

Document downloaded from:

<http://hdl.handle.net/10251/63685>

This paper must be cited as:

Hernández Orallo, E.; Herrera Tapia, J.; Cano Escribá, JC.; Tavares De Araujo Cesariny Calafate, CM.; Manzoni, P. (2015). Evaluating the Impact of Data Transfer Time in Contact-Based Messaging Applications. *IEEE Communications Letters*. 19(10):1814-1817. doi:10.1109/LCOMM.2015.2472407.



The final publication is available at

<http://dx.doi.org/10.1109/LCOMM.2015.2472407>

Copyright Institute of Electrical and Electronics Engineers (IEEE)

Additional Information

“© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# Evaluating the Impact of Data Transfer Time in Contact-based Messaging Applications

Enrique Hernández-Orallo, Jorge Herrera-Tapia, Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni

**Abstract**—In this paper we propose an analytical model based on Delay Differential Equations (DDEs) to evaluate the diffusion of messages in groups taking into account the transmission time of the messages. This model was validated through simulation studies using the ONE simulator.

Evaluation results show that considering the impact of data transfer time is of utmost importance, as when message size increases (for example when transmitting short videos), the diffusion is bounded by this transmission time, and the result is that the diffusion time increases slightly when the number of nodes increases (as opposed to the always decreasing diffusion time of the epidemic diffusion with no delay).

**Index Terms**—Opportunistic networks, Contact-based Messaging, Performance Evaluation, Epidemic diffusion

## I. INTRODUCTION

Messaging applications are used to instantly share and/or retrieve multimedia data, such as text, images, audio and video through mobile devices. Nevertheless, the problem of these messaging applications is twofold. First, a stable infrastructure is needed in order to provide this communication. Second, as messages are transmitted (and stored) in public networks, privacy is threatened. Furthermore, the messages can be traced in order to obtain its origin and destination, so the privacy of communication can be seriously affected. A solution to these problems is to use opportunistic networks. Instead of using the established Internet infrastructure, the communication in opportunistic networks takes place upon the establishment of opportunistic contacts among mobile nodes using direct communication (i.e. Bluetooth or WiFi Direct). By relying on well designed security mechanisms, opportunistic networks can increase the confidentiality and privacy of communications. Note, that direct communications are more robust to remote scanning and inspection.

The application of opportunistic networking to mobile devices such as smartphones has not been fully exploited, as communication is based in well established 3G/4G or WiFi infrastructure and cloud services. For example, instant messaging applications (such as Whatsapp, Vibe, etc.) use established local Internet connections in order to access to the centralised services responsible for message management. Their main advantage is supporting seamless communication between devices regardless of their location.

Based on the concept of opportunistic networks, new contact-based messaging applications have recently been developed, such as FireChat (developed by Open Garden -

<http://opengarden.com/FireChat/>). FireChat, a messaging application meant for festivals, that became popular in 2014 in Iraq following government restrictions on Internet use, and thereafter during the Hong Kong protests. There are other options, such as Briar (see <https://briarproject.org>) and Netsukuku (<http://netsukuku.freaknet.org>), and the so called Floating Content approach [1].

In this paper, we study the performance of these opportunistic contact-based messaging strategies in city squares or meeting places. A common approach is to combine a network simulation tool with realistic mobility traces. Nevertheless, simulation can be very time consuming and restricted to the limited scenarios of available mobility traces. In order to avoid these drawbacks, analytical models can provide a fast and broader performance evaluation. Two classes of analytical models have been developed: Markovian models [2]–[5] and deterministic models based on ODEs [3], [6]. Nevertheless, these models cannot evaluate the delay incurred on message transmission. Thus, we introduce a new deterministic model, based on Delay Differential Equations (DDEs). DDEs are similar to ODEs, but their evolution involves past values of the state variable. DDEs have been used for modelling population dynamics in many disciplines such as biology, ecology, epidemics [7], and network protocol analysis [8].

