Modeling message diffusion in epidemical DTN

Claudio Sá de Abreu, Ronaldo Moreira Salles*

Dept. of Computer Engineering, Military Institute of Engineering (IME), Praça Gen Tibúrcio 80, 22290-270 Rio de Janeiro, Brazil

A R T I C L E   I N F O

Article history:
Received 16 October 2012
Received in revised form 20 September 2013
Accepted 23 December 2013
Available online 4 January 2014

Keywords:
Delay and disruption networks
Intermittently connected networks
SIR mathematical model
Epidemic routing

A B S T R A C T

A Delay and Disruption-Tolerant Network (DTN) is a fault-tolerant network where end-to-end connections are not required for message transmissions between nodes. Usually, a DTN is implemented as a wireless mobile ad hoc network that can be applied, for instance, to rapidly build a basic telecommunication infrastructure in case of catastrophes and disasters, or to support communication in a disruptive military environment. It is important to model DTN behavior to better understand system dynamics and related physical laws, which may impact network performance. An accurate model will be useful to support the design of the network in such challenging scenarios and may allow to test design ideas before actually building the real system. This work proposes a mathematical model for message diffusion in epidemical DTN. Our approach is based on previous models for the spread of human epidemical diseases, namely SIR. Simulation results on message diffusion times in an epidemical DTN show that the model is accurate regarding expected values, however large deviations above and below average are also observed on diffusion times. We further study such deviations and provide insights on how to reduce and deal with them, making the model useful for DTN applications.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In a Delay and Disruption-Tolerant Network (DTN) messages are transmitted hop-by-hop from one network node to another until the destination is reached. Such networks follow the store-carry-and-forward paradigm [1]. DTN technology is getting more importance nowadays since it manages to establish connectivity in challenging network environments where other types of technology fail:

- Mobile Phone to Mobile Phone networks, where wireless channel impairments and mobility cause disconnections and message delays.
- Car networks (VANETS), where there is no continuous network service along all paths and places.
- Disaster and war zones, where communication infrastructure is partially or completely destroyed.
- Sensor Networks, where some devices are not always reachable to the sensor network coverage.
- Communication in remote zones with no fixed infrastructure.
- Interplanetary communications, where the line of sight communication between stations is difficult and delays are huge.

One of the most basic and important services provided in the above-mentioned scenarios is information diffusion. For instance, message diffusion can be used in VANETs to build up a car alert accident system to notify all cars that enter the affected regions. In war zones, a commander may want to issue an order to all his troops; in disaster zones, a message conveying the current state of the catastrophe may be issued to all local units attending the area; in sensor networks, contaminations detected by one sensor may be diffused to all other sensors or collecting nodes.

Regarding the delivery time and communication efficiency, good results are obtained when methods of epidemic diffusion of messages are used. Each node that
carries the message transmits copies of it during an established time interval to all nodes that still do not have the message, as initially proposed by Vahdat and Becker [2]. This type of DTN is also the fastest to deploy, because:

- It does not require any previous knowledge of network topology.
- There is no restriction on the capacity of the propagating nodes.
- It is not necessary to consider paths or history of contacts.

Fig. 1 illustrates in a small scale example how an Epidemical DTN works. Consider initially a single source and a single destination node. The arrows depicted in all four subfigures (Fig. 1(a)–(d)) indicate the moving direction of each node. In Fig. 1(a), the source node (blue dot) generates a message to the destination node (green dot). Fig. 1(b) and (c) show intermediate steps, where other DTN nodes are contacted and receive a copy of the message. Those nodes continue to propagate the epidemic to all their contacts until the message is eventually delivered to the destination (delivery time), as shown in Fig. 1(d). At this point two nodes (red dots) still do not have a copy of the message. In case of message diffusion to all network, the process continues until all nodes get a copy of the message (all nodes are blue), which constitutes the total message diffusion time in the network.

Epidemic diffusion of messages is also a useful transmission method when network topology is unknown or frequently changes over time due to the movement of mobile nodes. However, in epidemic diffusion, for high transmission rates, both battery consumption and memory usage are critical, given that each network node ends up transmitting more frequently and keeping almost all copies of messages in transit in their memory.

