via stairs that are accessed through four entrance doors at each the corner of the station. Four different lines pass through this station, with average train intervals ranging from 5 to 10 minutes.

In order to generate the scenario with PedSim, we first took real measures through direct observation of the people arriving and leaving the station at different times of the day. Particularly, the values obtained correspond to Thursday, 7 December 2017. We selected three different time periods, and for each of these periods, we measured the arrival time of the trains, the number of persons that get on and off these trains, and also the people arriving from the four entrance doors. In this case, compared to the Plaza scenario, the number of nodes inside the station is different in each period. Therefore, we ordered the three intervals from lower to higher number of nodes and renewal rates:

TABLE 2: Main parameters of the Station scenarios evaluated. Note that β_D and λ_D are in people/s, and A_T in seconds.

Scenario	β_D	δ_D	A_T	P_L	P_A	N_I	N_O
Low (19-20h)	0.09	0.04	132	5.14	11.43	320	144
Med (14-15h)	0.11	0.07	124	9.23	14.28	400	263
High (08-09h)	0.17	0.10	127	12.85	21.79	610	360

1) **Low (19-20h)**. This interval, obtained from 19:00 to 20:00 hours, is the time when most of the people arrive at the station to go back home, and it is the least crowded scenario.
2) **Medium (14-15h)**. The second interval (from 14:00 to 15:00) is characterized by a higher renewal rate.
3) **High (08-09h)**. Finally, this interval (from 08:00 to 09:00) corresponds to a rush hour, being the most crowded scenario, and thus having the highest renewal rate.

To characterise the three different time intervals, we define several metrics: the average arrival rate β_D and exit rate δ_D of the station (using the entrance doors), the average time between trains A_T, and the average number of people leaving and arriving on each train P_L, P_A. All these values are referred to the one-hour period evaluated. From these values, we can also obtain the actual number of people that arrived and left the station through the entrance doors ($N_I = \beta_D \cdot 3600$ and $N_O = \delta_D \cdot 3600$), as well as the number of people that have been in the station: $N = N_I + N_O$. The main parameters for these scenarios are shown in table 2

From this real scenario, we defined in PedSim a physical area with the obstacles and open spaces where people (pedestrians) can enter and leave, as shown in Figure 3(b). The generation and movement of pedestrians were implemented in PedSim as follows:

1) *People entering the station*. We specified how people enter the station through the main entrance doors and get to the platforms to wait for the train. Specifically, for each of the four main entrances, we generated new pedestrians according to a Poisson process with rate $\beta_D/4$, entering the station and passing the turnstiles. Then, each pedestrian is randomly directed (with equal probability) to one of the four platforms.
2) *Train arrivals*. The train arrival times are generated with the values obtained from the measurements. When a train arrives, the pedestrians waiting on the corresponding platform get on the train and *disappear* from the simulation (they leave the station on the train). At the same time, pedestrians get off the train and get into the platform. From this platform, and for each pedestrian, an output door of the station is randomly selected, so the pedestrian goes directly to it, leaving the station.

As in the previous scenarios, the movement of pedestrians follows the social force model with an average speed in the range of 0.3-1.5 m/s. Finally, for each interval, mobility traces are generated, and then these must be adapted to the

TABLE 3: Fixed simulation parameters.

Parameters	Value
Movement	Synthetic Subarea
Area	Defined by PedSim
Pedestrian Speed (m/s)	0.3 - 1.5
Interface	Bluetooth
Tx Range radio (m)	8
Tx Speed (Mbps)	2
Message Sizes	1KB to 6MB
Device Memory	1GB
TTL (min)	720 (default)
Simulation Time (seg)	3600

format accepted by the ONE simulator, as explained in the previous subsection.

D. OPPNET SIMULATION

Using the previously defined scenarios, we then evaluated the information dissemination in the Plaza and Station scenarios using OppNets. We assume that some kind of information (such as timetables, alerts, photos, videos, etc. from tourist information or from the subway company) is required to be transmitted to the pedestrians. This information is initially generated by two fixed nodes that are located near the two entry points, as shown in figure 3. In the Plaza scenario, they are located on opposite sides, near to the two attraction points. In the case of the Station scenario, they are located at the turnstiles areas (one in each edge of the station), that is considered the most convenient place, as all people must pass through the turnstiles. This way, nearby nodes can directly receive this up-to-date information, and, in case of *Epidemic* diffusion, spread this information to the rest of nodes in the place following an epidemic approach. Summing up, the fixed nodes are located in the positions with the highest opportunity of contacting with mobile nodes.

These fixed nodes are the origin of the information message. This message size is important for evaluating the performance; we therefore considered different messages sizes: from a small message (1KB) up to big messages (3-6MB) that could contain, for example, a short video. Messages are transmitted using Bluetooth, with a defined maximum range of 8 meters, and an average bandwidth of 2Mbps. Table 3 summarises the simulation parameters, including the fixed parameters that comprise both the defined scenario (already defined in the PedSim), such as area, node speed and simulation time; and the OppNets parameters, such as communication parameters and buffer size.

IV. TRACE MOBILITY CHARACTERISATION

Since Opportunistic Networks performance relies on the opportunity of contacts, it is important to characterise the structure of these contacts and, in general, the mobility of the nodes. Thus, we analysed the generated traces considering both temporal and spatial aspects.

A. TIME EVOLUTION OF NODES

In this section, we are going to model and characterise the number of nodes in the different scenarios. First, we consider the Plaza scenario using the main parameters detailed in table 1. We can obtain the accumulated number of people arriving at the plaza, considering also an initial number of pedestrians N_0, as:

$$N_A(t) = N_0 + P_R\lfloor t/A_R \rfloor \quad (1)$$

and the accumulated number of people exiting the plaza as:

$$N_E(t) = P_R\lfloor t/A_R \rfloor \quad (2)$$

Note that, by using the *floor* function ($\lfloor \, \rfloor$), we consider that the renewal takes place at the end of the interval. Summing up, the number of people that remains in the plaza at time t is $N_I(t) = N_A(t) - N_E(t)$, and, as expected, is N_0, and the whole number of nodes after t seconds is $N(t) = N_E(t) + N_I(t) = N_E(t) + N_0$.

Now, we evaluate the station scenarios. We can make the following simplifications for modelling the arrival and exiting of pedestrians:

- Station arrival rate: People entering the station by the access doors following a Poisson distribution with rate β_D (in each door we have $\beta_D/4$)
- Train arrival: In this case, people arrive when the train gets into the station (that is, it is a burst process). We consider an average number of people getting off for each train (arriving at the station) P_A, and that the time interval between consecutive trains is A_T.
- Train departure: People that are on the platforms leave the station by getting on the train, that is, P_L people with time interval A_T.
- Station exit rate: The people that get off the trains leave the station by one of the access doors, after τ seconds, which is the average exit time.

