

This item is the archived peer-reviewed author-version of:

Optimising data diffusion while reducing local resources consumption in Opportunistic Mobile Crowdsensing

Reference:

Hernández-Orallo Enrique, Borrego Carlos, Manzoni Pietro, Marquez-Barja Johann, Cano Juan Carlos, Calafate Carlos T.- Optimising data diffusion while reducing local resources consumption in Opportunistic Mobile Crowdsensing
Pervasive and mobile computing - ISSN 1574-1192 - 67(2020), 101201
Full text (Publisher's DOI): <https://doi.org/10.1016/J.PMCJ.2020.101201>
To cite this reference: <https://hdl.handle.net/10067/1706020151162165141>

Optimising data diffusion while reducing local resources consumption in Opportunistic Mobile Crowdsensing.

Enrique Hernández-Orallo^{a,*}, Carlos Borrego^b, Pietro Manzoni^a, Johann M. Marquez-Barja^c,
Juan Carlos Cano^a, Carlos T. Calafate^a

^a*Departamento de Informática de Sistemas y Computadores. Universitat Politècnica de València, Valencia, Spain.*

^b*Department of Mathematics and Computer Science. University of Barcelona, Barcelona, Spain.*

^c*IDLab Research Group, University of Antwerp - imec, 2000 Antwerp, Belgium.*

Abstract

The combination of Mobile Crowdsensing (MCS) with Opportunistic Networking (OppNet) allows mobile users to share sensed data easily and conveniently without the use of fixed infrastructure. OppNet is based on intermittent connectivity among wireless mobile devices, in which mobile nodes may store, carry and forward messages (sensing information) by taking advantage of wireless ad-hoc communication opportunities. A common approach for the diffusion of this sensing data in OppNet is the epidemic protocol, which carries out a fast data diffusion at the expense of increasing the usage of local buffers on mobile nodes and also the number of transmissions, thereby limiting scalability.

A way to reduce this consumption of local resources is to set a message expiration time that forces the removal of old messages from local buffers. Since dropping messages too early may reduce the speed of information diffusion, we propose a dynamic expiration time setting to limit this effect. Moreover, we introduce an epidemic diffusion model for evaluating the impact of the expiration time. This model allows us to obtain optimal expiration times that achieve performances similar to those other approaches where no expiration is considered, with a significant reduction of local buffer and network usage. Furthermore, in our proposed model, the buffer utilisation remains steady with the number of nodes, whereas in other approaches it increases sharply. Finally, our approach is evaluated and validated in a mobile crowdsensing scenario, where students collect and broadcast information regarding a university campus, showing a significant reduction on buffer usage and nodes message transmissions, and therefore, decreasing battery consumption.

Keywords: Mobile Crowdsensing, Opportunistic Networking, Epidemic diffusion

1. Introduction

Mobile Crowdsensing (MCS) has become an appealing paradigm where mobile devices (such as smartphones, tablets, and wearables devices), through the sensors integrated into these devices, enable monitoring and collecting data of interest [1, 2, 3]. This captured data can be shared through the use of mobile Opportunistic Networking (OppNet), not requiring the use of fixed infrastructure, being helpful to offload the cloud and to reduce the associated costs and energy consumption [4]. This combination of MCS and OppNet is usually referred to as Opportunistic Mobile Crowdsensing [5].

Opportunistic Networking is a networking paradigm where communications take place upon the establishment of ephemeral contacts among mobile nodes using direct communication (i.e. Bluetooth or WiFi Direct) [6, 7, 8]. The main advantage of OppNet is that it supports low cost and seamless communication between devices regardless of their location, allowing the establishment of local communication channels that can be used for applications such as mobile social networks, vehicular networks, disaster and rescue operations, gaming, among others.

Two main communication schemes for OppNet can be considered [9]: *destination-less* and *destination-oriented*. Destination-less scenarios assume that all network participants should receive the message, or that the receivers are not known *a priori*. On the contrary, destination-oriented scenarios assume that the destinations are well known beforehand. In this paper we focus our study on destination-less (or broadcasting) scenarios, where mobile devices can share their sensed information with other devices through the peer-to-peer transmission of messages.

Broadcast delivery is widely used in OppNet, mainly based on using the epidemic protocol [10], a diffusion protocol where nodes may store, carry and forward messages toward all the possible contacted devices. In this epidemic behaviour, when a mobile node gets in contact with another, they try to exchange messages during their contact. Additionally, the diffusion of messages depends mainly on three factors [11]: the users' mobility, the transfer time, and the local buffer management. User mobility determines the number of contacts and their duration, while the size of the exchanged messages and the channel throughput determines the required transfer time. The

*Corresponding author.

Email addresses: ehernandez@disca.upv.es (Enrique Hernández-Orallo), carlos.borrego@ub.edu (Carlos Borrego), pmanzoni@disca.upv.es (Pietro Manzoni), Johann.Marquez-Barja@uantwerpen.be (Johann M. Marquez-Barja), jucano@disca.upv.es (Juan Carlos Cano), calafate@disca.upv.es (Carlos T. Calafate)

epidemic protocol can obtain low delivery times at the cost of overloading local buffers and flooding the network with message transmissions. On the one hand, this memory and network requirements can be a severe drawback when working with nodes that have constrained resources, such as IoT (Internet of Things) devices and simple wearable devices. On the other hand, even if local memory is not a big problem (as in modern smartphones), a huge local buffer can create management problems, increasing local latencies. Furthermore, it can also generate an excessive number of message transmissions, increasing network usage and battery consumption [11]. Therefore, to cope with these problems, several alternatives have been proposed to reduce the overhead of the epidemic protocol [12, 13, 14, 15, 16] based mainly on limiting the living copies of the messages considering key factors like the message size, the diffusion time, the node’s centrality, reliability or similarity, and the imposed overhead, among others.

An important aspect of OppNet, particularly when sensed data is broadcast, is the local persistence of the data or information. Note that OppNet is based on the local storage of messages, that is a kind of best-effort service where messages are generated locally, and where their lifetime and diffusion depend on the node’s mobility and resources. Hence, if a message is removed from the nodes, or, if the nodes with the message leave the evaluated area, the diffusion is affected, and the sensed data can be lost. This persistence of information has been evaluated in open city squares by Desta et al. [17] using a custom mobility model with spatial analysis and Markov chains, assuming that nodes enter and leave the city square. Additionally, in previous works [18, 19], we have analytically modelled the performance of mobile OppNet in city squares or gathering points, taking into account several social aspects such as the density of people, the dynamics of people arriving and leaving a place, the size of the messages, and the duration of the contacts.

In this paper, we propose reducing the local consumption of resources in mobile devices to optimise the diffusion of crowdsensing data. This idea was partially introduced in our previous paper [20], although only focused on OppNet. Particularly, we focus our study on reducing the local buffer usage by setting a local expiration time that forces the removal of old messages, also leading to lower message transmissions thereby reducing network usage. In order to obtain this optimal expiration time, we develop an analytical model for evaluating message diffusion considering several critical temporal parameters, such as transmission time and local buffer expiration times. We also derive an expression for obtaining the local buffer occupancy depending on the network load. This

model allows us to perform a macroscopic evaluation of the dynamics of the epidemic diffusion and the impact of these parameters on the diffusion.

60 To this end, we follow the fluid approximation, which are models based on Ordinary Differential Equations (ODEs) that are commonly referred to as *Population Processes*. *Population Processes* is an analytical method commonly used to model the dynamics of biological populations [21], which has also been extensively used for modelling Opportunistic Networks [22, 23, 24, 25].

This model is used to obtain the optimal expiration time. Then, we show that, when using
65 this expiration time message, the broadcasting has a performance similar to the epidemic protocol. Notably, the model shows that, for obtaining a diffusion performance only 1% slower than epidemic diffusion, the required expiration time does not need to be too high, and that this time depends mainly on the number of nodes and the communication time. Furthermore, the model shows that, when using this optimal expiration time, the buffer utilisation remains steady with the number
70 of nodes, whereas in other proposed approaches it increases sharply, and therefore, making our approach scalable.

Finally, as a proof of concept, we introduce an Opportunistic Mobile Crowdsensing application. Particularly, we have performed a realistic evaluation where a group of students may collect and broadcast information on a university campus. The students' movement is from real traces obtained
75 at the National Chengchi University (NCCU) campus [26]. This scenario is evaluated through simulation showing a significant reduction on the buffer and network usage, being the most efficient data diffusion scheme when compared to other buffer management approaches.

Summing up, in the context of opportunistic mobile crowdsensing, in this paper we present a novel approach that uses an analytical model in order to obtain the expiration time that produces
80 a diffusion very close to the epidemic protocol while reducing network resource consumption.

The paper is organised as follows: after reviewing related work in Section 2, we outline in Section 3 the contact-based diffusion and its temporal aspects, introducing a performance model used for evaluating the diffusion and buffer usage. Section 4 presents the evaluation using the model and the method for obtaining the optimal expiration time. Section 5 discusses the simulation results
85 for a Mobile Crowdsensing application obtained using simulation. Finally, Section 6 presents the conclusions of the paper.

2. Related work

Mobile Crowdsensing benefits from the processing and communication capabilities of available smart devices, which has enabled the development of different types of cooperative sensing applications. Typically, sensors that register participant information (e.g., location, movements) and environmental data (e.g., images, sounds) are very common in these mobile devices. Crowdsensing relies on a large number of participants to collect data from the environment through its integrated sensors [2]. These data, depending on the application, can be sent to a server or disseminated among users, using different communication technologies.

2.1. Mobile CrowdSensing (MCS) and Opportunistic Networking (OppNet)

As our proposal is focused on the diffusion messages, we review some of the MCS proposals addressing transmission issues. A large number of solutions are based on WiFi and Cellular communications [27, 28], although we can also find proposals that are based on Bluetooth [29] due to its flexibility and low consumption features. Additionally, we find that most solutions opted for either a centralised topology [27] or a distributed topology [4] with only a reduced number of proposals choosing a hybrid approach [30]. Specifically, Song et al. [4] propose the use of MANET (Mobile Ad hoc NETWORK) or Device-to-Device (D2D) networking, through the communication and sharing of crowdsensing data by vehicles near the event (such as a pothole on the road), termed as a local crowd.

Some researchers have already proposed the use of Opportunistic Networking (also referred to as *Opportunistic Networks*) to collect and broadcast sensed data [31, 32]. Thus, OppNet can serve as a complementary technology to avoid or minimise the use of fixed infrastructure, as well as to offload it. However, using OppNet to broadcast sensed data can generate high memory consumption and even battery depletion, which can lead to reducing user collaboration and network performance [33]. Thus, the reduction of local resources is a key issue for the use of OppNet with MCS.