## II. A CONTACT-BASED MESSAGING APPROACH

The basis of contact-based messaging is to establish a short-range communication directly between mobile devices, and to store the messages in these devices in order to achieve a full dissemination of such messages. No messages are sent or stored in servers (that is, no cloud service is used). Instead, all information is stored on the mobile devices in a given area. Message spreading is based on epidemic diffusion, a concept similar to the spreading of infectious diseases, when an infected node (the one that has a message) contacts another node to infect it (transmit the message). Epidemic routing obtains the minimum delivery delay at the expense of increased usage of local buffer and increased number of transmissions. There are different variations to this diffusion scheme (that is, the infection process), that attempt to reduce resource utilisation e.g, 2-hop forwarding [2], probabilistic forwarding [9] and multiple copy [4].

The diffusion of messages among users is typically based on groups. Users can join (and leave) a group, and their members receive the messages that are sent to the group. The mechanisms for group management are, anyway, outside the scope of this paper. The dissemination scheme works as follows. Mobile devices have a messaging application that notifies and shows to the user the received messages for the

The authors are with the *Departamento de Informática de Sistemas y Computadores. Universitat Politècnica de València. Spain.* (contact email: ehernandez@disca.upv.es). This work was partially supported by *Ministerio de Economía y Competitividad, Spain* (Grant TEC2014-52690-R).

subscribed groups. The application is also cooperative: it must store the received messages and performs the diffusion of such messages to other nearby nodes. Each node has a limited buffer where it can store the messages it receives from other nodes. When two nodes establish a pair-wise connection, they exchange the messages they have in their buffer, and check whether some of the newly received messages are suitable for notification to the user device. In order to complete this exchange the devices are forced to stop. That is, the owners of the mobile devices control this exchange and wait until the message transmission is completed. This is a commonly used scheme in several existing short-range mobile protocols such as *Apple iOS Airdrop* and *Google Android Copleseence*. Finally, all nodes in the network (not only the members of the group) collaborate in storing and forwarding messages.

### III. PERFORMANCE MODEL

In this section we propose an analytical performance model to evaluate the dissemination of messages considering the data transmission time. The performance model presented here is based on *biological* epidemic models [10]. The number of nodes (population)  $N$  remains constant, and nodes move freely in a given area with a given contact rate between pairs  $\lambda$ . We assume short-range wireless communication (for example, Bluetooth), so network congestion and interferences do not have a strong impact. In the basic epidemic model [6], when a node carrying the message (a *infected* node) contacts with another node that does not have the message (a *susceptible* node), it transmits this message immediately, without further delay. From that moment on, both nodes carry the message. As we will prove, this model is valid only when the transmission time is low. Following the epidemic model notation, the population is divided in two groups: the *infected* nodes ( $I$ ) and the *susceptible* nodes ( $S$ ), so population remains constant:  $N = I + S$ .

We extend this basic model by considering the transmission time of the messages and the existence of groups with  $N_g$  members ( $N_g \leq N$ ). The message communication time can be expressed as  $T_c = T_s + m/Bw$ , and includes two components: a fixed set-up time  $T_s$  for two nodes to establish a connection; and a transmission time  $m/Bw$ , that depends on the message size ( $m$ ) and the available bandwidth ( $Bw$ ). In this model, when an infected node contacts a susceptible node, both nodes establish a connection and transmits a message. During the communication time  $T_c$ , the nodes involved cannot infect other nodes, so a new class of nodes is introduced: the communicating ( $C$ ) nodes. Thus, when a contact occurs,  $C$  is increased by two, one coming from the class of infected nodes, and another one from the class of susceptible nodes. When transmission ends, these two nodes are moved to the infected nodes class. Note that, in this model, the infected class is defined as the nodes that can infect other nodes (represented by letter  $J$ ). Thus, the dynamics of this system are expressed through the following DDEs:

$$\begin{aligned} S'(t) &= -\lambda S(t)J(t) \\ C'(t) &= 2\lambda S(t)J(t) - 2\lambda S(t - T_c)J(t - T_c) \\ J'(t) &= 2\lambda S(t - T_c)J(t - T_c) - \lambda S(t)J(t) \end{aligned} \quad (1)$$