From these values, we can obtain the accumulated number of people arriving at the station, also considering an initial number of pedestrians N_0, as:

$$N_A(t) = N_0 + \beta_D t + P_A\lfloor t/A_T \rfloor \quad (3)$$

and the accumulated number of people exiting the station as:

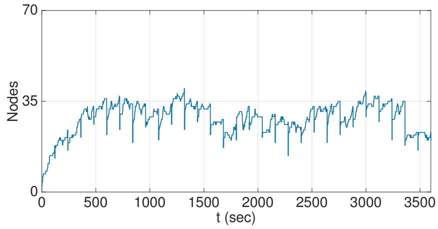
$$N_E(t) = P_A\lfloor (t+\tau)/A_T \rfloor + P_L\lfloor t/A_T \rfloor \quad (4)$$

Summing up, the number of people that remains in the station at time t is $N_A(t) - N_E(t)$, and working out we have:

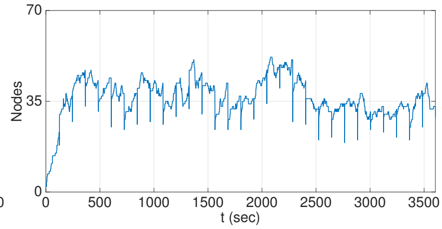
$$N(t) = N_0 + \beta_D t + (P_A - P_L)\lfloor t/A_T \rfloor - P_A\lfloor (t-\tau)^+/A_T \rfloor \quad (5)$$

Note that, in these equations, we consider that trains arrive at the end of the A_T intervals. Therefore, for time 0, we assume that a train has just passed, and so some people that are initially at the station will exit at time τ; the term $(t-\tau)^+$, when $t-\tau$ is negative, becomes 0. Using equation 5, we can obtain the average number of nodes that are in the station up to time T as:

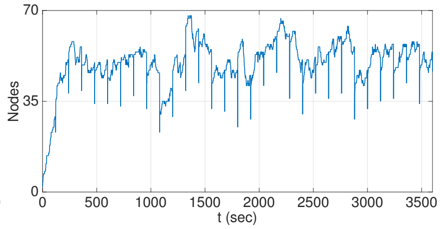
$$\bar{N}(T) = \frac{1}{T}\int_0^T N(t)dt \quad (6)$$



(a) Low (19-20h)



(b) Medium (14-15h)



(c) High (08-09h)

FIGURE 4: Number of nodes along time for the Station scenarios.

TABLE 4: Number of contacts and their average duration for the different hour ranges in the Plaza scenarios.

Scenario	Number of Contacts	Average Duration (s)
No renewal	67742	12.29
Low renewal	64333	11.64
Medium renewal	65985	11.16
High renewal	62084	9.89

TABLE 5: Number of contacts and their average duration in three time ranges for the Station scenarios.

Scenario	Number of Contacts	Average Duration (sec.)
Low (19-20h)	2912	53.76
Medium (14-15h)	5618	46.50
High (08-09h)	11002	39.72

Using this expression, with $\tau = 180s$, and the characterisation values from table 2, the average number of nodes for the different scenarios are the following: Low (19-20h): 17.7s; Medium (14-15h): 29.1s, and High (08-09h): 40.9s.

Finally, figure 4 shows the number of nodes depending on time for the different time ranges, obtained from the mobility traces. We can see the variation of the number of nodes depending on the train arrivals. The pattern is irregular, depending mainly on the measured values of the people arriving and leaving by train. From these mobility traces, we also calculated the average number of nodes for the scenarios obtaining the following values: Low (19-20h): 28.5s; Medium (14-15h): 36.5s, and High (08-09h): 46.1s. The difference of these values from the ones obtained from the previous analytic model reflects the effect of real measurement variations, where the analytic model works solely with the average values.

B. CONTACTS CHARACTERISATION

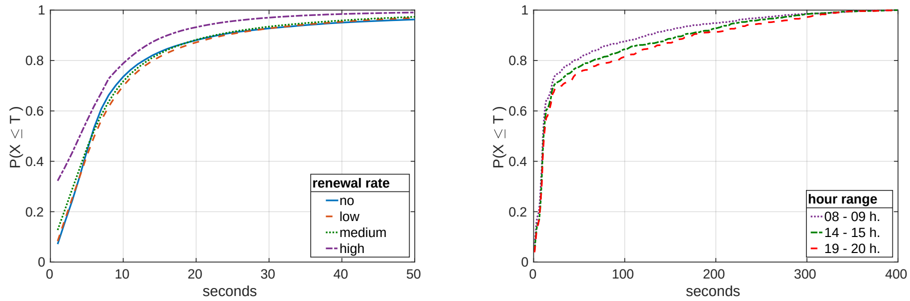
Now, we focus our evaluation on characterising the contacts. These contacts were obtained using the ONE simulator, considering a granularity of 1s. So, given the position of the nodes, the simulator determines, at every second, which nodes are in range or not, thereby allowing us to obtain the whole number of contacts and their duration, in addition to their location. The main figures, number of contacts, and average duration of the different renewal scenarios, are summarised in tables 4 and 5. As we can see, for the Plaza scenarios, the number of contacts remains constant, since the number of nodes in the area remains constant; nevertheless, when the renewal rate increases, the contacts have a lower duration; this reflects, as expected, the higher mobility of nodes. On the contrary, for the Station scenarios during rush hours, we can see that the number of contacts is greater but with a lower duration. Comparing both scenarios, the average contact duration for the station scenario is higher since now pedestrians must wait for the trains.

We also obtained the cumulative distribution function (CDF) of the contact duration (that is, $P(X \leq T)$), as shown in figure 5. For the Plaza scenarios, we can see that nearly 80% of all contacts have a duration lower than 10s for all the renewal rates, excluding the high renewal, where contact durations are lower. However, we can see a dichotomy in the contact duration for the station scenario since nearly 70% of the contacts have a duration lower than 20s, but there is a significant proportion of contacts between 20 and 300s. This clearly reflects the two class of contacts in the station: short-duration contacts between people moving in the station, and long-contact durations for people staying on the platforms. These duration patterns, as we will see in the evaluation section, have a huge impact on the delivery probability.

Considering the temporal distribution of contacts, figures 6 and 7 show the number of started contacts at a given time, and the number of node pairs that are in contact. We can see that, when the renewal of nodes is produced in the different scenarios, it generates a higher variability on the number of contacts.

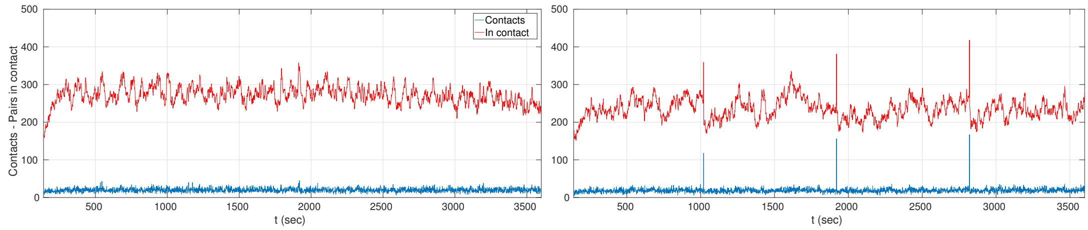
We also study the spatial characteristics of the contacts, as shown in the contour maps in figure 8, that represents the accumulated number of contacts per m^2. We can distinguish the pedestrians paths, where the number of contacts is low (and of short duration), and the high number of contacts concentrated in the points of interest or the station platforms, where the pedestrians stay and wait for a while observing the monuments or for the train arrivals. Furthermore, these contacts have long durations.

Finally, in order to identify the communities and the interrelations between nodes, we obtained the contact graph for four different scenarios, as shown in figure 9. These graphs were generated using the *Gephi* tool [37], and the nodes were distributed according to the Fruchterman–Reingold algorithm [38]. For the Plaza scenario with no renewal (see figure 9(a)), we cannot identify communities, as all nodes re-

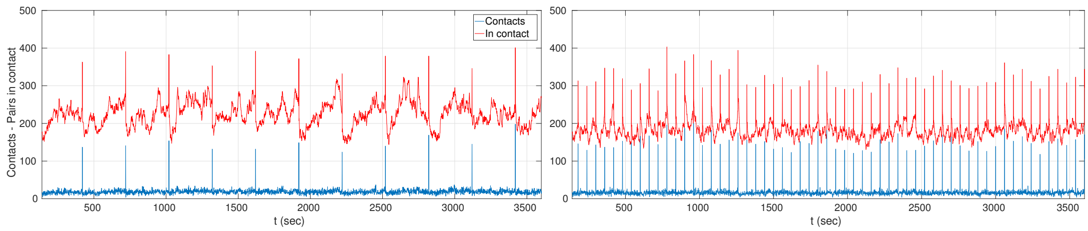


(a) Plaza scenarios. (b) Station scenarios.