2.2. Information Persistence and Propagation in OppNet

An important aspect of OppNet is the local persistence of information. Note that OppNet is based on the local storage of messages, that is a kind of best-effort service where messages are generated locally, and their lifetime and diffusion depend on the mobility and resources of mobility.

Hence, if a message is removed from the nodes, or if the nodes carrying messages leave the studied

area, the diffusion is affected. This persistence of information has been evaluated in open city squares by Desta et al. [17] using a custom mobility model, spatial analysis, and Markov chains, and assuming that nodes enter and leave the city square.

Recently, Pajevic et al. [34] have studied the conditions of content survival in such opportunisti-
120 tic networks, considering the user mobility patterns, as well as the time during which users keep forwarding the content, deriving an approximation based on stochastic differential equations. This persistence of information is also evaluated in previous works by the aforementioned authors of [17, 34] considering several communication aspects that impact the performance of message dissemination, such as the communication time and the duration of the contacts. In [Vehicular Networks](#),
125 [these issues have also been studied in \[35, 36, 37\]](#).

2.3. Analytical Performance of OppNet and MCS

The analytical modelling and performance evaluation of OppNet is one of the key challenging problems (according to a recent survey [38]). A common approach is to combine a network simulation tool with realistic mobility traces [8]. Nevertheless, simulation can be highly time-consuming
130 and restricted to the limited scenarios of the available mobility traces. Analytical models can avoid these drawbacks providing a fast and broader performance evaluation. It has been shown that under the assumption of a given contact rate between mobile nodes, the evolution of the number of infected nodes (nodes with messages) can be modelled as a Markov chain [12]). These Markovian models have been effectively used for modelling several aspects of DTN (Delay-Tolerant
135 Networking) and Opportunistic Networking [22, 15, 39, 40, 41]. Nevertheless, the resulting Markov chain is only amenable for simple models, so several approximations have been proposed as a simpler solution for these processes, such as branching processes, fluid approximation, or stochastic differential equations [42].

The temporal aspects of OppNet have been taken into account by the authors, introducing
140 several analytical dynamic models based on Delay Differential Equations (DDEs). Particularly, in [18], we analytically modelled the performance of mobile OppNet in city squares or gathering points (open systems), taking into account several social aspects such as the density of people, the dynamics of people arriving and leaving a place, the size of the messages, and the duration of the contacts. This model was extended considering also fixed nodes, and showing the diffusion
145 coverage in several realistic scenarios [19].

In this paper, we follow the fluid approximation, which are models based on Ordinary Differential Equations (ODEs) that are commonly referred to as *Population Processes* [21]. This approach has been extensively used for modelling Opportunistic Networks [22, 23, 24, 25]. Specifically, Haas and Small [22] presented a model based on epidemiological processes for a network that used animals (whales) as data carriers to store and transfer messages (an approach similar to DTN). Zhang et al. [23] derived ODE equations for the study of the dynamics of various forwarding and recovery DTN schemes, such as epidemic and 2-hop, among others. De Aubreu et al. [24] introduced a mathematical approach for messages diffusion in opportunistic networks using the epidemic protocol. This approach is based on well-known models for the spreading of human epidemical diseases, e.g. SIR (Susceptible, Infectious and Recovered) models. One of the main conclusions of their analysis (mathematical model and its respective simulation) is that SIR models are quite accurate for the average behaviour of epidemical DTN.

Xu et al. [25] propose a detailed analytical model to analyse the epidemic information dissemination in mobile social networks. It is also based on SIR models including rules that concern user's behaviour, especially when their interests change according to the information type, and it can have a considerable impact on the dissemination process. After large simulations, they have demonstrated the accuracy of their model. The epidemic model has also been used to model other diffusion models considering heterogeneous pairwise contact rates [43].

Differently from the aforementioned proposals, other authors relied on different mathematical methods. Whitbeck et al. [41] present an analytical study describing the performance of the epidemic protocol, arguing that intermittently-connected mobile networks can be modelled as edge-Markovian dynamic graphs. The authors propose a new model for epidemic propagation based on such graphs, and calculate a closed-form expression that links the best achievable delivery ratio to common OppNet parameters such as message size, maximum tolerated delay, and link lifetime. In this research work, the authors have shown that, given a certain maximum delay and node mobility, bundle size has a major impact on the delivery ratio. Finally, there are other proposals such as the ones presented by Su et al. [44], and Feng et al. [45], who evaluate the message dissemination behaviour of the epidemic protocol by focusing on the mobility patterns of the nodes. In such works, authors explain the relationship between factors such as speed, mobility model, and node density in the target regions.

2.4. Buffer management and resource consumption

Besides the mobility of the nodes, the performance of OppNet 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 protocol adopted, as detailed above. Regarding the second aspect, the impact of removing messages on local buffers was initially studied by Zhang et al. [23] for the epidemic diffusion. The authors introduced a basic epidemic model (*Limited-time forwarding*) where the expiration time is drawn from an exponential distribution with a given rate. Actually, the model considers, instead of a given expiration time, a simple exit rate that models when nodes delete their local message (that is, as known in the epidemic process, a death rate). In our recent works [46, 11], we show that it is important to implement certain mechanisms to improve buffer management in the context of epidemic diffusion, showing that the best results were obtained when local buffer management follows a combination of smallest message forwarding and largest message dropping. Social aspects can also be considered in the management of local buffers and in the message forwarding strategy. In this context, Zhang et al. [47] 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.

Another important factor in the design of OppNet is the consumption of resources, mainly local memory and the number of messages transferred. The goal is to make information diffusion effectively using the minimum possible energy. With this goal in mind, Borrego et al. [16] propose a broadcast dissemination protocol for messages that is efficient with respect to message latency and message dissemination per unit of energy using Optimal Stopping Theory to select the best message storers. Previous approaches, such as the one presented by Goundan et al. [48], propose a protocol to also reduce the number of transmissions waiting for an opportunity to reach multiple nodes, while another work by Gao and Cao [49] relied on the social centrality metric to ensure an effective relay selection for also reducing message transmissions.

2.5. Conclusions

To conclude, in this paper we present a novel approach for minimising the consumption of network resources that is essentially different from the previous works. It is based on using an analytical model to estimate the diffusion of a message, so nodes can calculate an optimal expiration

205 time in order to achieve diffusion of crowdsensing data that is very close to the epidemic protocol, yet with a significant reduction of network resources consumption. Additionally, our model also gives a deep analysis of how the different temporal aspects of epidemic diffusion impact the dynamics of message spreading.

This paper extends our previous contribution [20], which introduced a simpler model only
210 suitable for Opportunistic Networks. In this paper, we have revised and improved the analytical model with a deep analysis of the components that impact the diffusion. Now, it is not only focused on reducing the buffer utilisation but also on reducing network usage, which is a key issue in mobile crowdsensing applications. We also include more experiments aimed at understanding how the expiration time impacts the diffusion. Particularly, section V is completely new, detailing
215 the efficiency of our diffusion method in a realistic mobile crowdsensing simulated scenario.

3. Contact-based diffusion and model

In this work, we focus on Opportunistic Mobile Crowdsensing applications, which are contact-based messaging applications that may sense data, and can establish a short-range communication link among devices in order to exchange this sensed data, storing these data locally in order to
220 achieve their full broadcast. In particular, we can assume a wireless peer-to-peer (P2P) network of nodes that connect opportunistically. No data (messages) are sent or stored in servers; instead, all information is stored on the mobile devices in a given area.

Therefore, in this section, we first detail how this diffusion is performed, and the temporal parameters that affect the performance of such diffusion. Then, we propose an analytical model
225 considering these temporal parameters based on *population processes or models*. We also study the buffer occupancy under different network loads, which will be used in the evaluation section. Finally, we end with the validation of the model.

3.1. Temporal aspects of the epidemic diffusion

Message spreading is based on epidemic diffusion, a concept similar to the spreading of infectious
230 diseases, where an infected node (the one that has a message) contacts another node to infect it (transmit the message). If there are not local buffer constraints, epidemic protocol obtains the minimum delivery delay at the expense of increased usage of buffers, and an increased number of transmissions.

Message dissemination takes place as follows. Mobile devices are running a crowdsensing application, which is responsible for capturing and broadcasting the sensed data, and finally notifying the user of the broadcast information captured from other devices. Each node has a limited buffer where the messages in transit can be stored. When two nodes establish a pairwise connection, they exchange the messages they have in their buffers and check whether some of the newly received messages are suitable for notification to the user. All nodes that have the crowdsensing application collaborate in storing and forwarding messages.

Message diffusion is affected by several temporal aspects. The most important ones are the communication time, and the contact duration. The communication time is the required time to transmit a set of messages, and depends on the available bandwidth and the size of the messages. This communication time can include the set-up time to establish a connection, that in some technologies it is not negligible (for example, Bluetooth). Therefore, the communication time can be expressed as $T_c = T_s + T_t$, where T_s is a fixed set-up time, and T_t is the total transmission time of the messages, such as $T_t = \sum_i m_i / Bw$, where m_i are the message sizes and Bw the available bandwidth.

The contact duration (T_d) depends on the mobility of the nodes and the communication range. This contact duration limits the number of messages to be transmitted when a contact occurs. If the required communication time is lower than the current contact duration ($T_t \leq T_d$) all messages will be transmitted successfully. On the contrary, if $T_t > T_d$ only a portion of the messages will be transmitted (and even no message would be transmitted if the contact duration were too short). The communication time also impacts the diffusion speed, since when two nodes are transmitting they cannot transmit other messages to other nodes, and so the diffusion is delayed when communication time is long.

As stated previously, the fast message spreading of epidemic diffusion comes at the expense of high utilisation of local buffers, and a high number of transmissions impacting the overall performance. Each mobile node has a limited buffer where the messages are stored for future exchanges. Several approaches can be considered for the local buffer management. *Passive* approaches are based on storing all received messages until no room is available. Then, when a new message is received, the local node should decide which message(s) are to be removed in order to provide space for newly received messages [11]. On the other hand, *active* approaches try to remove messages

when they are no longer required. The problem here is how to determine that diffusion has com-
265 pletely finished. Recovery schemes, such as the ones proposed by Hass and Small [22], where the
destination, upon receiving the message, sends back to all the nodes an *erase* message notification
to remove the local copies, are only valid for unicast transmission. The extension of this removal
method to broadcast transmission is not amenable as it would require to determine what is the
last node that receives the message.

270 Therefore, in the case of broadcast diffusion, we must use temporal approaches. The most
common method is using the TTL (Time-to-Live), that is, the time during which a message exists
in the network following its initial transmission. All messages with a lifetime greater than their
TTL are directly removed from all the local buffers, so no further diffusion of this message is
performed.