It is important to model epidemic behavior in order to assist the design and management of an epidemic DTN. With an appropriate model it is possible, for instance, to
test design ideas before actually building the real system, and also to derive and maintain the optimal parameter settings for the desired DTN operation in terms of communication efficiency, memory usage and latency (message delivery time). Such model constitutes the main motivation behind this work.

One major mechanism to control the lifetime of a message in an epidemiological DTN is message TTL (Time-To-Live). DTN performance depends on message TTL since memory usage, latency and communication efficiency are all affected by this parameter. TTL may also be used as a QoS parameter for some DTN applications and scenarios where a delay bound to message delivery time is necessary. For instance, in a military scenario, commander messages to all troops have fixed lifetimes otherwise the whole operation may be compromised by late deliveries. Hence, TTL plays an important role on the performance and modeling of message diffusion in DTNs.

Due to the similarity observed between the epidemic transmission of messages in a DTN and the spread of infectious diseases, several works, like [3,4], employed analytical models of epidemic diffusion in populations as a base for DTN modeling. Almost all those works were derived from the SIR model [5,6], which is a three-state-model for epidemiological disease where an individual may be in one of three possible states, in relation to the contamination: Susceptible, Infected or Recovered.

Regarding epidemiological models applied to the DTN area, the work in [3] introduced important analytical results for node contact rates and for the time distribution of these contacts, under the assumptions that: network is homogeneous and sparse, node transmission range is much smaller than the surface area, transmission interference and impairments can be neglected. Those assumptions were preserved in our work for the different scenarios studied.

However, it was observed from simulations that the above-mentioned models are only accurate for the average case, i.e. when results of epidemic evolution (e.g. number of infected nodes over time) were averaged for a large number of different scenario settings (e.g. different source nodes – initial conditions). In some cases the epidemic spread in one setting is much slower than in another more favorable setting providing large deviations above and below the mean, either in diffusion times, nodes contact times or number of infected nodes over time. Such deviations were not taken into account on previous works limiting the applicability of the existing models to average cases only.

This work starts studying how analytical models for the spread of epidemic diseases could be applied to the DTN scenario, similarly to what has been done in some previous works. However, it goes further in the analysis of expected results and their deviations, creating a better framework for the design of a practical model to study DTN dynamics. The proposed model is validated through simulation results.

The contributions of this work are:

- The employment and validation of a mathematical model for epidemiological DTNs based on an epidemiological model for the spread of infectious diseases (for the average case).
- The discussion and investigation about the reasons behind the large deviations observed in simulation results. Such deviations were ignored or sometimes were observed but not treated in previous works, like [7–9].
- The study of important DTN performance parameters from the proposed model: the message TTL, the node contact rate (β), the number of nodes and the work area dimensions.
- The proposal of a simple framework to predict the behavior of an epidemiological DTN in both mean and variance of message diffusion times (or number of available/propagating nodes over time).
- The suggestion of practical procedures to improve the functionality of epidemiological DTNs from the results obtained with the model.

The rest of the paper is organized as follows. Section 2 gives a general view of recent developments on DTN mathematical modeling. In Section 3, the Epidemical DTN Basic Model (EDTN) is investigated, while in Section 4 some important characteristics of epidemiological DTNs are observed from the model. In Section 5, the model is validated through simulations for the average case, however deviations in the number of available/propagating nodes (nodes without/with the message) over time are also observed for all scenarios. Section 6 deals with such deviations and discusses how to calculate them in order to improve model precision. A statistical model is proposed to predict DTN behavior regarding the initial node contact times and their variance. Section 7 presents a summary of the main parameters derived from the model, and finally, the paper is concluded in Section 8.

2. Related work and historical review

In 1906, Hamer [10] postulated that the evolution of an epidemiological disease depends on the contact rate between infected and susceptible individuals. This principle, now known as the mass interaction principle, becomes one of the most important concepts of mathematical epidemiology. The main idea was that the spread of an epidemic disease in a population is proportional to the density of susceptible by the density of infected individuals. The Hamer principle was originally formulated using a discrete time model, but in 1908 Sir Ronald Ross, who discovered the vector of malaria transmission, generalized the model for continuous time in his work about malaria dynamics.

In 1927 and 1932, Kermack and McKendrick in [5,6] improved the continuous model created by Sir Ronald Ross with the threshold principle, establishing that the introduction of infected individuals in a community may not cause an epidemic, unless the density of susceptible individuals is above a certain critical value. This principle, in conjunction with the mass action principle, are the basis of a model called SIR (Susceptible, Infected, Recovered), which is one of the most important models in modern mathematical epidemiology [11].