FIGURE 5: Cumulative distribution function of the contact duration.

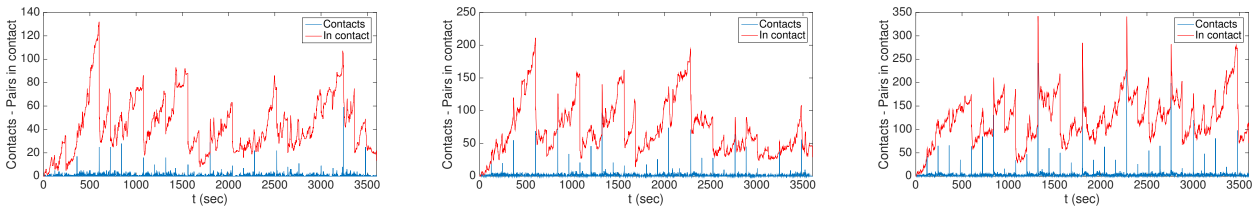


(a) No Renewal. (b) Low Renewal.



(c) Medium Renewal. (d) High Renewal.

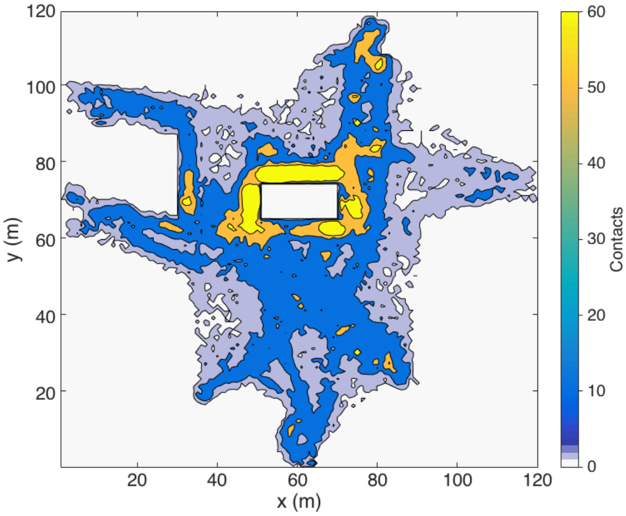
FIGURE 6: Number of contacts and pairs of nodes in contact for the Plaza scenarios.



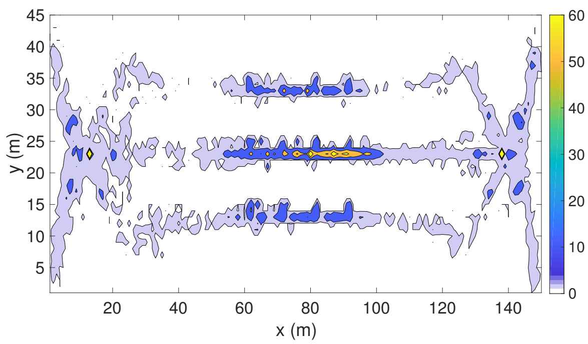
(a) Low (19-20h). (b) Medium (14-15h). (c) High (08-09h).

FIGURE 7: Number of contacts and pair of nodes in contact for the Station scenarios.

main in the Plaza and have contacts between them. Thus, the distribution of contacts, and the degree of the nodes (number of contacts) are very uniform. Nevertheless, for the scenario with mid renewal (figure 9(b)), we can identify communities of about fifty nodes, that corresponds to the number of nodes renewed every five minutes. Additionally, the members of these communities are also connected to the members of the nearby communities, reflecting the renovation of half of the nodes in the plaza. We can also identify the fixed nodes, as the two darker nodes in the centre, that have a connection with nodes in all the groups. When the renewal is increased, in Figure 9(c) we can identify a similar pattern, in which nodes are connected mainly to their neighbours. Lastly, for the station scenario (figure 9(d)), the contact graphs of all scenarios are very similar. We can identify several communities, which corresponds to people waiting in the platforms in the different train time intervals. Moreover, due to this larger waiting time, the nodes of these communities have more contacts (the nodes are greater), and with longer durations (the edges between them are thicker). The fixed nodes are clearly identified as the darker nodes at the centre of the graph.



(a) Plaza scenario - Medium Renewal.



(b) Station scenario - Medium (14-15h).

FIGURE 8: Contour maps of the different scenarios representing the accumulated number of contacts per m^2. The number of contacts has been truncated to 60, since there are some locations, such as the sender nodes, with more than 300 contacts. From this map we can see that most of the contacts occur mainly in the the points of interest, on the station platforms and near the sender nodes. Out of curiosity, the white rectangle in the centre of the Plaza scenario corresponds to a fountain (no contacts are possible).

In conclusion, the generated traces highlight the temporal and spatial characteristics that are very realistic and impossible to obtain using current synthetic mobility models. We will see that these aspects will have a strong impact on the diffusion of information.

V. IMPROVING LARGE-MESSAGES DIFFUSION

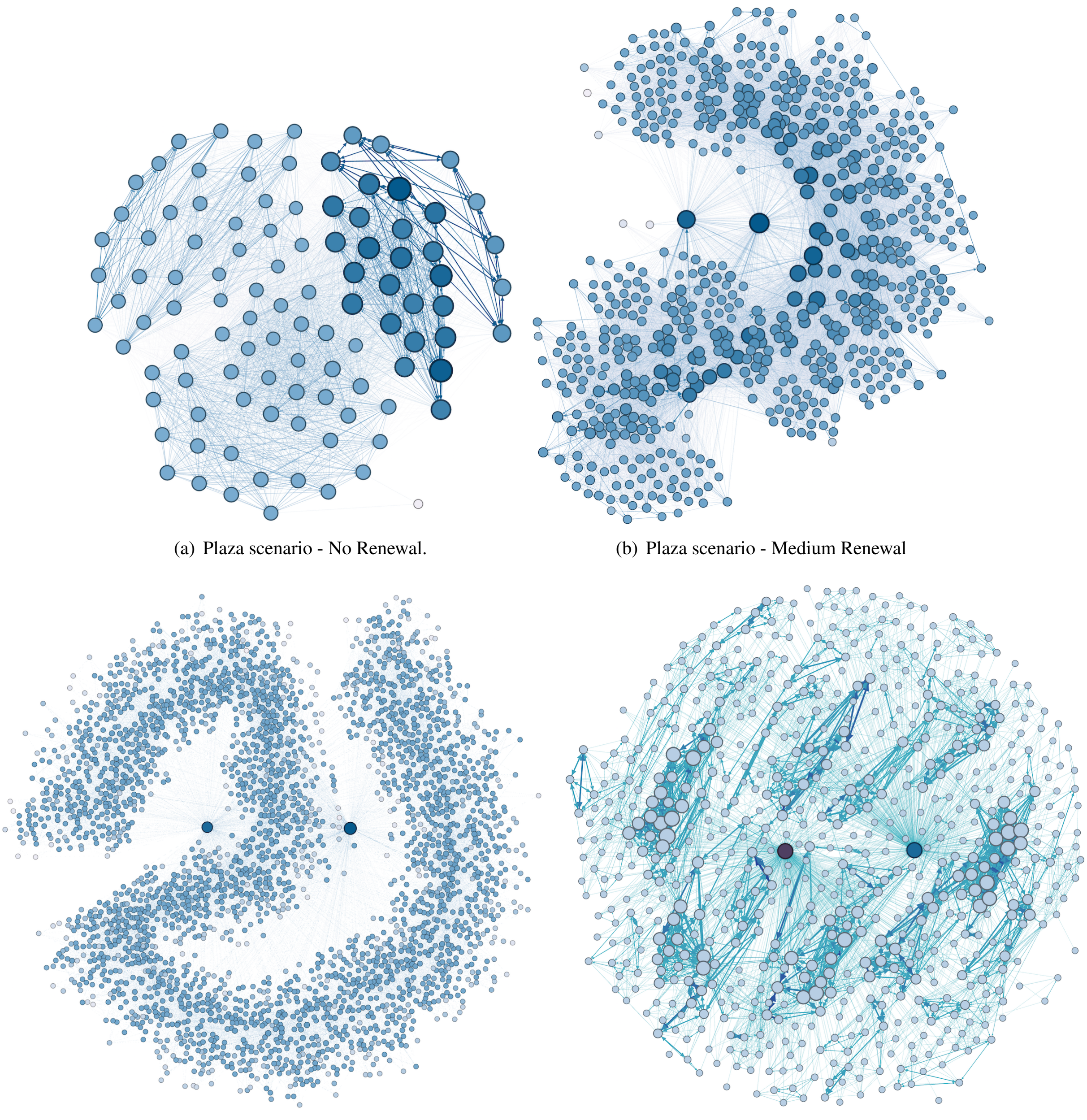
The main problem with the diffusion of large messages is that the transmission time is high, despite that most contacts have a low duration. So, in most cases, they cannot be successfully transmitted. For example, a 6MB message, assuming a Bluetooth connection as the one used in the simulations, will take roughly $6MB \cdot 8/2Mbps = 30s$, and, as shown in the CDF of the contact duration for the Plaza scenarios (see figure 5(a)), only 10% of the contacts have a duration greater than 30s. More critical are the contact durations of the fixed nodes, which are the source of the messages. For example, in the Station scenarios, the fixed nodes are at the entrance of the station, and so contact durations are of about 20s, making it impossible to successfully transmit these large messages, and meaning that the diffusion of the information cannot start, as detailed in the experiment section.