275 Another temporal approach method is to consider a local expiration time T_e , which is the time
the message is to be stored in the local buffers following its reception. Nevertheless, this expiration
time has also an impact on the diffusion of the messages, since it implies that, after a time, a
message is removed for future exchanges, reducing the number of copies in the network. In order
to minimise the effect of this reduction on the number of copies, we introduce a model to evaluate
280 the impact of this expiration time on the diffusion process, which can also be used to obtain the
optimal expiration time that minimises the impact on the diffusion.

3.2. Diffusion model

The diffusion model presented in this paper is based on *population processes or models*. Pop-
ulation models are well-known mathematical models that study population dynamics and they
285 are extensively used to model biological populations such as the spread of parasites, viruses, and
diseases [21]. Specifically, our model is based on *biological* epidemic models [42], where individuals
can be infected when a contact occurs with other infected nodes. In our case, an infected node
refers to a node that has a message, and infection, when a node that has a message transmits it to
another node.

290 For the proposed model, we assume a set of N mobile nodes (population) that move freely in an
area with a given contact rate between pairs λ (also known as pairwise meeting rate). We assume
a short-range communication range (for example, Bluetooth or WiFi direct).

Table 1: Notation table for the model.

Symbol	Definition
N	Population
λ	Pairwise contact rate.
T_c	Communication time. The required time to transmit a message.
T_e	Expiration time. Time a message is stored in a node following its reception.
S_c	Class of Susceptible nodes communicating (receiving the message).
S_w	Class of Susceptible waiting nodes (i.e. waiting to receive the message).
I_c	Class of Infected nodes communicating (i.e. transmitting the message).
I_e	Class of Infected nodes where the message has expired (i.e. removed from the node).
I_a	Class of Infected nodes that still have the message (and are not communicating).
I	All infected nodes (i.e. $I(t) = I_c(t) + I_e(t) + I_a(t)$).
h	Step size for the Euler's method.
t_c, t_e	Delay indexes for T_c and T_e .
D_t	Delivery time. Time when nodes get the message.
$C(t)$	Number of messages transmitted up to time t (equivalent to number of local copies).

Following the epidemic model notation¹, the population is divided mainly into two main classes: the *infected* nodes (I) and the *susceptible* nodes (S), so population remains constant: $N = I + S$.
 295 In the basic epidemic model [23], the number of nodes in a place remains constant, and when a node carrying the message (an *infected* node) contacts with another node that does not have the message (called the *susceptible* node) it transmits this message immediately. From that moment on, both nodes carry the message. We extend this basic model by considering the communication and expiration times of the messages.

300 In our proposed model, when an infected node contacts a susceptible node, both nodes establish a connection and the transmission of the message starts for a given communication time T_c . During this communication time, the nodes involved cannot infect other nodes, so two new sub-classes of nodes are introduced: the susceptible nodes communicating (S_c), and the infected nodes communicating (I_c). Thus, when a contact occurs, S_c and I_c are increased by one, one coming

¹See Table 1 for a reference of the notation used in the model.

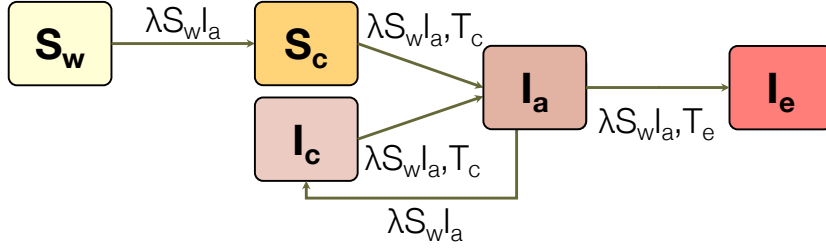


Figure 1: The diffusion model used, with the transitions between classes with the expiration time of messages

305 from the class of infected nodes, and another one from the class of susceptible nodes. When the transmission ends, these two nodes are moved to the infected nodes class. The susceptible nodes that are not communicating are denominated susceptible waiting nodes S_w .

Messages have an expiration time of T_e , which is defined as the time that a message is stored in a node since it has been received. Note that $T_e > T_c$ in order to allow the complete reception of the message. After this time, the message is removed, and this node passes to a new subclass: the expired infected nodes I_e . Further on, the subclass of nodes that can still infect is denominated infected alive nodes: I_a . Summing up, we have three different subclasses of infected nodes, but only the alive ones can infect other nodes. On the other hand, expired infected nodes cannot infect further nodes since the message is removed from their local buffers. Thus, this implies a
 315 new transition from class I_a to class I_e , with a delay of T_e . In detail (see Figure 1), the transitions between classes are the following²:

- $(S_w \rightarrow S_c, \lambda S_w I_a, 0)$: a susceptible waiting node contacts a node with the message to start the reception of the message with contact rate λ .
- $(I_a \rightarrow S_c, \lambda S_w I_a, 0)$: infected nodes contact a node with no message to start the transmission
 320 of the message with contact rate λ .
- $(S_c \rightarrow I_a, \lambda S_w I_a, T_c)$: the reception of the message ends after T_c seconds, so now the node is infected.
- $(I_c \rightarrow I_a, \lambda S_w I_a, T_c)$: the transmission of the message ends after T_c seconds, so the node returns to the infected class.

²The notation used for the transitions is the following: $(A \rightarrow B, \text{transition rate}, \text{delay})$, where A and B are classes, and the third value represents the delay incurred, that can be 0 (no delay), T_c or T_e .

- $(I_a \rightarrow I_e, \lambda S_w I_a, T_e)$: the message is removed from the nodes after the expiration time, and then it moves to the I_e class.

Using these transitions, we can model the dynamics of this system using the following Delay Differential Equations (DDEs):

$$\begin{aligned}
S'_w(t) &= -\lambda S_w(t) I_a(t) \\
S'_c(t) &= \lambda S_w(t) I_a(t) - \lambda S_w(t - T_c) I_a(t - T_c) \\
I'_c(t) &= \lambda S_w(t) I_a(t) - \lambda S_w(t - T_c) I_a(t - T_c) \\
I'_a(t) &= 2\lambda S_w(t - T_c) I_a(t - T_c) - \lambda S_w(t) I_a(t) \\
&\quad - \lambda S_w(t - T_e) I_a(t - T_e) \\
I'_e(t) &= \lambda S_w(t - T_e) I_a(t - T_e)
\end{aligned} \tag{1}$$

From Eq. (1) we can obtain the whole number of susceptible nodes and infected nodes as $S(t) = S_w(t) + S_c(t)$ and $I(t) = I_c(t) + I_a(t) + I_e(t)$. Particularly, Eq. (1) can be solved numerically using Euler's method, with a step size of h and time $t_i = hi$:

$$\begin{aligned}
S_w[i+1] &= S_w[i] + h(-\lambda S_w[i] I_a[i]) \\
S_c[i+1] &= S_c[i] + h(\lambda S_w[i] I_a[i] - \lambda S_w[i - t_c] I_a[i - t_c]) \\
I_c[i+1] &= I_c[i] + h(\lambda S_w[i] I_a[i] - \lambda S_w[i - t_c] I_a[i - t_c]) \\
I_a[i+1] &= I_a[i] + h(2\lambda S_w[i - t_c] I_a[i - t_c] - \lambda S_w[i] I_a[i] \\
&\quad - \lambda S_w[i - t_e] I_a[i - t_e]) \\
I_e[i+1] &= I_e[i] + h(\lambda S_w[i - t_e] I_a[i - t_e])
\end{aligned} \tag{2}$$

with the following history: $S_w[0] = N - 1, S_c[i] = I_c[i] = I_e[i] = 0 \quad \forall i \leq 0$ and $I_a[i] = 0 \quad \forall i < 0, I_a[0] = 1$. The delay indexes are obtained as $t_c = \lceil T_c/h \rceil$ and $t_e = \lceil T_e/h \rceil$.

We can derive other metrics from these equations. Concretely, from $I(t)$ we can obtain the number of times a message is transmitted up to time t as the number of nodes that have been infected, that is $C(t) = I(t) - 1$. Note that if the message is spread among all the nodes this value will be $N - 1$. We can also obtain the delivery time D_t , which is the time when a given number of nodes M get the message. This can be obtained from the total of infected nodes, calculating the smallest value of i that makes $I[i] \geq M$, so D_t will be the time hi :

$$D_t = hi_m, \quad i_m = \min\{i : I[i] \geq M\} \tag{3}$$

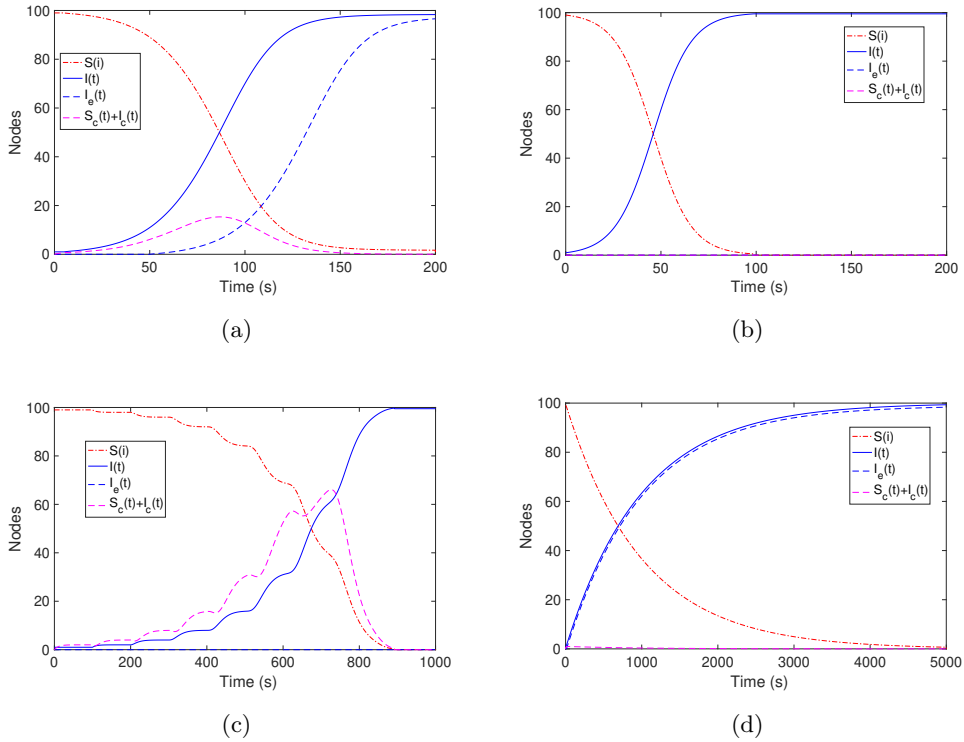


Figure 2: Evolution of the susceptible nodes $S(t)$, infected nodes $I(t)$, expired infected nodes $I_e(t)$ and communicating nodes ($S_c(t) + I_c(t)$) for different values of T_c and T_e : a) $T_c = 5s, T_e = 50s$; b) $T_c = 0s, T_e = \infty$; c) $T_c = 50s, T_e = \infty$; d) $T_c = 5s, T_e = 6s$

3.3. Dynamics of the modelled system

330 Regarding the dynamics of the modelled system, we study its evolution depending on the different temporal parameters. The impact of the contact rate and the number of nodes has been extensively studied in [23, 44, 45, 18], showing a faster message diffusion when the contact rate and number of nodes are higher. Therefore, for all the evaluations, the contact rate is set to $\lambda = 0.001$, and the considered number of nodes (N) is 100. Figure 2 shows the evolution of the number of
335 infected nodes, susceptible nodes, expired infected nodes, and transmitting nodes depending on time (using Eq. (1)).