During the following decades, the model of Kermack and McKendrick was extended and improved, leading to
more sophisticated models, which have applications in many different types of spreading diseases. In recent years, mathematical modeling of diseases has become increasingly important in the control of current epidemics [12]. Nevertheless, SIR based models have also been applied in other areas whenever the studied scenario has similarities with the dynamics involving the spread of diseases.

In 2000, Vahdat and Becker [2] developed techniques to deliver messages in partially connected networks, like MANETs. Such work employed epidemic routing.

In 2002, Kevin Fall proposed the first model of what we now call DTN [13]. In that work and the following [13,1], the major concerns were with the concept and necessary structure for DTN implementation.

In 2005, Groenevelt et al. [3] proposed a mathematical model for DTN based on previous models for epidemic spread of diseases in populations. One of the main contributions of that work was the presentation of a closed form expression to estimate the parameter $\beta$ (see Eq. (15) in Section 5), the contact rate among nodes. This result was obtained by demonstrating that contact times between nodes are exponentially distributed if some assumptions hold: network is homogeneous and sparse, node transmission range is much smaller than the surface area, transmission interference and impairments can be neglected. The expression still applies for different node movement models, such as the Random Way Point, Random Direction and Brownian.

In 2006, Haas and Small [7] used differential equations and concepts applied to Epidemiology to model a network that used whales as DTN data mules. That work did not mention the problems related to the delay between the first contamination (source data node) and the spread of the epidemic to the population. It also did not take into account the conditions that support the modeling of a message diffusion in DTN as an epidemic process in mathematical epidemiology.

The work in [8] developed a model to study the performance of ad hoc networks according to the movement of nodes. Such model can be extended to the DTN scenario and is useful for obtaining the efficiency of communication in this context.

Zhang et al. in [9] used the approach derived by Groenevelt et al. [3] as basis to propose and investigate a unified ODE framework to study the performance of various forwarding and recovery DTN schemes. They derived the ODE model as limiting processes of Markovian models and employed the equations to obtain a rich set of closed-form formulas regarding the packet-delivery delay, number of copies sent and buffer occupancy under various schemes. Those equations were validated through simulations, and good matches were obtained between the model prediction and simulations results. However, as done in previous works, the results were obtained for the average case only, and the deviations in forwarding times were also not treated by their approach.

In 2008, Walker et al. [14] studied the performance of DTNs considering nodes moving according to the random walk mobility model. Such work introduced the localized random walk mobility model, derived some of its properties and used it to construct models for simple DTN scenarios. The authors further checked the predictions against simulation results.

In 2010, Lu and Hui in [15] proposed an energy-efficient n-epidemic routing protocol for DTN, which was based on a mathematical model similar to the one studied in [7]. However, the work did not consider possible deviations of the model for some specific scenarios.

In 2010, Jacquet et al. [16] derived a theoretical upper bound for the information propagation speed in two dimensional large scale networks, and multi-dimensional networks. Such theoretical bounds are useful to improve the understanding about the fundamental properties and performance limits of DTNs, as well as to evaluate and/or optimize the performance of specific routing algorithms.

The model used in that analytical study encapsulates many popular mobility models, such as random way-point, random walk and Brownian motion, and provides a general framework for the derivation of analytical bounds on the information propagation speed in DTNs. It explicitly takes into account the spatial dimension as well as nodes density and mobility.

In 2012, Shahbazi et al. in [17] proposed ways to improve DTN performance using stationary nodes, however all their analysis came directly from some particular experimental results, they were not concerned with the construction of a more general mathematical model.

Khabbaz et al. in [18] made a complete survey on recent developments and persisting challenges in DTN. In that work, among other important conclusions is the observation that still there is no adequate analytical model for the performance evaluation of DTNs. This topic persists as an open problem.

Wang and Haas in [19] proposed a methodology where a random graph was used to model the dynamic connectivity among network nodes. The work derived three main metrics: the probability distribution of the number of infected nodes, the average number of infected nodes and the probability of all nodes being infected, all three metrics being functions of time. Simulation results were very close to the results predicted by their model for the averages of the three metrics. Despite their contribution, that work did not take into account the standard deviation of each metric on each time instant.