with  $S(t) = 0, J(t) = 0 \quad \forall t < 0$ . As stated previously,  $J(t)$  represents the nodes able to infect other nodes at time  $t$ . Since half of the communicating nodes are also infected at time  $t$ , then, the actual number of infected nodes at time  $t$  is:

$$I(t) = J(t) + 0.5C(t) \quad (2)$$

Analysing this model, we can see that it is bounded by two components. When the communication time is near to zero ( $T_c = 0$ ) it is equivalent to the basic epidemic model, that have a simple analytical solution [6]:

$$I_\lambda(t) = \frac{N}{1 + (N - 1)e^{-\lambda N t}} \quad (3)$$

And when the contact rate tends to be high, the number of infected nodes doubles after each message transmission with a period  $T_c$ , and can be approximated by the following function:

$$I_{T_c}(t) = \min(2^{\lfloor t/T_c \rfloor}, N) \quad (4)$$

Note that  $I_\lambda(t)$  and  $I_{T_c}(t)$  are minimum values, so  $I(t)$  of expression 2 will be always greater or equal to  $I_\lambda(t) + I_{T_c}(t)$ . These components will be clearly evidenced in the experiments presented in the evaluation section.

From expression 2, we can obtain the number of nodes of the group ( $G(t)$ ) that received the message as:

$$G(t) = \frac{N_g}{N} I(t) \quad (5)$$

Finally, note that this model evaluates the diffusion of one message. Thus, considering the exchange of several messages when a contact occurs will increase this transmission time.

Another metric of interest is the delivery time  $T_d$ , that is the time when all the nodes in a group receive the message. Although  $G(t)$  gives an asymptotic value for the number of nodes in a group carrying the message, we introduce an expression for obtaining the average delivery time  $E[T_d]$ . Given that  $I(t)$  is the number of nodes that received the message up to time  $t$ , the probability that  $x$  members of a group have received a message from  $I(t)$  is given by the hypergeometric distribution, that models the probability of  $x$  from  $K$  successes, in a fixed-size sample  $n$  drawn without replacement from a finite population ( $N$ ). Thus, we have that:

$$H(t) = H(x = N_g; N, n = I(t), K = N_g) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \quad (6)$$

is the *cumulative* probability that all  $N_g$  members of a group have received the message when  $I(t)$  nodes have the message. In order to obtain the expected delivery time, we calculate the values of  $t_i$  when  $I(t)$  is increased by one (that is, a new node receives a message), so  $t_i$  corresponds to the delivery time of the  $i^{th}$  node of  $I(t)$ . Then, the expected value of  $T_d$  is:

$$E[T_d] = \sum_{i=2}^N t_i * (H(t_i) - H(t_{i-1})) \quad (7)$$

where  $H(t_i) - H(t_{i-1})$  is the probability for event  $t_i$

In order to study the message diffusion process, we evaluate the two components that generate this delivery time:

- 1) *Message communication delay* ( $D_m$ ), that is caused by the message transmission time. In the absence of contact delay, the number of infected nodes doubles after each message transmission as stated in expression 4, where  $t$  represents the communications delay. Solving for  $t = T_d$ , we have  $D_m = T_c \log_2(I(T_d))$ , where  $I(T_d)$  is the number of nodes that have the message up to the delivery time  $T_d$ .
- 2) *Contact waiting delay* ( $D_c$ ), that is caused by waiting for new contacts. This component does not have a direct expression, and it is calculated as the difference between  $E[T_d]$  and  $D_m$ .