Therefore, in order to improve the diffusion of these larger messages, we propose a modification to the *Epidemic* protocol, named *EpidemicX2*, based on fragmenting the original message into two parts that are then transmitted by each of the fixed nodes. Basically, in this way, the fixed nodes have more opportunities to deliver these initial messages to the nearby nodes. Mobile devices can receive and assemble both parts in the locations where the number of contacts and their durations are longer (that, as shown in the contour maps in figure 8, corresponds to the points of interest in the plaza, and the platforms in the station). Following the previous example, a 6MB message can be fragmented into two parts, so the transmission time of each part is reduced to 15s. This time is lower than the duration of the contacts with the fixed nodes, allowing the beginning of the diffusion of these fragmented messages, and also increasing the possibility of successfully transmitting parts of the message when a contact between mobile nodes occurs. As will be shown in the experiments, this approach considerably increases the delivery ratio for pieces of information with sizes greater than 2MB, while also reducing the delivery time.

The implementation of the *EpidemicX2* protocol is straightforward in the mobile application, not requiring any modification in the transport or network layers[1]. First, the infrastructure will divide the information into two messages, with a convenient identification. These messages will be redirected to the fixed nodes, and a different order to transmit this messages is assigned. That is, the message transmitted first by each fixed node will will be different. This is important since, if both fixed nodes started with the transmission of the same message, the other message would not be transmitted in most cases. The mobile application is responsible of receiving, storing and relaying the two messages, as if it were the standard *Epidemic* protocol. Nevertheless, only when both messages are received, will the application notify to the user that the information was successfully received.

In the evaluated scenarios we considered only two fixed nodes, which are the locations with the highest contact opportunities. So, our initial proposal is to divide the message into two fragments. Nevertheless, this basic protocol can be extended to have further fragmentation, and thus improve the diffusion of very large messages, but at a cost of increasing the message overhead.

[1]Note that, in these layers, a large message is already fragmented into small packets to fit the network MTU (Maximum Transmission Unit) in order to improve the transmission performance. Nevertheless, the reception of the message is not successful until all the packets are received.



(a) Plaza scenario - No Renewal.

(b) Plaza scenario - Medium Renewal

(c) Plaza scenario - High Renewal

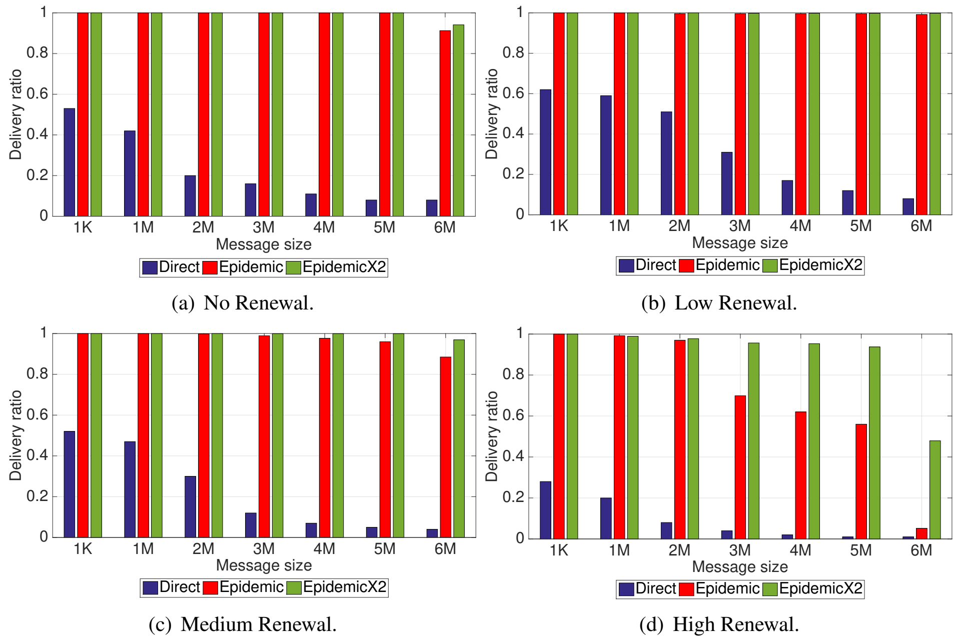
(d) Station scenario - Medium (14-15h).

FIGURE 9: Contact graphs for four different scenarios. Each line (edge) represents a contact between two nodes (that is, when two nodes are in range for a possible transmission). The thickness of the edges is proportional to its contact duration, and the size and darkness of nodes are proportional to their number of contacts (in graph theory, the *weighted node degree*).

Finally, the evaluation of the *EpidemicX2* protocol is performed as follows: The ONE is configured to start the diffusion of two different messages with half the size, also defining the fixed nodes as the initial senders of these messages. Then, the two messages are disseminated using the default *Epidemic* protocol. The ONE is also configured for sending the messages following a FIFO scheme, assuring that each fixed node first transmits a different part of the message. When a node receives the two message parts, it is considered that the information is delivered, taking into account only the reception time of the second part. This way, we can obtain the same metrics as for the already implemented protocols.

VI. PERFORMANCE EVALUATION

In this section, we describe the performance evaluation of information diffusion in the scenarios described in the previous sections. We consider three different performance metrics. Particularly, the most significant metric is the *delivery ratio*, which is obtained as the number of messages correctly delivered divided by the total number of generated messages. It is also important to determine how fast are these messages



(a) No Renewal. (b) Low Renewal.