Figure 2a shows a typical epidemic diffusion evolution. Regarding the expired infected nodes, we can see that at $t=50s$ (expiration time) their increase is exponential, therefore reducing the communicating nodes. We can compare this evolution with the results obtained when the communication time is zero ($T_c = 0$), and when no message is removed from the local buffer (that

is, when the expiration time is infinity, $T_e = \infty$). The results are shown in Figure 2b, where we can see that the diffusion is faster since communication time is not considered. Also, there are no communicating nodes since this time is assumed to be zero. Therefore, the diffusion is equivalent to the basic epidemic model, which has a simple analytical solution [23]:

$$I^\lambda(t) = \frac{N}{1 + (N - 1)e^{-\lambda Nt}} \quad (4)$$

When the communication time is longer (and considering $T_e = \infty$), we can see in Figure 2c that the diffusion is slower (practically the diffusion time is multiplied by 10), and that the number of infected nodes is doubled after each message transmission with a period T_c (a kind of saw-teeth patterns on the curves). This diffusion can be approximated by the following function [50]:

$$I^c(t) = \min(2^{\lfloor t/T_c \rfloor}, N) \quad (5)$$

Finally, when the expiration time is very near to T_c (in our example $T_e - T_c = 1$), we can see in Figure 2d that the diffusion is extremely slow. In this case, the diffusion is basically performed node by node, as message life in the nodes does not allow to forward it to other nodes. That is to say, its performance is downgraded to the one obtained with the *Direct Delivery* protocol [51]. Therefore, the diffusion is exponential, and the accumulated number of nodes that have received the message can be obtained as follows:

$$I^e(t) = N(1 - e^{-\lambda t}) \quad (6)$$

This way, expressions $I^\lambda(t), I^c(t)$ and $I^e(t)$ reveal three different diffusion patterns that will impact the effective message diffusion $I(t)$. Summing up, given enough time the diffusion will cover all the nodes. Even when the expiration time is very close to the communication time, the
340 diffusion goes on as the message still remains in the source node, although very slowly.

3.4. Buffer occupancy

In order to study the buffer occupancy, we have to consider a network load that may consist of a set of periodically generated messages, which contain the sensed data that are intended to be spread among a set of N nodes. Specifically, the data sensing and message generation process is
345 assumed to follow a Poisson process with an arrival rate of β .

If a message that has started its diffusion at time t has a delivery time D_t , and considering that its average life is its expiration time T_e , then, according to Little's law, we have that the average

number of message copies per time unit between t and $t + D_t$ is $N \cdot T_e / D_t$. Therefore, as messages are created with a rate of β , adding new copies during its delivery time of D_t , the average number of local copies of messages in the nodes can be obtained as follows:

$$E[C] = (\beta \cdot D_t) N \cdot T_e / D_t = N\beta T_e \quad (7)$$

Then, dividing by the number of nodes, the average per-node buffer occupancy is simply $E[Q] = \beta T_e$. Note that this expression is the result of assuming that all messages have the same expiration and delivery time.

For a more realistic evaluation, we also consider a non-homogeneous workload with L different messages types, each with different arrival rates β^i and communication times t_d^i (that is, with different messages sizes m_i). Then, for each message type, we can obtain the average number of copies using Eq. (7), so the average number of messages in the network and the local buffers are obtained as:

$$E[C] = N \sum_{i=0}^L \beta^i T_e^i \quad E[Q] = \sum_{i=0}^L \beta^i T_e^i \quad E[B] = \sum_{i=0}^L m_i \beta^i T_e^i \quad (8)$$

Finally, we also include in the previous expressions the local buffer usage in bytes $E[B]$. Note that this local buffer usage, as we will show in the following sections, impacts on the network utilisation since a lower buffer usage implies a reduction on the number of bytes transmitted.

3.5. Model validation

The model introduced in this section is validated using a procedure similar to the ones described by Zhang et al. [23], and our previous work[18]. The goal is to compare the results obtained using the analytical models with those obtained by simulation, specifically the entire set of infected nodes $I(t)$ up to a given generated time. The simulator is driven by contacts and uses the same input parameters as the model. Contacts are generated with an inter-contact distribution following an exponential distribution with mean $1/\lambda$.

The validation procedure is outlined in Algorithm 1. It is based on a set of 1000 random tests, where each test uses a set of different input values that are randomly generated. For each test i , the test program generates the parameters randomly and obtains $I(t)$ from our model, using the Euler's method (function `ModelEuler`). Using the same parameters, the simulation is repeated 500 times (function `SimDiffusion`), generating for each simulation contacts with a rate λ up to time T_{sim} in order to obtain a mean of the number of infected nodes ($\overline{I^S}$). Note that each simulation

Algorithm 1 Validation Process. This function returns a vector with the relative error of each performed validation test.

```

1: function VALIDATE_MODEL return  $\epsilon$ 
2:    $h = 0.01$ ; // time step for Euler's solution of DDE model
3:   TESTs = 1000; SIMs = 500;
4:   for  $i = 1$  to TESTs do
5:     // Generate randomly the input parameters
6:      $T_{sim} \sim \mathcal{U}(10, 10000)$ ;  $\lambda \sim \mathcal{U}(0.001, 0.1)$ ;
7:      $N \sim \mathcal{I}(50, 1000)$ ;
8:      $T_c = \mathcal{U}(0.1, 100)$ ;  $T_e = \mathcal{U}(T_c + 1, T_{sim})$ ;
9:     //  $\mathcal{U}(a, b)$  is the uniform distribution (over interval  $(a, b)$ )
10:    // and  $\mathcal{I}(a, b)$  the uniform integer distribution.
11:     $I^m = \text{ModelEuler}(N, T_c, T_e, \lambda, T_{sim}, h)$ ;
12:    for  $j = 1$  to SIMs do
13:       $I_j^s = \text{SimDiffusion}(N, T_c, T_e, \lambda, T_{sim})$ ;
14:    end for
15:     $\epsilon_i = (I^M - \overline{I^s}) / \overline{I^s}$ ;
16:  end for
17: end function

```

365 represents a realisation of the process, and it depends on the distribution of the contacts, so this must be repeated to assure statistical significance. That is, in each simulation using the same contact rate λ different sequences of contacts are generated. Finally, we obtain a relative error for each simulation ϵ_i .

Summing up, after running these 1000 tests, the obtained mean percentage error with 95%
370 confidence intervals was $\bar{\epsilon} = 0.82\%$ [0.24-1.24]. These results validate the model presented in this paper.

4. Evaluation

In this section we evaluate the diffusion of mobile crowdsensing data and the buffer consumption reduction using the models proposed in Section 3. We consider a communication range of $7.5m$
375 (average Bluetooth range), and a contact rate of $\lambda = 0.001s^{-1}$, that is, a pairwise contact rate of about 3.6 contacts/h. This contact rate depends on several factors such as the size of the area, the

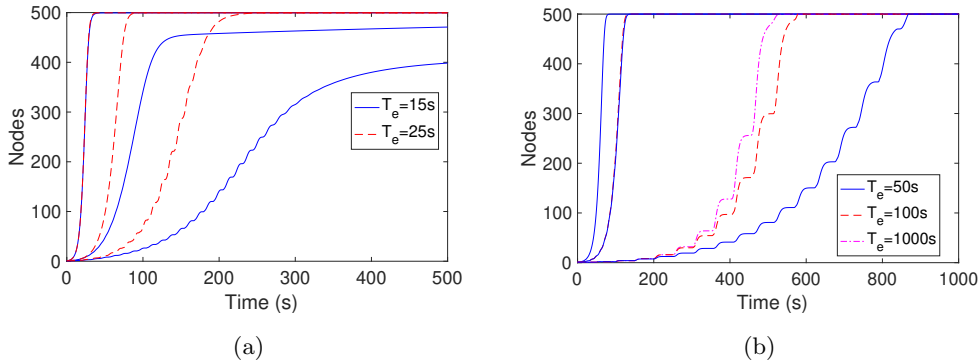


Figure 3: Message diffusion depending on communication and expiration time. a) For tight expiration times (from left to right $T_c = 1, 5, 10s$); b) For larger communication and expiration times (from left to right $T_c = 5, 10, 50s$)

mobility of the nodes, and the communication range among others, but it is independent of the number of nodes (note that λ is the pairwise contact rate).

4.1. Impact of communication and expiration times

380 Firstly, we focus our study on the impact that communication and expiration times have on the diffusion process. We consider an area with $N = 500$ people carrying a mobile device obtaining the whole number of infected nodes depending on time ($I(t)$) for different values of T_e and T_c , as shown in Figure 3. The first plot (Figure 3a) shows the results for tight expiration times ($T_e = 15s, 25s$). Three communication times are considered: 1s corresponding to a very small message; and 5s
385 and 10s corresponding to medium size messages. We can clearly see, as expected, that the longer the communication time is, the slower the diffusion is. Furthermore, we can also evidence the effect of tighter expiration times, which also reduces the diffusion of messages. This is particularly significant for $T_c = 10s, T_e = 15s$ where we can see the effect of this expiration time on the diffusion.