Finally, the DDEs of expression 1 does not have a simple analytical solution. Thus, we can solve the DDEs numerically using Euler's method, with a step size of  $h$  and time  $t_i = hi$ :

$$\begin{aligned} S_{i+1} &= S_i + h(-\lambda S_i J_i) \\ C_{i+1} &= C_i + h(2\lambda S_i J_i - 2\lambda S_{i-tc} J_{i-tc}) \\ J_{i+1} &= J_i + h(2\lambda S_{i-tc} J_{i-tc} - \lambda S_i J_i) \end{aligned} \quad (8)$$

with  $C_i = 0, J_i = 0 \quad \forall i \leq 0, J_1 = 1$  and  $tc = \lceil T_c/h \rceil$ . Since  $G_i = (N_g/N)(J_i + 0.5C_i)$ , iterating over the previous expression while  $G_i \leq N_g$  allows obtaining the infected nodes of a group ( $G_i$ ) up to the time when all nodes of the group are infected.

This model was validated using a procedure similar to the one described in [6] comparing the results obtained with our analytical model with those obtained using a custom simulator that implements the contact-based message diffusion.

#### IV. EVALUATION RESULTS

This section is devoted to evaluating the diffusion of messages based on the model introduced in the previous section. For the following evaluations we consider a square area with side  $l = 100m$ . In this area we consider 500 individuals ( $N = 500$ ) that move freely, following the *Random Waypoint* (RWP) model with a mean speed  $E[V] = 0.5m/s$  and no pause time. The communication range ( $r$ ) is 7.5 m with mean bandwidth of  $Bw = 2.1Mb/s$  and setup time  $T_s = 0.1s$ , that are practical values of Bluetooth 2.0, Class 2 communication. In [2], it is shown that in a bounded domain (such as the one adopted in this paper) the pairwise inter-contact distribution between pairs of nodes following a RWP model is nearly exponential with a contact rate given by the following expression:  $\lambda \approx \frac{2.7366rE[V]}{l^2}$  when  $r \ll l$ . Using this expression, we obtain  $\lambda = 0.001s^{-1}$ .

The goal of the first experiment, is to compare (and validate) the results of our model with the ones obtained with the ONE simulator. For our experiment, we used epidemic routing, all nodes are destinations of the message (that is,  $N_g = N$ ) and the mobility model was set to RWP. Regarding the communication time, the message size  $m$  was set to 512KB, so  $T_c = 2.1s$ . The simulation was repeated 100 times, in order to obtain the mean and confidence intervals. In figure 1 we can see the cumulative number of nodes with copy ( $I(t)$ ), using the DDE model, the custom simulator and the ONE simulator (with mean and 95% confidence intervals). The results show that the DDE model results are close to the experimental

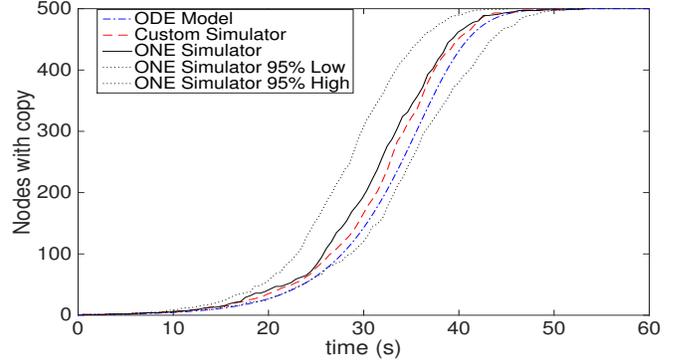


Fig. 1: Number of nodes with copy, for a message size  $m = 512KB$ .

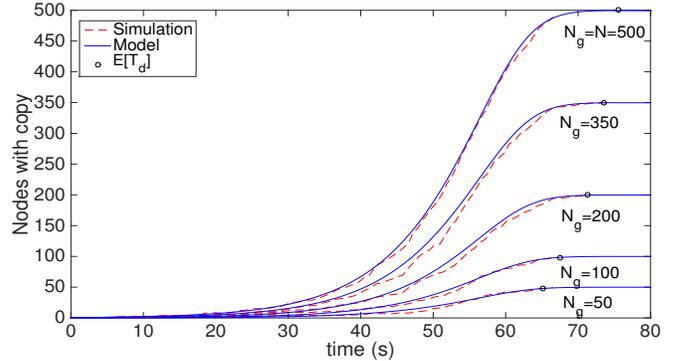


Fig. 2: Message diffusion for different group sizes ( $N_g = 50, 100, 200, 350$ ). The marks in the plots represent the average delivery time  $E[T_d]$ .

result, validating again our model. More experiments were performed with the ONE simulator with different sizes and mobility parameters obtaining similar results.