(c) Medium Renewal. (d) High Renewal.

FIGURE 10: Delivery Probability of message diffusion protocols in the four Plaza scenarios.

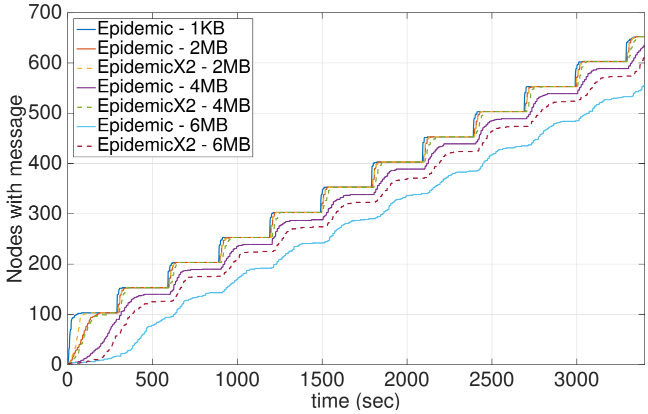


FIGURE 11: Evolution of the number of nodes with a copy of the message in the Plaza medium renewal scenario.

delivered, and so we also obtained the average *delivery delay*. Finally, for evaluating the *overhead* of the different protocols, we obtained two values: (1) the number of messages created, that comprises all started transmissions of messages, and (2) the number of messages relayed, that is, all the messages successfully transmitted.

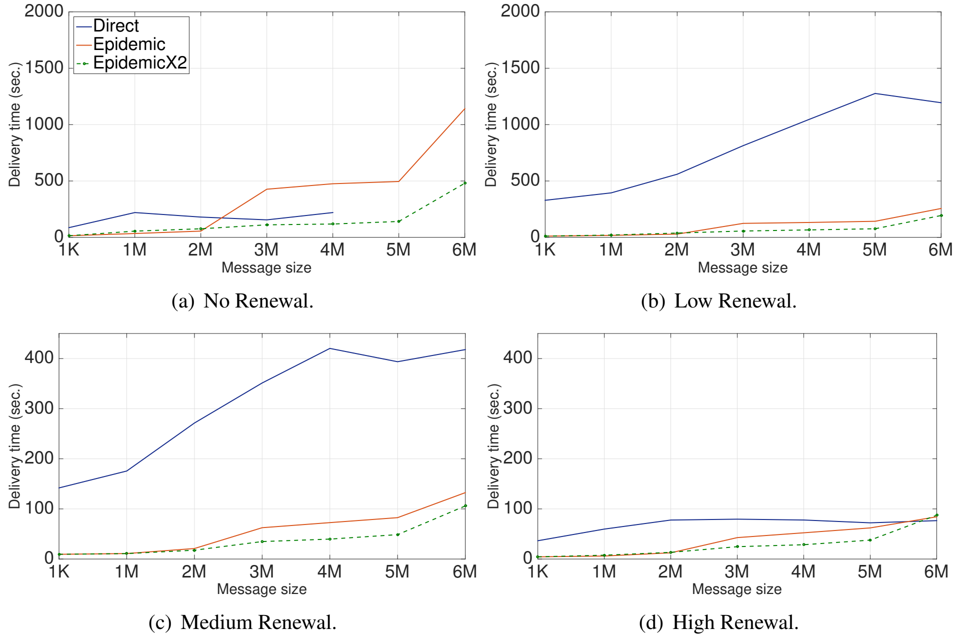
A. PLAZA SCENARIOS

We start with the Plaza scenarios. Figure 10 shows the delivery ratio depending on message size for the four different scenarios. First, we can see that *direct delivery* has a poor performance compared to the epidemic protocols, with delivery rates under 0.5 for 1K, and clearly decreasing its efficiency for greater message sizes, since they require greater transmission times.

For the *Epidemic* protocol, we can see that the efficiency for messages sizes under 2MB is almost 1, that is, nearly all nodes receive the message. For greater message sizes, we can see that the delivery ratio is reduced, particularly for the medium and high renewal scenarios. Increasing the renewal rate (from medium rate) implies that some nodes leave the place before the message can be delivered to them. In more detail, the results for a 6MB message size show that the delivery probability is reduced dramatically. The main reason is the transmission time of the messages, which is now roughly $6\text{MB} \cdot 8/2\text{Mbps} = 24\text{s}$. Taking into account that more than 90% of the contacts in the plaza have a duration of less than 24 seconds (as shown in figure 5(a)), this implies that very few messages are transmitted successfully from the fixed points (as shown with the direct delivery values, that are near zero), so the diffusion of the message cannot start.

In order to clarify the dynamics of this message diffusion, in figure 11 we plot the evolution of the number of nodes with messages for the medium renewal scenario. Regarding the *Epidemic* protocol, for small-sized messages (1KB) where the delivery rate is near to one, we can see that diffusion is performed very fast, spreading the message when people enter the plaza (the curve profile is a staircase). For greater message sizes, the diffusion becomes slower (levelling down the staircase pattern).

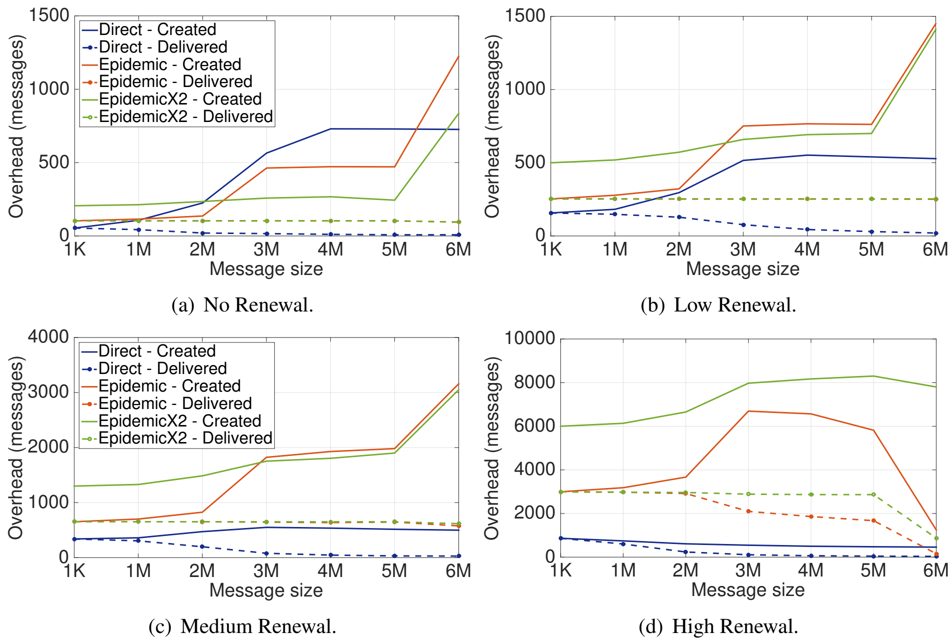
As detailed in section V, in order to improve the diffusion of these larger messages, we propose the *EpidemicX2* protocol. We can see in Figure 10 that the delivery rate is increased compared to the *Epidemic* protocol, particularly for messages sizes greater than 5MB. Furthermore, in figure 11, regarding the *EpidemicX2* protocols (4MB and 6MB), we can see that the diffusion is faster than the corresponding diffusion for the



(a) No Renewal. (b) Low Renewal.

(c) Medium Renewal. (d) High Renewal.

FIGURE 12: Average Delivery Time of message diffusion protocols in the four Plaza scenarios. When the delivered messages is under 10, the delivery time is not representative, and is thereby omitted in the graphs



(a) No Renewal. (b) Low Renewal.