Figure 3b shows the results for longer communication and expiration times. In this case, we
390 additionally consider a longer communication time of $t_c = 50s$, which could correspond to the transmission of large messages. For $t_c = 5s$ and $t_c = 10s$, and setting $t_e = 50s$ there is a significant improvement over the results shown in the previous graph. Nevertheless, increasing this expiration time shows no significant improvement in the diffusion. The results for $T_c = 50s$ are quite different. In these curves, we can clearly see the effect of message communication time, and how the number
395 of infected nodes doubles approximately every 50 seconds ($T_c = 50s$). The impact of T_e can be

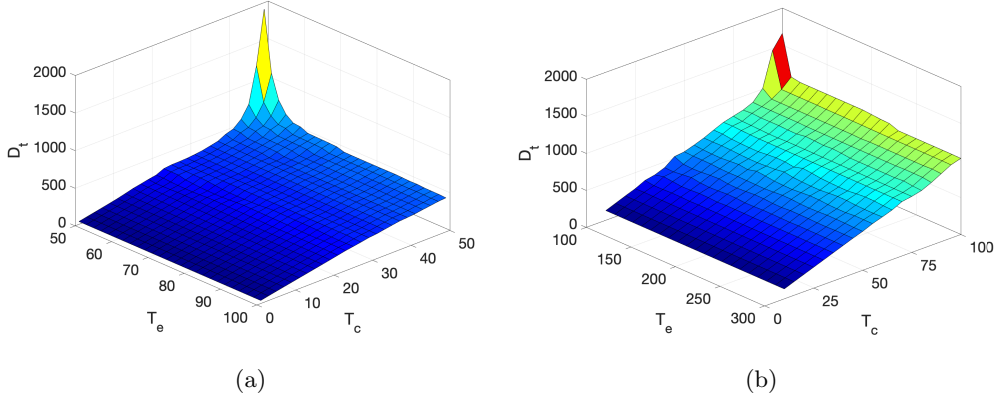


Figure 4: Message delivery time depending on communication and expiration times. a) For short expiration times ($50 < T_e \leq 100$, $1 \leq T_c \leq 50$); b) For larger expiration and communication times ($100 < T_e \leq 300$, $1 \leq T_c \leq 100$)

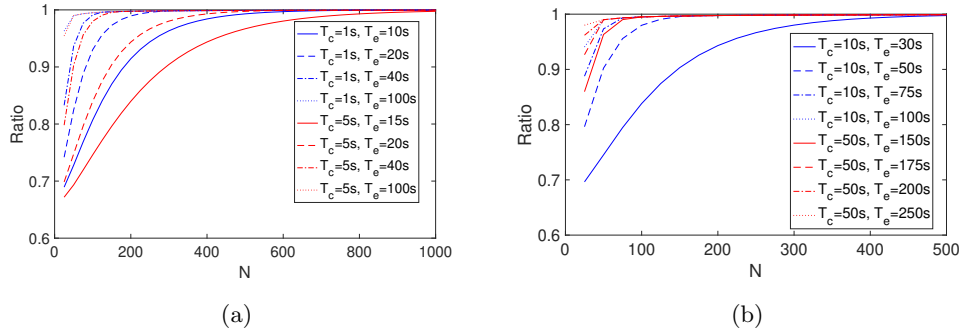


Figure 5: Message diffusion depending on communication and expiration times. a) For short communication and expiration times; b) For longer transmission and expiration times. In this plot, the values for $N > 500$ are omitted since they equal to one.

compared with the results obtained when no message is removed (that is, when $T_e = 1000s$). Relative tight expiration times have a strong impact on the diffusion is these larger messages.

Now, we study the delivery time (D_t), which is a very significant indicator of how fast the diffusion is performed. Figure 4 shows D_t depending on both communication and expiration
 400 time. The first plot (Figure 4a) reveals a great increase in the delivery time when the expiration and communication times are both around 50s. This is also confirmed for larger expiration and communications times values, as shown in Figure 4b for times around 100s. These cases reflect the slowness of the *direct delivery* diffusion. Nevertheless, excluding these extremes cases, the impact of communication time is clearly higher than the impact of expiration time.

405 Finally, Figure 5 shows the impact of the number of nodes on the diffusion of messages. For these experiments, we obtained the number of nodes that have the message $I(t)$ up to time $t=1000s$, using different sets of values of the communication and expiration times. The number of nodes N was varied from 25 to 1000 and the ratio of diffusion was obtained as $I(t)/N$. In general, we can see that the greater the number of nodes, the better the diffusion. Furthermore, the impact of
410 a tighter expiration time is greater for small number of nodes as we can see in Figure 5a for the curves with $T_e = 10s$, $T_e = 20s$ and $T_e = 40s$. This is also confirmed for longer communication times $T_c = 10s$ and $T_e = 20s$ in Figure 5b. This behaviour is reasonable since when the number of nodes is small, the effect of short expiration times is the reduction of the nodes with a copy of the message, therefore slowing the diffusion.

415 We repeated the previous experiments (not shown here) varying the contact rate λ . As expected, as this rate only determines the speed of the diffusion and not the diffusion behaviour, the same conclusions can be extracted from these experiments. Indeed, mathematically, we can see that in the set of equations in Eq. (1) the λ value is a factor that multiplies all the equations of the model.

4.2. Optimising local buffer usage

420 The results we have presented in the previous subsection show that the expiration time has a significant impact on the diffusion, particularly for short values, and for a reduced number of nodes. Nevertheless, it is shown that with only a small increase on the expiration time, the message diffusion improves substantially, also maintaining the local buffer utilisation low.

The method for obtaining the optimal expiration time is based on comparing the diffusion
425 results of our model when the expiration time is not considered (that is, when there are no message removals) with the results obtained when this time is considered. Specifically, considering a given error percentage of E , the goal is to obtain the minimum value of T_e that makes the diffusion $E\%$ slower than the no expiration time case³. This value is obtained iteratively, starting from an initial value of $T_e = T_c$, and incrementing it until the diffusion is at least $E\%$ slower than the one
430 obtained with the epidemic diffusion when no message is removed from local buffers.

The algorithm for obtaining this expiration time is outlined in Algorithm 2. In the line 3, we obtain the vector of infected nodes considering $T_e = \infty$ and then, the delivery time for these N

³For example, if $E = 1\%$, then we obtain T_e for a diffusion 1% slower.

Algorithm 2 Algorithm for obtaining the optimal expiration time.

```

1: function OPTIMAL_EXPIRATION_TIME( $N, T_c, \lambda, T_{sim}, \Delta_e, E$ ) return  $T_e$ 
2:    $h = 0.01$ ;
3:    $I = \text{ModelEuler}(N, T_c, \infty, \lambda, T_{sim}, h)$ ;
4:    $D_t = hi_{i_m}, i_m = \min\{i : I[i] \geq N\}$ ;
5:    $T_e = T_c$ ;
6:   do
7:      $T_e = T_e + \Delta_e$ ;
8:      $I_e = \text{ModelEuler}(N, T_c, T_e, \lambda, D_t, h)$ ;
9:      $\epsilon = (I(i_m) - I_e(i_m)) / I(i_m)$  //  $i_m$ =diffusion time index
10:  while  $\epsilon > (E/100)$ 
11: end function

```

nodes (using Eq. (3)). The loop in lines 6 to 10 iteratively increments the values of T_e while the obtained error (ϵ) is greater than the required one E . Note how the error is obtained in the line 9, where we calculate the relative error between the number of nodes with the message at time D_t , which corresponds to the index i_m .

Using this algorithm, Figure 6 shows the optimal expiration time depending on the number of nodes and communication time. The first two plots (Figure 6a and 6b) show the results for the contact rate already used in the previous experiments ($\lambda = 0.001$) with two error values: 5% and 1%, respectively. In general, we can see that the required expiration times are not lengthy, even for a 1% slower diffusion. As expected, we can see that longer communication times requires a longer expiration time. Nevertheless, it is important to note that for short communication times, the required expiration time can be very short (that is, shorter than 100s in most of the cases). When the number of nodes is small (less than 100), the required expiration time is also long, since fewer nodes can store the message. Nevertheless, when the number of nodes is higher than 200 the impact is very low.

If the contact rate is increased to $\lambda = 0.001$ and using an error of 1%, we can see in Figure 6c that there is a greater dependence on the number of nodes. It is worth to note that when the number of nodes is higher than 250, the diffusion is so fast that the optimal T_e is almost T_c . On the other hand, when the contact rate is reduced ($\lambda = 0.001$), there is a general increase in the optimal T_e , particularly when the number of nodes is below 200.

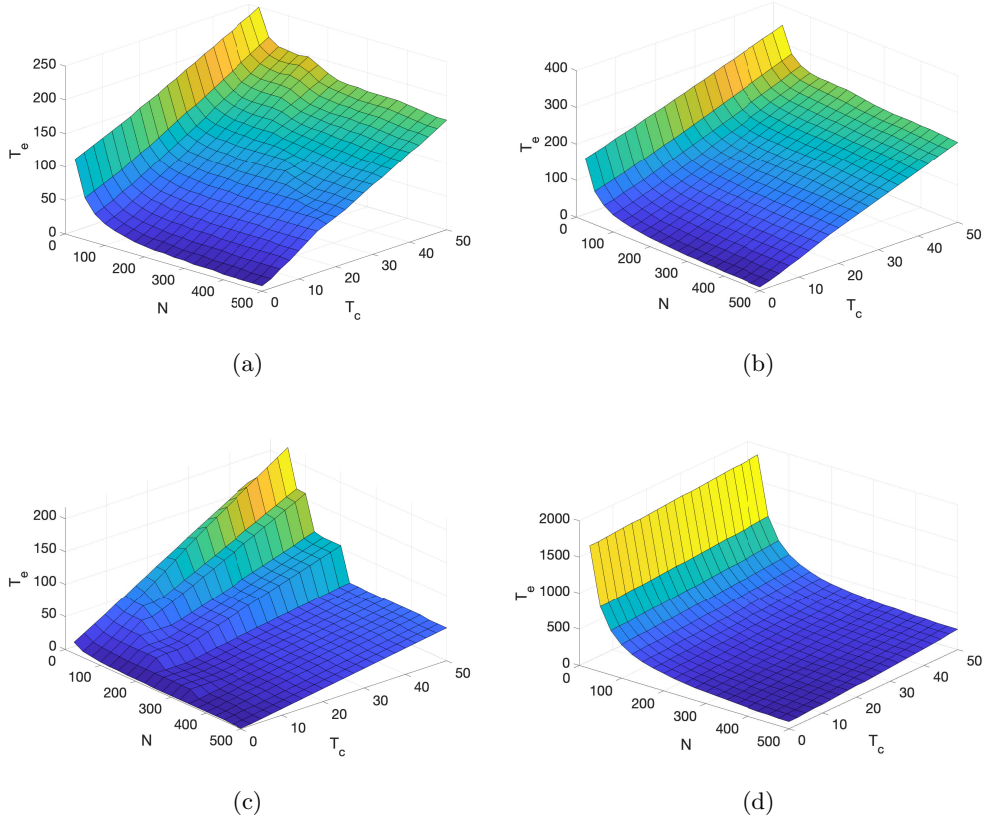


Figure 6: Optimal expiration time depending on expiration time and number of nodes. The number of nodes ranges from 25 to 500 and the communication time from 0 to 50. a) for an error E of 5% and $\lambda = 0.001$; b) for an error of 1% and $\lambda = 0.001$; c) for an error of 1% and higher contact rate ($\lambda = 0.01$); d) for an error of 1% and very small contact rate ($\lambda = 0.0001$);