The next experiment evaluates the impact of group size on message diffusion. The message size was set to 1MB (so  $T_c = 4.1s$ ) and the number of members in the group was set to 50, 100, 200 and 350 members, respectively. Figure 2 show the results for the model and our custom simulator. First of all, we can see that the curves for the simulation and our model are similar, validating the results of the model. Regarding diffusion, the pattern is very similar for all curves, being characteristic of a typical exponential spread of the message, and the average delivery time increases with the group size.

In the third experiment, we evaluate the impact of message size on diffusion in a group ( $N_g = 100$ ). Specifically, we plot  $I(t)$  and  $G(t)$ , until all nodes in the group are infected (receive the message). The results are shown in figure 3. The first three curves have the following message sizes:  $m = 1KB$  (a typical text message);  $m = 512KB$  (a photo, for example); and finally  $m = 2MB$  (a short video). The diffusion of messages follows an exponential curve typical of epidemic diffusions, following expression 3. The curve at the right corresponds to a message size of  $m = 10MB$ . In this curve, we can see clearly the effect of the message communication time, and how the number of infected nodes doubles approximately every  $T_c$  seconds ( $T_c \approx 40s$ ). This effect is very important because, when the communication time is high, message spreading is delayed by this time and not by the contact rate. In fact, we can see that the number of infected nodes doubles after each

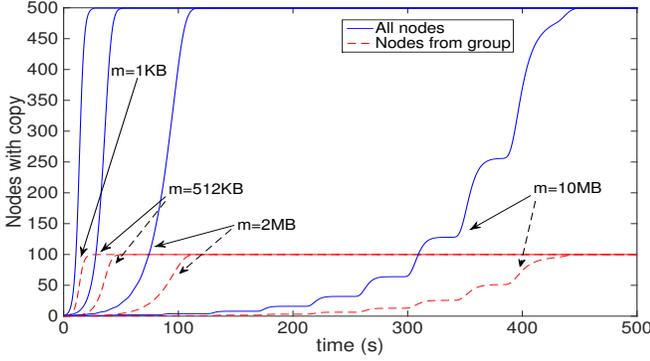


Fig. 3: Diffusion of one message for different sizes.

period, as represented by expression 4.

The goal of this final experiment is to evaluate the *average* delivery time  $E[T_d]$  in a group depending on the number of nodes, and for different group sizes. Figure 4 shows the results for two message sizes (1MB and 10MB). For  $m = 1\text{MB}$ , the delivery time decreases exponentially when  $N < 500$ . For higher values of  $N$ , the delivery time decreases very slowly, so the epidemic diffusion to these nodes does not produce a significant improvement. For  $m = 10\text{MB}$  (a diffusion bounded by the message communication delay), we can see that the delivery time is higher, and it does not decrease with the number of nodes; in fact, it slightly increases with the number of nodes. In order to explain this behaviour we plot the two components of the delivery time in figure 4b, for group size  $N_g = 50$ . We can see that for  $m = 1\text{MB}$ , the *message communication delay* is nearly constant, so the whole diffusion time decreases when the *contact waiting delay* decreases. Nevertheless, for  $m = 10\text{MB}$ , we can see that although the contact waiting delay decreases exponentially for  $N < 200$ , then, as the message communication delay increases, it slightly increases the whole delivery time.

Finally, repeating the previous experiments for different network parameters and message sizes, we can see that when the message delay  $D_m$  is greater than approximately 75% of  $E[T_d]$ , the effect of the message communication constrained delay is clearly evidenced, thus producing an increase of the delivery time when the number of nodes increases.

## V. CONCLUSIONS

In this paper, we study the diffusion of messages in a group, based on direct contact transmission, using an DDE based model. This model takes into account the communication time which is shown to be a key aspect in the diffusion of messages. No previous evaluation model takes into account this delay.