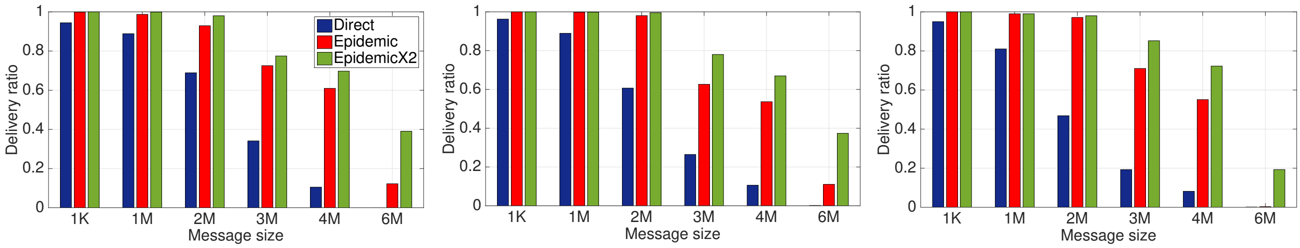
(c) Medium Renewal. (d) High Renewal.

FIGURE 13: Overhead of message diffusion protocols in the four Plaza scenarios.

standard epidemic protocols. Summing up, the experiments show that this approach considerably increases the delivery ratio for delivering information with large sizes.

Now, we evaluate the average delivery time (see figure 12). Note that these values have been calculated only with the messages that have been delivered. This way, when the delivery ratio is very low, that is, when only a few messages were received, the representativity of these results is minimal. Therefore, when the number of delivered messages is under 10, these values are omitted from the graphs.

In general, we can see that the delivery time increases with the message size, particularly for both epidemic protocols when messages sizes are greater than 3MB. Depending on the scenario, the results show that this delivery time decreases with the renewal rate. Particularly, for the scenario with medium or high renewal, as the nodes stay less time in the



(a) Low (19-20h). (b) Medium (14-15h) (c) High (08-09h).

FIGURE 14: Delivery ratio of message diffusion protocols in the three Station scenarios

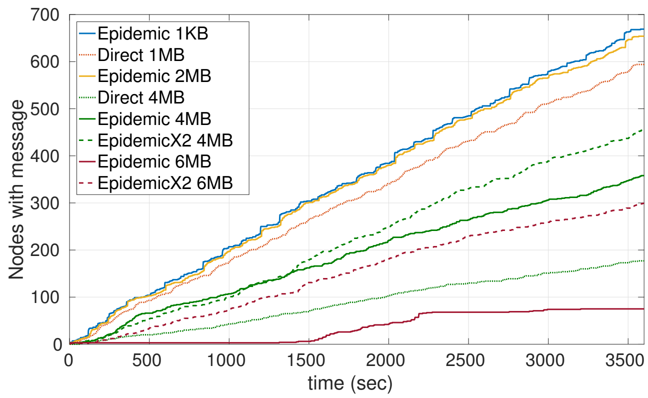


FIGURE 15: Evolution of the number of nodes with a copy of the message in the Station (Medium (14-15h) scenario).

place, the time (and opportunity) of receiving the message is reduced, and so is the delivery time. For example, in the high renewal scenario, the average delivery time remains below 100s. This makes sense since 60s is the average time that a node stays in the place in this scenario. On the contrary, in scenarios where the nodes stay longer, they have more time to deliver the message.

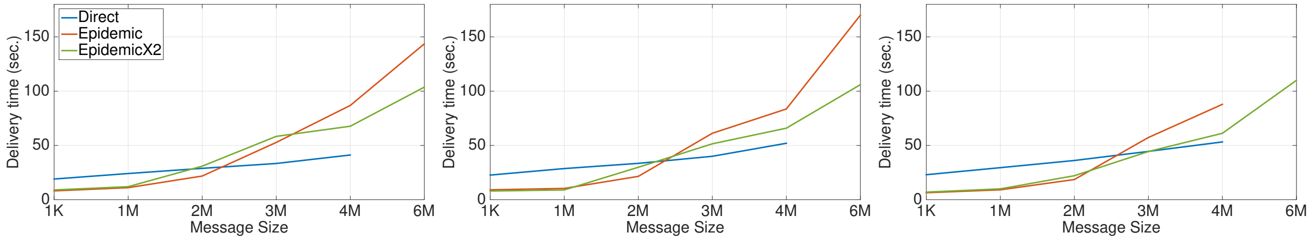
Finally, we evaluate the overhead results shown in figure 13. First, we analyse the results of the Direct Delivery, and we can see that the number of created messages increases with the message size, despite the number of delivered messages diminishes. The main reason is that, for large messages, the transmission time increases, and the delivery is reduced, increasing the number of (unsuccessful) retransmissions. For the epidemic protocols, we can see that, for small messages (up to 2MB), most messages are successfully transmitted as the number of created and delivered messages are about the same. Note that the *EpidemicX2* protocol creates at least twice the number of messages, but the delivered messages curve only increases when the second message is received. For this protocol, we can see that, for message sizes lower than 3M, the overhead is greater than for the other ones; however, for higher sizes, the overhead is reduced since the delivery ratio is greater, reducing the number of retransmissions. Besides, depending on the scenario, we can see that the overhead increases for scenarios with higher renewal, since the number of nodes and their mobility increases. Regarding the high renewal scenario results (see figure 13(d)), it is worth mentioning that the reduction of the created messages for 6MB messages is an effect of the longer transmission time, that reduces the number of contacts, and therefore the number of messages created.

B. STATION SCENARIOS

This subsection is intended to evaluate the information diffusion in the different Station scenarios. We start by evaluating the delivery ratio results shown in figure 14. We can see that, for very short messages (1KB), the delivery ratio is almost 1 for all protocols. The main reason is that the transmission time is very low, and so the message can be transferred even when the contact time is also very low. Increasing the message size has an evident impact on the performance of protocols, especially on the Direct Delivery, since the contact duration with the fixed nodes is sometimes not long enough to transmit the message completely, so most nodes do not receive the message. Particularly, the results for the 6MB message size show that the delivery ratio is reduced dramatically, and, even for the rush hour (High (08-09h) scenario), the *Epidemic* protocol cannot deliver any message. *EpidemicX2* increases considerably the delivery ratio for delivering contents with sizes greater than 2MB, and even allows the delivery of 20% of the 6MB-contents in the High (08-09h) scenario.

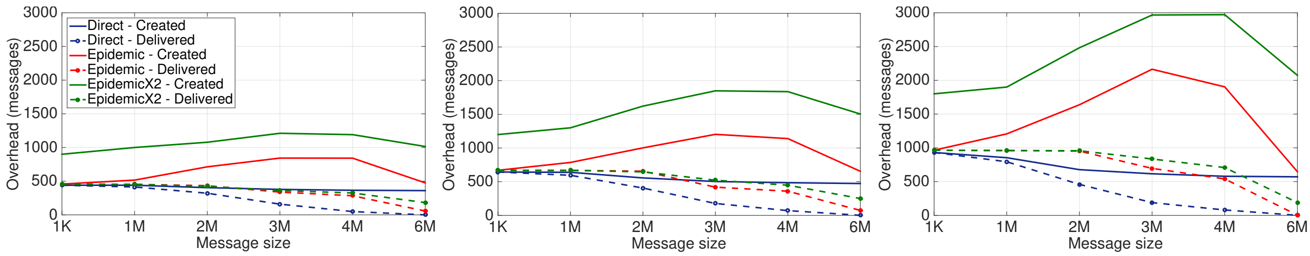
Figure 15 shows the diffusion of messages throughout time. For example, for the *Epidemic* protocol (messages sizes 1K and 2MB), the diffusion is steady along the nodes, although for direct delivery (1MB message size) this diffusion is slightly reduced. For 4MB messages, the rate of diffusion is lower, particularly for the direct delivery protocol. Finally, for 6MB messages, it is clearly shown that the diffusion for the *Epidemic* protocol did not start until 1500s time (when the fixed node can deliver the message); from then on, the diffusion is very low. Instead, since *EpidemicX2* divides this message in two, it permits to start the diffusion at the beginning of the experiment, and thereby increase the diffusion rate.