Now, we evaluate the buffer consumption under a given network workload. This workload tries to model the typical data diffusion of a crowdsensing application where shorter messages are far more frequent than larger ones (that is, it follows a Zipf's law [52]). We consider that nodes can capture three different types of data and spread this information in messages according to the following sizes and frequencies: (1) a short message with simple sensor measurements transmitted every hour ($m_1 = 1\text{KB}$), (2) a medium-size message, which can contain more complex sensor measurements such as audio or small photos, transmitted every 18 hours ($m_2 = 1\text{MB}$), and (3) a large message, which can contain very complex sensed data such as a video or a high-resolution picture, transmitted every 96 hours ($m_3 = 10\text{MB}$). These frequencies and sizes are based on the statistics of *Whatsapp* message usage from [53]. Then, considering a Bluetooth communications

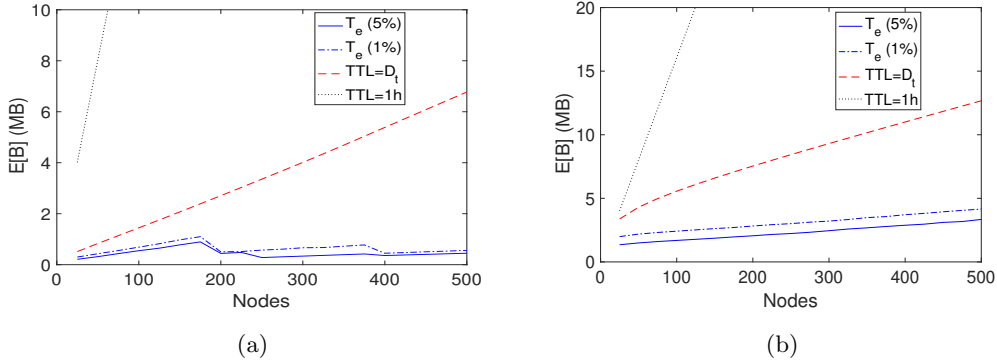


Figure 7: Average buffer utilisation depending on the number of nodes ($E[B]$ in MB) a) For $\lambda = 0.001$; b) For $\lambda = 0.0001$;

Table 2: Workload used for evaluating the network resources consumption. U is the number of active users.

Size	Period	T_c	β
1KB	1h	0.1039s	$0.2778 \times 10^{-3} \cdot U$ msg/s
1MB	18h	4.0946s	$0.0154 \times 10^{-3} \cdot U$ msg/s
10MB	96h	40.0458s	$0.0029 \times 10^{-3} \cdot U$ msg/s

channel with an average bandwidth (Bw) of 2.1Mb/s, and setup time $T_s = 0.1s$, we can work out the communication time as $T_c = 8 \cdot m_i / Bw + T_s$. Regarding the whole network message generation (that is, the workload), we consider a set of active nodes U , which are the ones that capture data and generate the messages following the previous frequencies. Then, the arrival process is multiplied by the number of active nodes U , which in our experiments is N . Summing up, the workload is detailed in Table 2.

We compared three different temporal message removal approaches: 1) the optimal T_e for 5% and 1% error, obtained using Algorithm 2; 2) a fixed TTL of one hour; and, 3) a TTL obtained as the expected diffusion time (D_t) of the message using our model (that is, the message lives the necessary time to be completely spread). The results depending on the number of nodes are shown in Figure 7, where the buffer usage is obtained using expression $E[B]$ in Eq. (8).

First, for $\lambda = 0.001$ (Figure 7a), we can see that using a fixed TTL approach the buffer usage increases dramatically, reaching a peak value of 200MB for $N=500$ (not shown in the plot due that is extremely outside of the figure scale). These results clearly show that this unconstrained

epidemic diffusion can overload the local buffer when the number of nodes is high, so this is not a suitable approach for removing local messages when the message workload is high, and therefore, local buffer usage should be limited using other methods. Focusing now on the other approaches, we can see that the buffer utilisation for optimal T_e is almost constant with N , with an average
480 buffer usage of 0.6MB. Instead, using $TTL = D_t$ we can see a steady increase of buffer usage almost linear with the number of nodes.

The results for a lower contact rate $\lambda = 0.0001$ in Figure 7b show an increase on buffer utilisation, but with a similar pattern. There is a slight increase of memory usage $E[B]$ with N , due mainly to the storage requirements of large messages that require more time to be transmitted,
485 and therefore their optimal T_e are greater.

In conclusion, the proposed approach reduces the local consumption of the buffer drastically. The practical implementation of the optimal T_e is not complex. For example, when a new message is created in order to be broadcast, the sender node calculates the optimal T_e using Algorithm 2, which can be locally implemented. The obtained T_e value can be included in the message diffusion,
490 so the receiver nodes know the message expiration time. Even if local nodes do not have an exact value of the contact rate, they can make an estimation of these parameters using, for example, a linear regression method as detailed in [54].

5. Evaluation of a Realistic Mobile Crowdsensing Scenario

In this section, we evaluate the efficiency of our diffusion method in a realistic simulated sce-
495 nario, where students in a university campus collect and share information using their mobile phones. The idea behind is that students use their mobile phones to sense and monitor the status and availability of university resources, such as the library, sports facilities, canteen, and so on. For example, students could share the current noise and lighting levels at the library, or even photos or videos about how nice the food looks today.

This scenario is simulated using an OppNet simulator and real movement traces of mobile
500 users. Note that, although analytical models are a handy tool to provide a thorough analysis of the dynamics and diffusion of a message, using a real mobile trace with a simulator is necessary to validate the performance of the proposed approaches [8]. Basically, OppNet simulations are contact-based, that is, the simulator reads a trace with the positions of the nodes, generates some

505 workload, detects the nodes that are in contact (that is, within a given range), and performs the exchange of messages. This is the approach of well-known OppNet simulators such as the ONE [55] and OMNeT++ [56]. Based on these ideas, we have implemented in Matlab a custom contact-based simulator with the aim of evaluating the diffusion of our proposed diffusion model considering the expiration time⁴. This simulator can use the same mobility traces as the ONE
510 simulator. Therefore, in order to validate our simulator, we checked the results of transmitting one message using several scenarios, and compared the obtained results with the ones obtained using the ONE simulator (using the epidemic simulation), achieving the same results. Particularly, we used the NCCU trace [26], which comes from an experiment at the National Chengchi University (NCCU) campus, and where GPS position data were collected during two weeks (336 hours) using
515 an Android app installed in the smartphones of 115 students in an area of 3.764km x 3.420km. The total number of contacts was of 81.112, with an average contact rate of 4.8×10^{-6} c/s, and a mean contact duration of 244s.

The simulation aims to evaluate how periodical sensed data is broadcast among the nodes (students) of the NCCU trace, obtaining the buffer utilisation, ratio of diffusion and bytes transmitted
520 (network load) for different local buffer management strategies. In this scenario, we consider a number of 20 active users ($U = 20$), which are the ones that actively capture data and generate messages. The pattern of the messages sent are the same than in the previous analytical experiments (Table 2). Each message was generated randomly following a Poisson process with arrival intensity β , and message sizes also randomly generated following a normal distribution with
525 $\mu = \sigma = size$. Given this message pattern generation, the number of different messages generated for each simulation is around 7500 ($L \approx 7500$).

The first experiment obtains the temporal dynamics of buffer utilisation, diffusion and bytes transmitted from running simulations using the same workload and three different buffer approaches: 1) Our Optimal T_e expiration message removal; 2) no buffer limit with TTL=3600h,
530 which is considered to be the approach that obtains the best diffusion, and 3) buffer limit set to 35MB with TTL=3600h. In this approach, when a new message is received and there is not enough free buffer space, the required oldest messages are removed to allocate space for the new message.

⁴In case the paper was published, the code of this simulator and also the analytical models would be freely available for the research community on the GitHub site of our research group (<https://github.com/GRCDEV/OppMCS/>). This implementation, which is provided as a proof of concept, can be easily re-coded in other languages

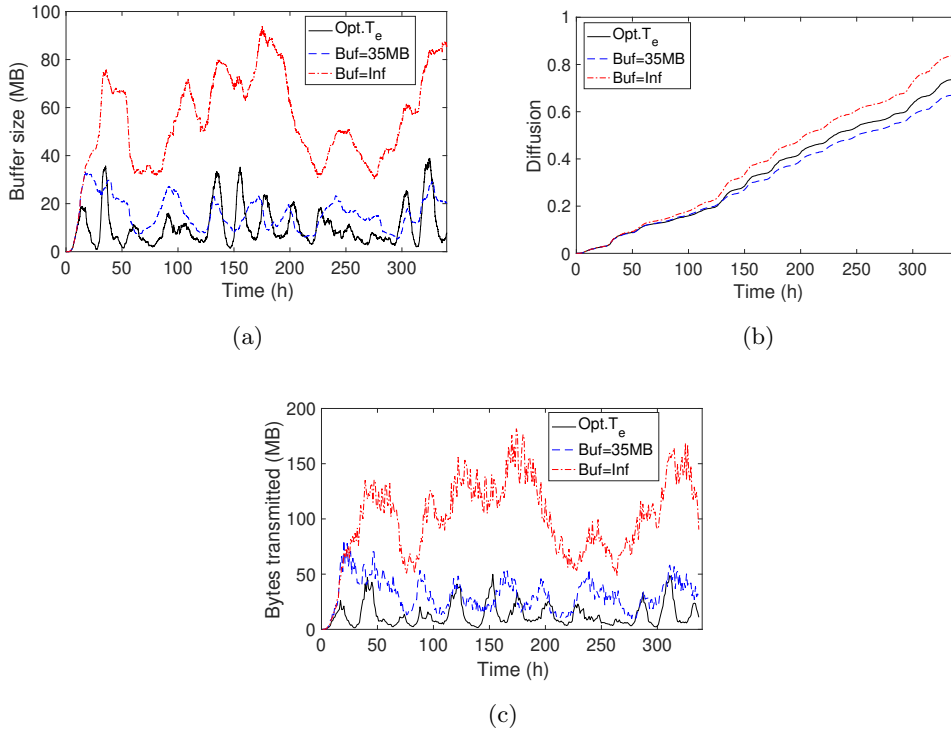


Figure 8: Results of simulating the diffusion of sensed data in the NCCU scenario. a) Buffer occupation throughout time; b) Overall diffusion ratio of all messages; c) Average number of bytes transmitted by each mobile using a sampling period of 1 hour.