Based on the evaluation results, we can see that, when the message communication time ( $T_c$ ) is high, the overall message diffusion is bounded by this time (when message transmission delay is greater than 75% of the delivery time) and the effect is that the diffusion time increases slightly when the number of nodes increases (as opposed to the always decreasing diffusion time of the epidemic diffusion with no delay). In this situation, the epidemic diffusion is not optimal

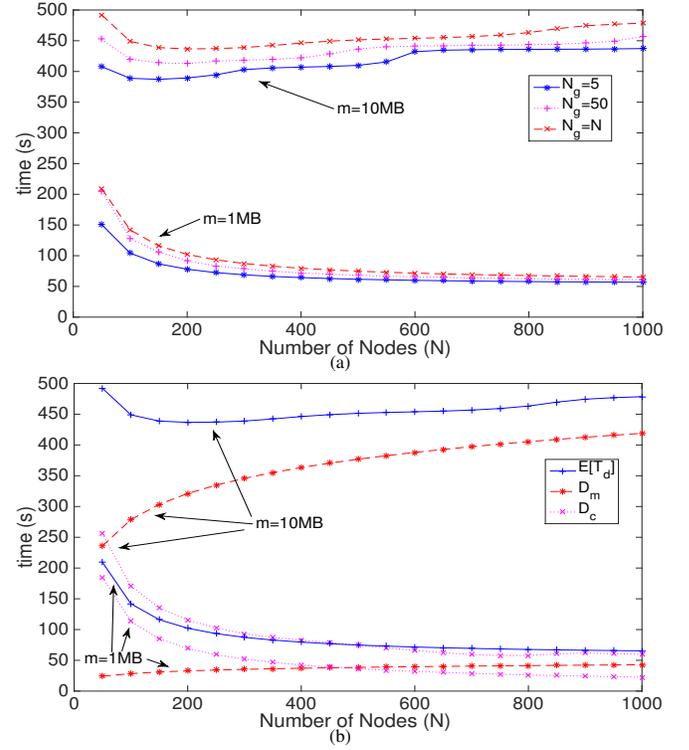


Fig. 4: Delivery time depending on the number of nodes. a) Results for message sizes 1MB and 10MB. For the curves labeled ( $N_g = 5, N_g = 50$ ) the nodes in the group are constant (5 and 50) and for  $N_g = N$ , the size of the group is  $N$  (all nodes). b) Components of delay for group size  $N_g = 50$ .

since it loses time transmitting the message to all the nodes, so alternative diffusion protocols must be considered.

## REFERENCES

- [1] J. Kangasharju, J. Ott, and O. Karkulahti, "Floating content: Information availability in urban environments," in *8th IEEE Conf. on Pervasive Computing and Communications Workshops*, 2010, pp. 804–808.
- [2] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, pp. 210–228, Oct. 2005.
- [3] Z. J. Haas and T. Small, "A new networking model for biological applications of ad hoc sensor networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 1, pp. 27–40, Feb. 2006.
- [4] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 1, pp. 77–90, Feb. 2008.
- [5] E. Hernández-Orallo, M. D. Serrat, J.-C. Cano, C. M. T. Calafate, and P. Manzoni, "Improving selfish node detection in MANETs using a collaborative watchdog," *IEEE Communications Letters*, vol. 16, no. 5, pp. 642–645, 2012.
- [6] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.
- [7] Y. Kuang, *Delay Differential Equations: With Applications in Population Dynamics*. Elsevier, 1993.
- [8] M. Alizadeh, A. Javanmard, and B. Prabhakar, "Analysis of dctp: Stability, convergence, and fairness," in *Proc. ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems*, 2011, pp. 73–84.
- [9] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 7, no. 3, pp. 19–20, Jul. 2003.
- [10] L. J. S. Allen, *Mathematical Epidemiology: Lecture Notes in Mathematics*. Springer Verlag, 2008, vol. 1945, ch. An Introduction to Stochastic Epidemic Models, pp. 81–130.