Figure 16 shows the results obtained when evaluating the average delivery time, which, in this particular scenario, is the average time that people must wait to obtain the information when they enter the station. Note that, as in the



(a) Low (19-20h). (b) Medium (14-15h). (c) High (08-09h).

FIGURE 16: Average Delivery Time of message diffusion protocols in the three Station scenarios. When the number of delivered messages is under 10, the delivery time is not representative, and is thereby omitted in the graphs.



(a) Low (19-20h). (b) Medium (14-15h). (c) High (08-09h).

FIGURE 17: Overhead of message diffusion protocols in the three Station scenarios

Plaza scenarios, when the number of delivered messages is under 10, the delivery time is omitted as not being representative. We can see that, as expected, as the size of the message increases, the delivery time also increases. It is important to note that the reason for having higher average delivery times for epidemic (compared to the direct delivery ones) is allowing for more time to receive the message, thus increasing the delivery ratio. For example, for a 2MB-message in the Medium (14-15h) scenario, direct delivery has a delivery ratio of 0.56, while with *Epidemic* it increases to 0.98. Therefore, the number of nodes that received the message increased from 375 to 650 (14 to 15 hours). Finally, we can see that *EpidemicX2* reduces the delivery time by fragmenting the message. Regarding the different scenarios, the results are very similar.

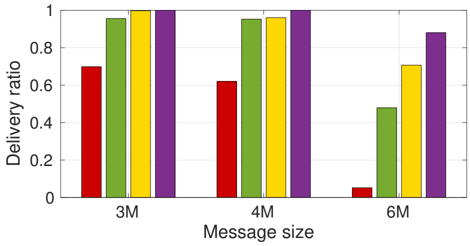
We finally analysed the performance in terms of overhead (see figure 17). We can see that, for a message size of 1KB, all messages are successfully delivered as both values are the same. When the size increases to 1MB, some messages are not delivered. Finally, for 6MB, most of them are not delivered. This can be seen as a worst-case situation for Opportunistic Networks, as most of the opportunities to transmit are wasted. In the case of our *EpidemicX2* protocol, the overhead is incremented since there are more messages created by its fragmentation policy, and at the same time more messages are delivered. Nevertheless, if we take into account the number of bytes transmitted, the overhead is equivalent to the *Epidemic* protocol. Finally, depending on the scenario, we can see that the overhead is increased for scenarios with the greater number of nodes, particularly for the rush hour (High (08-09h)).

In conclusion, considering all the scenarios, we can see that both people renewal and the number of nodes has a serious impact on the delivery performance. First, in scenarios where the number of nodes remains constant (the Plaza scenarios) increasing the renewal rate reduces the diffusion of messages. This is especially relevant for big messages, where the diffusion is very low. Additionally, in scenarios where the number of nodes is also incremented (the Station scenarios), the diffusion is also reduced. A way to increase this diffusion is the proposed *EpidemicX2* protocol that increases the delivery rate while also reducing the average delay. Although this supposes an increase in the number of messages generated, if we take into consideration only the number of bytes transmitted, the overhead is negligible.

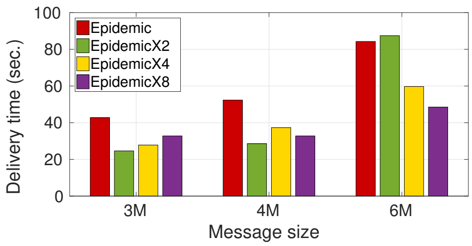
C. IMPACT OF MESSAGE FRAGMENTATION

In the previous *EpidemicX2* protocol evaluations, we considered fragmenting the message in two parts that are transmitted by the two fixed nodes. Now, we study the impact of fragmenting the original message in more fragments (particularly, in four and eight fragments, named *EpidemicX4* and *EpidemicX8*).

In general, the experiments performed for the different scenarios showed that the delivery ratio increases when the fragmentation is higher, particularly for the largest messages. As a sample of these experiments, we present the results of two scenarios in figures 18 and 19. We can see that the most significant increment of the delivery ratio is from *Epidemic* to *EpidemicX2*; for *EpidemicX4* and *EpidemicX8* the increase

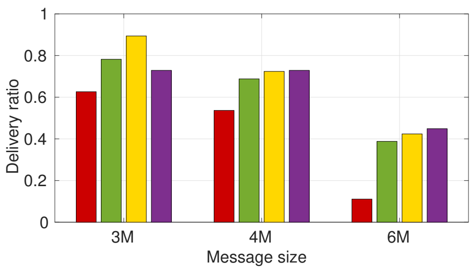


(a) Delivery Ratio.

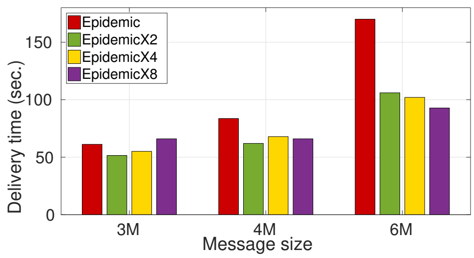


(b) Delivery Time.

FIGURE 18: Delivery Ratio and Delivery Time in the Plaza High Renewal scenario with different message fragmentation strategies.



(a) Delivery Ratio.



(b) Delivery Time.

FIGURE 19: Delivery Ratio and Delivery Time in the Station Medium (14-15h) scenario with different message fragmentation strategies.

remains marginal in most of the cases. The only exception is the Plaza scenario, in which figure 18(a) shows that the delivery ratio for 6MB messages increases substantially for *EpidemicX4*.

Regarding the delivery time (see figures 18(b) and 19(b)), the experiments showed that, in general, this time is slightly reduced from *Epidemic* to *EpidemicX2*, but for *EpidemicX4* and *EpidemicX8* there is not a clear trend. As the delivery time depends on the delivery ratio, in some cases an increase on the delivery rate can produce an increase on the time for receiving all the parts of the message.

The overhead of *EpidemicX4* and *EpidemicX8*, as expected, increases by a factor of nearly two and four, respectively, compared to the *EpidemicX2* overhead. Therefore, since the performance increase of the *EpidemicX4* and *EpidemicX8* protocols is not substantial superior to the one for the *EpidemicX2* protocol, we consider that *EpidemicX2* achieves a good trade-off between performance, efficiency and complexity of implementation.

Finally, another approach to improve the diffusion is adding more fixed nodes. Nevertheless, this is a costly option, as more infrastructure needs to be deployed. Furthermore, as messages would be mainly transmitted from fixed nodes to mobile nodes, it will not exploit opportunistic networking, which is the networking paradigm embraced by this paper.

VII. CONCLUSIONS

In this paper we proposed a methodology to improve the degree of realism when evaluating Opportunistic Networks. Specifically, we focused on the evaluation of crowded spaces with people renewal, i.e. situations where users that can either enter or leave the evaluated scenario. To the best of our knowledge, this is a problem that remains mostly untackled despite being one of the most suitable application areas for these networks.