The buffer limit is set to a value equivalent to the mean buffer utilisation of the Optimal T_e approach in order to compare approaches with similar buffer consumption. The idea is to compare three different temporal buffer management policies. Other alternative management strategies when the buffer is limited, such as dropping the largest message or even randomly dropping of messages, were evaluated with no significant difference from the results presented in the paper.

Figure 8a shows the average local buffer occupation along time. Firstly, we can see the impact of the largest messages in the buffers when these messages are created (the utilisation peaks). Secondly, the optimal and limited buffer approach has a similar use of buffers, whereas the diffusion with no buffer limit increases substantially the buffer use. This aspect clearly determines the overall ratio of diffusion as shown in Figure 8b. This ratio is obtained as $R_{msgs}/(N \cdot L)$, where R_{msgs} is the total number of messages correctly delivered. As expected, when there is no buffer limitation, the diffusion is improved. Nevertheless, we can see that our approach outperforms the limited buffer

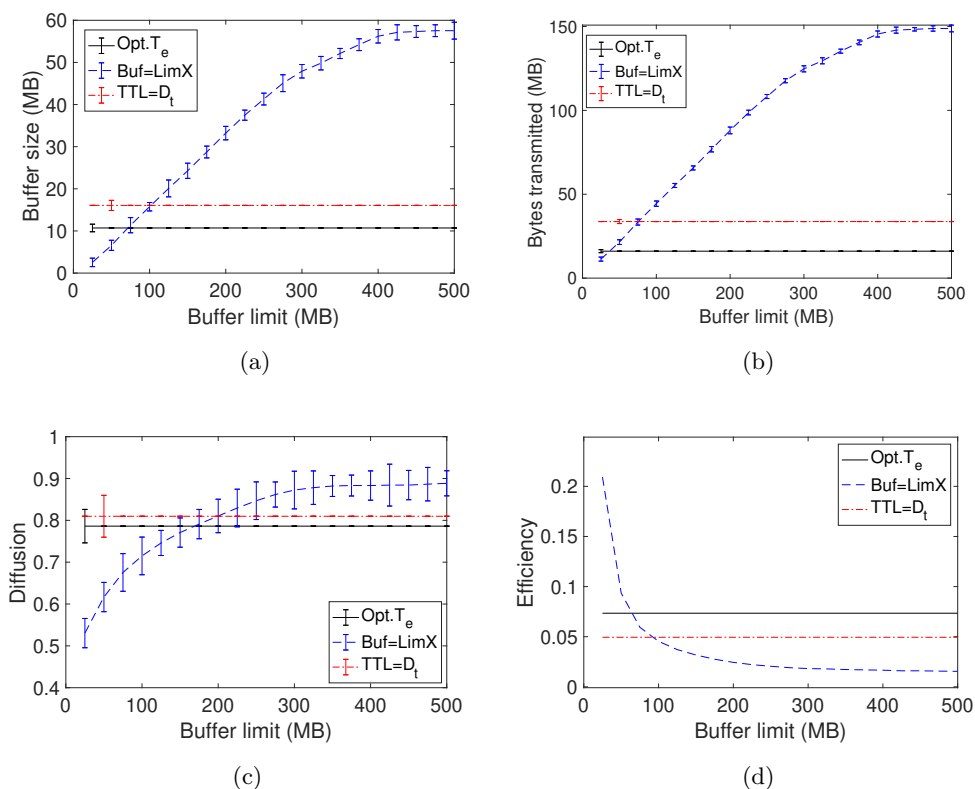


Figure 9: Extended evaluation depending on local buffer size. Note that only the curve for limited buffer depends on this buffer limit. The other lines, Optimal T_e and D_t , which have no buffer limit, are plotted as a straight line. Graphs a), b) and c) show 95% confidence intervals. Note that for the Opt. T_e and TTL= D_t curves, their confidence intervals are only shown once at the left. a) Average Buffer occupation; b) Average bytes transmitted per hour; c) Average diffusion ratio of all messages; d) Efficiency as the quotient between diffusion and average buffer size.

545 approach using the same local buffer. Finally, Figure 8c shows the average of the number of bytes transmitted by each mobile considering periods of one hour (for example, a value of 11.91MB at hour 50 for the optimal T_e means that each mobile has transmitted an average of 11.91MB during this hour). Thus, this figure clearly represents how local communication resources are used. As expected, we can see that this transmission follows the pattern of the buffer size figurename 8a.

550 Furthermore, we can see that the number of bytes transmitted by our proposal is significantly lower than the other two approaches, particularly when it is compared to the epidemic with no buffer constraint. Summing up, these results confirm the expected reduction on both the buffer and network utilisation.

Finally, we performed an extensive evaluation depending on the buffer size. Specifically, Figure 9

555 shows the average values for the buffer size, bytes transmitted and diffusion of messages, which is aimed at evaluating the efficiency of the different buffer approaches. The values obtained for these plots are the result of repeating ten times the simulations with different random generated workloads⁵ in order to obtain average and confident values. Three approaches were evaluated: 1) our Optimal T_e method, 2) TTL set to D_t and 3) buffer limited from 25 to 500MB. Note that this
 560 upper buffer size limit obtains results equivalent to the case when there is no buffer limitation, and thus, it is not required to evaluate greater buffer limits.

Figure 9a shows the average buffer sizes of all nodes. Note that only the curve with a buffer limit (Buf=LimX) depends on the values of the buffer limit. Particularly, this curve shows a sharp increase usage of the buffer when the buffer limit ranges from 25 to 350MB, levelling out
 565 for values greater than 400MB, denoting that practically no message is dropped from the buffers. The buffer usage of the other two approaches is shown as a straight line. Figure 9b shows the average number of transmitted bytes considering per mobile using a sampling period of one hour (that is, the equivalent to the previous Figure 8c). The results of this figure confirm an important reduction on the network usage for our approach when compared with the other approaches. The
 570 average diffusion of all messages is shown in Figure 9c where we can see that, for obtaining a diffusion equivalent to the one obtained with the Optimal T_e approach (0.79), for the limited buffer approach it is required a limit of at least 200MB, which supposes an average size buffer of around 30MB, that is three times greater.

Furthermore, in order to measure the efficiency of the different approaches, we obtained the
 575 quotient between the diffusion and the average buffer size occupation. The results are shown in Figure 9d, where we can see that excluding the values for smaller buffer limits, the Optimal T_e is the most efficient approach, obtaining a good diffusion for a reduced local buffer usage.

Summing up, the experiments described in this section has confirmed the results obtained using the model: we can obtain a data diffusion similar to the epidemic diffusion with a reduced local
 580 buffer and network usage. In addition, this reduction in the number of bytes transmitted can have a significant impact in terms of energy consumption. These issues are particularly important for mobile crowdsensing applications since they can generate a lot of data, which can overload either the local buffers and the network, and therefore, reduce the diffusion of the sensed data and also

⁵Note that the mobility is always the same, which is defined by the trace.

drain mobile's batteries.

585 **6. Conclusions**

In this paper, we presented a solution, based on adapting the expiration time of messages, to reduce local buffer consumption for the diffusion of Mobile Crowdsensing data. We demonstrated the necessity to determine the optimal expiration time in order to have a good diffusion, which we obtained using an analytical model of this diffusion. [The results showed that we could obtain a 1% slower diffusion with a very significant reduction of buffer usage and network utilisation.](#)
590 [Furthermore, when compared to other temporal approaches, the buffer utilisation is almost linear with the number of nodes.](#)

In conclusion, we have shown that the combination of Mobile Crowdsensing, Opportunistic Networking data diffusion, and local expiration time, can avoid the typical buffer overload when
595 using epidemic diffusion, obtaining nearly the same diffusion effectiveness. Furthermore, being the buffer and network utilisation almost linear with the number of nodes, makes our approach very scalable without the need of using cloud-based infrastructure for storing and spreading the sensed data; consequently, it can reduce the cost and energy consumption of crowdsensing applications.

[As future work, we plan to consider not only the temporal aspect of the messages but also social aspects such as message popularity or social ties between nodes to improve the local buffer management and the message diffusion.](#) Furthermore, some parts of this protocol are planned to be implemented in our GRChat Messaging Application [57] as a proof of concept.
600

Acknowledgments

This work was partially supported by the *Ministerio de Ciencia, Innovación y Universidades*,
605 Spain, under Grant RTI2018-096384-B-I00. Also, this work has been partially performed in the framework of the European Union's Horizon 2020 project 5G-CARMEN co-funded by the EU under grant agreement No. 825012. The views expressed are those of the authors and do not necessarily represent the project. The Commission is not liable for any use that may be made of any of the information contained therein.

610 References

- [1] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A. Campbell, A survey of mobile phone sensing, *Communications Magazine*, IEEE 48 (9) (2010) 140–150.
- [2] W. Zamora, C. T. Calafate, J.-C. Cano, P. Manzoni, A survey on smartphone-based crowdsensing solutions, *Mobile Information Systems 2016* (2016) 26.
- 615 [3] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, P. Bouvry, A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities, *IEEE Communications Surveys Tutorials* 21 (3) (2019) 2419–2465.
- [4] C. Song, M. Liu, X. Dai, Remote cloud or local crowd: Communicating and sharing the crowdsensing data, in: *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, 2015, pp. 293–297.
- 620 [5] J. G. P. Rodrigues, A. Aguiar, C. Queirós, Opportunistic mobile crowdsensing for gathering mobility information: Lessons learned, in: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 1654–1660.
- [6] L. Pelusi, A. Passarella, M. Conti, Opportunistic networking: data forwarding in disconnected mobile ad hoc networks, *Communications Magazine*, IEEE 44 (11) (2006) 134–141.
- 625 [7] S. Trifunovic, S. T. Kouyoumdjieva, B. Distl, L. Pajevic, G. Karlsson, B. Plattner, A decade of research in opportunistic networks: Challenges, relevance, and future directions, *IEEE Communications Magazine* 55 (1) (2017) 168–173.
- [8] J. Dede, A. Förster, E. Hernández-Orallo, J. Herrera-Tapia, K. Kuladinithi, V. Kuppusamy, P. Manzoni, A. bin Muslim, A. Udugama, Z. Vatandas, Simulating opportunistic networks: Survey and future directions, *IEEE*
- 630 *Communications Surveys and Tutorials* 20 (5) (2018) 1547–1573.
- [9] A. Udugama, J. Dede, A. Förster, V. Kuppusamy, K. Kuladinithi, A. Timm-Giel, Z. Vatandas, My smartphone tattles: Considering popularity of messages in opportunistic data dissemination, *Future Internet* 11 (2) (2019).
- [10] A. Vahdat, D. Becker, Epidemic routing for partially-connected ad hoc networks, Technical report number cs-200006, Duke University (2000).
- 635 [11] J. Herrera-Tapia, E. Hernández-Orallo, A. Tomás, P. Manzoni, C. T. Calafate, J. Cano, Selecting the optimal buffer management for opportunistic networks both in pedestrian and vehicular contexts, in: *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*, 2017, pp. 395–400.
- [12] R. Groenevelt, P. Nain, G. Koole, The message delay in mobile ad hoc networks, *Performance Evaluation* 62 (2005) 210–228.
- 640 [13] A. Lindgren, A. Doria, O. Schelén, Probabilistic routing in intermittently connected networks, *SIGMOBILE Mob. Comput. Commun. Rev.* 7 (3) (2003) 19–20.
- [14] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Spray and Wait: An efficient routing scheme for intermittently connected mobile networks, in: *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking, WDTN '05*, ACM, New York, NY, USA, 2005, pp. 252–259.
- 645 [15] T. Spyropoulos, K. Psounis, C. Raghavendra, Efficient routing in intermittently connected mobile networks: The multiple-copy case, *Networking*, *IEEE/ACM Transactions on* 16 (1) (2008) 77–90.

- [16] C. Borrego, J. Borrell, S. Robles, Efficient broadcast in opportunistic networks using optimal stopping theory, *Ad Hoc Networks* 88 (2019) 5 – 17.
- [17] M. Desta, E. Hyytia, J. Ott, 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*, 2013, pp. 147–154.
- [18] E. Hernández-Orallo, M. Murillo-Arcila, C. T. Calafate, J. C. Cano, J. A. Conejero, P. Manzoni, Analytical evaluation of the performance of contact-based messaging applications, *Computer Networks* 111 (2016) 45 – 54.
- [19] E. Hernández-Orallo, M. Murillo-Arcila, J. C. Cano, C. T. Calafate, J. A. Conejero, P. Manzoni, An analytical model based on population processes to characterize data dissemination in 5G opportunistic networks, *IEEE Access* 6 (2018) 1603–1615.
- [20] E. Hernández-Orallo, C. Borrego, C. Calafate, J.-C. Cano, Using local expiration timers to reduce buffer utilisation when using epidemic diffusion, in: *IEEE (Ed.), Global Telecommunications Conference (GLOBECOM 2019)*, 2019.
- [21] T. G. Kurtz, *Approximation of Population Processes*, SIAM, 1981.
- [22] Z. J. Haas, T. Small, A new networking model for biological applications of ad hoc sensor networks, *Networking, IEEE/ACM Transactions on* 14 (1) (2006) 27–40.
- [23] X. Zhang, G. Neglia, J. Kurose, D. Towsley, Performance modeling of epidemic routing, *Computer Networks* 51 (10) (2007) 2867 – 2891.
- [24] C. S. De Abreu, R. M. Salles, Modeling message diffusion in epidemical DTN, *Ad Hoc Networks* 16 (2014) 197–209.
- [25] Q. Xu, Z. Su, K. Zhang, P. Ren, X. S. Shen, Epidemic information dissemination in mobile social networks with opportunistic links, *Emerging Topics in Computing, IEEE Transactions on* 3 (3) (2015) 399–409.
- [26] T.-C. Tsai, H.-H. Chan, NCCU trace: social-network-aware mobility trace, *Communications Magazine, IEEE* 53 (10) (2015) 144–149.
- [27] Y. Yao, L. T. Yang, N. N. Xiong, Anonymity-based privacy-preserving data reporting for participatory sensing, *IEEE Internet of Things Journal* 2 (5) (2015) 381–390.
- [28] C. Greco, M. Kieffer, C. Adjih, NeCoRPIA: Network coding with random packet-index assignment for mobile crowdsensing, in: *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 6338–6344.
- [29] P. Nguyen, K. Nahrstedt, Context-aware crowd-sensing in opportunistic mobile social networks, in: *2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, 2015, pp. 477–478.
- [30] D. Zhao, H. Ma, S. Tang, X. Li, COUPON: A cooperative framework for building sensing maps in mobile opportunistic networks, *IEEE Transactions on Parallel and Distributed Systems* 26 (2) (2015) 392–402.
- [31] X. Wu, K. N. Brown, C. J. Sreenan, Analysis of smartphone user mobility traces for opportunistic data collection in wireless sensor networks, *Pervasive and Mobile Computing* 9 (6) (2013) 881 – 891.
- [32] T. E. Amah, M. Kamat, K. A. Bakar, S. O. A. Rahman, M. H. Mohammed, A. M. Abali, W. Moreira, A. Oliveira, Collecting sensed data with opportunistic networks: The case of contact information overhead, *Information* 8 (3) (2017).
- [33] T. E. Amah, M. Kamat, K. A. Bakar, W. Moreira, A. Oliveira, M. A. Batista, Preparing opportunistic networks

- 685 for smart cities: Collecting sensed data with minimal knowledge, *Journal of Parallel and Distributed Computing* (2019).
- [34] L. Pajevic, V. Fodor, G. Karlsson, Ensuring persistent content in opportunistic networks via stochastic stability analysis, *ACM Trans. Model. Perform. Eval. Comput. Syst.* 3 (4) (2018) 16:1–16:23.
- [35] V. Palma, E. Mammi, A. M. Vegni, A. Neri, A fountain codes-based data dissemination technique in vehicular
690 ad-hoc networks, in: 2011 11th International Conference on ITS Telecommunications, 2011, pp. 750–755.
- [36] A. Mostafa, A. M. Vegni, R. Singoria, T. Oliveira, T. D. C. Little, D. P. Agrawal, A V2X-based approach for reduction of delay propagation in vehicular ad-hoc networks, in: 2011 11th International Conference on ITS Telecommunications, 2011, pp. 756–761.
- [37] A. M. Vegni, T. D. C. Little, A message propagation model for hybrid vehicular communication protocols, in:
695 2010 7th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP 2010), 2010, pp. 382–386.
- [38] M. Khabbaz, C. Assi, W. Fawaz, Disruption-tolerant networking: A comprehensive survey on recent developments and persisting challenges, *Communications Surveys Tutorials*, IEEE 14 (2) (2012) 607–640.
- [39] E. Hernandez-Orallo, M. Serrat Olmos, J.-C. Cano, C. Calafate, P. Manzoni, CoCoWa: A collaborative contact-
700 based watchdog for detecting selfish nodes, *Mobile Computing*, IEEE Transactions on 14 (6) (2015) 1162–1175.
- [40] M. Karaliopoulos, Assessing the vulnerability of DTN data relaying schemes to node selfishness, *Communications Letters*, IEEE 13 (12) (2009) 923–925.
- [41] J. Whitbeck, V. Conan, M. Dias de Amorim, Performance of opportunistic epidemic routing on edge-markovian dynamic graphs, *Communications*, IEEE Transactions on 59 (5) (2011) 1259–1263.
- 705 [42] L. J. S. Allen, *Mathematical Epidemiology: Lecture Notes in Mathematics*, Vol. 1945, Springer Verlag, 2008, Ch. An Introduction to Stochastic Epidemic Models, pp. 81–130.
- [43] G. M. de Souza Dias, J. F. de Rezende, R. M. Salles, Mathematical modeling of delivery delay for multi-copy opportunistic networks with heterogeneous pairwise encounter rates, *Information Sciences* 475 (2019) 142–160.
- [44] J. Su, A. Chin, A. Popivanova, A. Goel, E. de Lara, User mobility for opportunistic ad-hoc networking, in:
710 *Mobile Computing Systems and Applications*, 2004. WMCSA 2004. Sixth IEEE Workshop on, 2004, pp. 41–50.
- [45] Z. Feng, 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, IPDPSW '12*, IEEE Computer Society, Washington, DC, USA, 2012, pp. 1484–1493.
- [46] J. Herrera-Tapia, E. Hernández-Orallo, A. Tomás, P. Manzoni, C. Tavares Calafate, J.-C. Cano, Friendly-
715 sharing: Improving the performance of city sensing through contact-based messaging applications, *Sensors* 16 (9) (2016).
- [47] Y. Zhang, 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, MobiHoc '09*, ACM, New York, NY, USA, 2009, pp. 345–346.
- 720 [48] A. Goundan, E. Coe, C. Raghavendra, Efficient broadcasting in delay tolerant networks, in: *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, 2008, pp. 1–5.
- [49] W. Gao, G. Cao, User-centric data dissemination in disruption tolerant networks, in: 2011 *Proceedings IEEE*

INFOCOM, 2011, pp. 3119–3127.

- 725 [50] E. Hernandez-Orallo, J. Herrera-Tapia, J.-C. Cano, C. Calafate, P. Manzoni, Evaluating the impact of data transfer time in contact-based messaging applications, *Communications Letters, IEEE* 19 (10) (2015) 1814–1817.
- [51] M. Grossglauser, D. N. C. Tse, Mobility increases the capacity of ad hoc wireless networks, *IEEE/ACM Transactions on Networking* 10 (4) (2002) 477–486.
- [52] L. A. Adamic, B. A. Huberman, Zipf’s law and the internet, *Glottometrics* 3 (1) (2002) 143–150.
- 730 [53] <http://www.statista.com/chart/1938/monthly-whatsapp-usage-per-user>, An Average WhatsApp User Sends Messages per Month (15/09/2015).
- [54] C. Borrego, E. Hernández-Orallo, N. Magaia, General and mixed linear regressions to estimate inter-contact times and contact duration in opportunistic networks, *Ad Hoc Networks* 93 (2019) 101927.
- [55] A. Keränen, J. Ott, T. Kärkkäinen, The one simulator for dtn protocol evaluation, in: *Proceedings of SIMU-Tools’09*, 2009, pp. 55:1–55:10.
- 735 [56] Z. Zhao, B. Mosler, T. Braun, Performance evaluation of opportunistic routing protocols: A framework-based approach using omnet++, in: *Proceedings of the 7th Latin American Networking Conference, LANC ’12*, ACM, New York, NY, USA, 2012, pp. 28–35.
- [57] E. Hernández-Orallo, D. Fernández-Delegido, J. Herrera-Tapia, J. Cano, C. Calafate, P. Manzoni, Grchat: A contact-based messaging application for the evaluation of information diffusion, in: *Proceedings of the 6th International Conference on Advanced Communications and Computation, INFOCOMP*, 2016.
- 740