The contributions of this paper are the following:

1) We have introduced a new methodology to generate realistic mobility scenarios in order to evaluate the performance of OppNets, based on the combination of realistic pedestrian simulators (PedSim) with OppNet simulators (The ONE).
2) We showed that the diffusion of information in these scenarios is mainly affected by two factors: mobility and message size. Higher node densities accelerates the diffusion, whereas, for the same node density, we discovered that the higher the renewal rates, the lower the delivery ratio. Moreover, we evidence the relation between message size and contact duration, which is also a key factor, showing that it can lead to very low performance when message size is large.
3) Finally, we proposed the *EpidemicX2* protocol, a variation of the *Epidemic* protocol that, compared to the standard Epidemic Protocol, increases the delivery ratio and reduces the average delivery time for large-sized messages. We also tested the *EpidemicX2* protocol considering and evaluating different schemes for partitioning and combining the messages in order to increase the delivery ratio.

REFERENCES

[1] L. Pelusi, A. Passarella, and M. Conti, “Opportunistic networking: data forwarding in disconnected mobile ad hoc networks,” *IEEE communications Magazine*, vol. 44, no. 11, 2006.
[2] C.-M. Huang, K.-c. Lan, and C.-Z. Tsai, “A survey of opportunistic networks,” in *Advanced Information Networking and Applications-*

Workshops, 2008. AINAW 2008. 22nd International Conference on. IEEE, 2008, pp. 1672–1677.

[3] J. Dede, A. Förster, E. Hernández-Orallo, J. Herrera-Tapia, K. Kuladinithi, V. Kuppusamy, P. Manzoni, A. bin Muslim, A. Udugama, and Z. Vatandas, "Simulating opportunistic networks: Survey and future directions," *IEEE Communications Surveys and Tutorials*, vol. PP, no. –, 2017.

[4] V. Vukadinović, Ó. R. Helgason, and G. Karlsson, "An analytical model for pedestrian content distribution in a grid of streets," *Mathematical and Computer Modelling*, vol. 57, no. 11–12, pp. 2933 – 2944, 2013.

[5] L. Pajevic and G. Karlsson, "Characterizing opportunistic communication with churn for crowd-counting," in *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2015, pp. 1–6.

[6] C. Gloor, "PEDSIM - Pedestrian Crowd Simulation," url-http://pedsim.silmaril.org, 2016.

[7] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proceedings of the 2nd international conference on simulation tools and techniques*, 2009, p. 55.

[8] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2001, pp. 1360–1369.

[9] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, 2007.

[10] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867 – 2891, 2007.

[11] K. Garg, S. Giordano, and A. Förster, "A study to understand the impact of node density on data dissemination time in opportunistic networks," in *Proceedings of the 2Nd ACM Workshop on High Performance Mobile Opportunistic Systems*, ser. HP-MOSys '13. New York, NY, USA: ACM, 2013, pp. 9–16.

[12] Y. Lin, X. Wang, L. Zhang, P. Li, D. Zhang, and S. Liu, "The impact of node velocity diversity on mobile opportunistic network performance," *Journal of Network and Computer Applications*, vol. 55, pp. 47 – 58, 2015.

[13] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara, "User mobility for opportunistic ad-hoc networking," in *Mobile Computing Systems and Applications, 2004. WMCSA 2004. Sixth IEEE Workshop on*, Dec 2004, pp. 41–50.

[14] Z. Feng and K.-W. Chin, "A unified study of epidemic routing protocols and their enhancements," in *Proceedings of the 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum*, ser. IPDPSW '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 1484–1493.

[15] E. Hernández-Orallo, J. Herrera-Tapia, J.-C. Cano, C. T. Calafate, and P. Manzoni, "Evaluating the impact of data transfer time in contact-based messaging applications," *IEEE Communications Letters*, vol. 19, no. 10, pp. 1814–1817, 2015.

[16] V. Vukadinovic and S. Mangold, "Opportunistic wireless communication in theme parks: A study of visitors mobility," in *Proceedings of the 6th ACM Workshop on Challenged Networks*, ser. CHANTS '11. New York, NY, USA: ACM, 2011, pp. 3–8.

[17] W. Moreira and P. Mendes, "Impact of human behavior on social opportunistic forwarding," *Ad Hoc Networks*, vol. 25, Part B, pp. 293 – 302, 2015.

[18] Y. Zhang and J. Zhao, "Social network analysis on data diffusion in delay tolerant networks," in *Proceedings of the Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '09. New York, NY, USA: ACM, 2009, pp. 345–346.

[19] J. Herrera-Tapia, A. Förster, E. Hernández-Orallo, A. Udugama, A. Tomas, and P. Manzoni, "Mobility as the main enabler of opportunistic data dissemination in urban scenarios," in *Proceedings of the 16th International Conference on Ad-Hoc Networks and Wireless, AdHocNow 2017*. Springer International Publishing, 2017, pp. 107–120.

[20] P. Hui, A. Lindgren, and J. Crowcroft, "Empirical evaluation of hybrid opportunistic networks," in *2009 First International Communication Systems and Networks and Workshops*, Jan 2009, pp. 1–10.

[21] N. Banerjee, M. D. Corner, D. Towsley, and B. N. Levine, "Relays, base stations, and meshes: Enhancing mobile networks with infrastructure," in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, ser. MobiCom '08. New York, NY, USA: ACM, 2008, pp. 81–91.

[22] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, March 2003, pp. 1543–1552.

[23] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2002, pp. 1079–1088 vol.2.

[24] J. Kangasharju, J. Ott, and O. Karkulahti, "Floating content: Information availability in urban environments," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*, 29 2010-april 2 2010, pp. 804 –808.

[25] M. Desta, E. Hyytia, J. Ott, and J. Kangasharju, "Characterizing content sharing properties for mobile users in open city squares," in *Wireless On-demand Network Systems and Services (WONS), 2013 10th Annual Conference on*, March 2013, pp. 147–154.

[26] C. Bettstetter, C. Wagner *et al.*, "The spatial node distribution of the random waypoint mobility model." *WMAN*, vol. 11, pp. 41–58, 2002.

[27] A. Mei and J. Stefa, "Swim: A simple model to generate small mobile worlds," in *INFOCOM 2009, IEEE*, April 2009, pp. 2106–2113.

[28] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: A new mobility model for human walks," in *INFOCOM 2009, IEEE*, april 2009, pp. 855 –863.

[29] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working day movement model," *Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models - MobilityModels '08*, pp. 33–40, 2008.

[30] T.-C. Tsai and H.-H. Chan, "NCCU trace: Social-network-aware mobility trace," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 144–149, 2015.

[31] F. A. Silva, C. Celes, A. Boukerche, L. B. Ruiz, and A. A. Loureiro, "Filling the gaps of vehicular mobility traces," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. ACM, 2015, pp. 47–54.

[32] Ó. Helgason, S. T. Kouyoumdjieva, and G. Karlsson, "Opportunistic communication and human mobility," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1597–1610, July 2014.

[33] J. Herrera-Tapia, E. Hernández-Orallo, A. Tomás, P. Manzoni, C. Tavares Calafate, and J.-C. Cano, "Friendly-sharing: Improving the performance of city sensoring through contact-based messaging applications," *Sensors*, vol. 16, no. 9, p. 1523, 2016.

[34] C. Souza, E. Mota, L. Galvao, P. Manzoni, J. C. Cano, and C. T. Calafate, "FSF: Friendship and selfishness forwarding for delay tolerant networks," in *Computers and Communication (ISCC), 2016 IEEE Symposium on*. IEEE, 2016, pp. 1200–1207.

[35] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," Duke University, Tech. Rep., 2000.

[36] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E*, vol. 51, no. 5, pp. 4282–4286, 1995.

[37] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009*, 2009.

[38] T. M. J. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software-Practice And Experience*, vol. 21, no. 11, pp. 1129–1164, 1991.