In 2013, Venkadatri et al. in [20] proposed the use of M2M communication(18,8),(989,994) for epidemic DTNs.

The works in [7–9] provided some interesting analytical models that could be employed to evaluate the performance of DTNs. In different ways, those works reached some of the same equations presented in Section 3 of our work, specially the ones in (8). Those equations are important to better understand the dynamics involving DTN transmissions as a whole, but are not enough to explain the differences observed on contact times when the model is compared with simulation results obtained for some specific scenarios. The understanding of such deviations is one of the main goals of this work.
3. The Epidemical DTN Basic Model

In this section, a mathematical model for epidemical DTNs will be investigated based on previous models for the epdemical spread of diseases in populations.

Regarding the human spread of diseases, it is well known that some diseases are transmitted from one individual to another by direct contact or proximity, like flu or cholera. In general, when a significant portion of the population is infected and the rate of growth of the infected population is high, it can be said that an epidemic is happening. There are several mathematical studies investigating the dynamics of epidemic spread of diseases, but the original one used as reference by almost all others was first proposed in [6], named SIR. That model is based on the assumption that in an approximately constant population, an individual in a given time can be only in one of three possible states: Susceptible (S), Infected (I) and Recovered (R). A set of differential equations show how the states S, I and R vary over time.

The dynamics involving the transmission of a message through nodes in an epidemical DTN (called message forwarding) and the spread of an epidemical disease, when an infected individual infects others, enjoy several similarities. The agent in the first process is the message carried by a node to be transmitted to other nodes, while in the second process is a virus or some other plague carried by a person. In both cases, the agent stays in a carrier for some time before the carrier becomes unavailable (recovered). Because of this similarity, the SIR model is used as a basis for creating a mathematical model for epidemical DTN, considering that one node, in a given time, can only be in one of three possible states:

- **available (D):** The node is available to receive a message;
- **propagating (P):** The node has received a copy of the message and is spreading it until TTL (Time To Live) expires;
- **unavailable (I):** The node had already received the message and it has been delivered, or its TTL has expired.

From the above discussion and following what has been done before to model the spread of epidemical diseases [5–7,9], it is possible to derive a simple model that hopefully approximates the dynamics of message transmissions in epidemical DTNs.

One of the main assumptions of the model is that the number of nodes on each state (D, P and I) can be approximated by continuous and differentiable functions of time, \( t \): \( D(t), P(t) \) and \( I(t) \). Other assumptions are listed below:

- The number of nodes \( N \) is constant during message propagation.
- At any time instant all nodes are equally likely to meet, i.e. the effects of the spatial dimension are neglected.
- When a contact occurs between a node in \( P \) and a node in \( D \), the message transferring probability is \( 1 \).
- When a message is received, the node is immediately able to transfer it to all reachable available nodes during message lifetime.
- A node makes only one message forwarding at each time.
- A node makes an average of \( \beta N \) contacts per unit time, where \( \beta \) is the contact rate of the network.
- The nodes in propagating state go to unavailable state with rate \( \gamma P \).

The diagram below shows the sequence of states.

According to Fig. 2, each arrow indicates the transition rates from one state to the following. Between D and P, as one node makes \( \beta N \) contacts per unit time, \( P \) nodes make \( \beta NP \) contacts per unit time. Then, since \( \frac{D}{N} \) is the proportion of available nodes in a given time, we can conclude that, in each time unit, \( \beta NP \frac{D}{N} = \gamma DP \) nodes are infected. Thus, \( \beta DP \) is the rate of state change for available nodes per unit of time.

Between P and I, the transition rate is \( \gamma P \), which is the rate of depletion, i.e. the number of nodes that stop transmitting the message per unit time. So, \( 1/\gamma \) is the average time that a node remains transmitting the message.

Given that \( N \) is the total number of nodes, we have for each time \( t \): \( N = D(t) + P(t) + I(t) \). To simplify notation we use:

\[
N = D + P + I \quad (1)
\]

From model construction assumptions, it is possible to derive (1) over time, obtaining:

\[
\frac{dD}{dt} + \frac{dP}{dt} + \frac{dI}{dt} = 0 \quad (2)
\]

![Fig. 2. Three/two state models.](image-url)
According to the transition rates presented in Fig. 2, we have:

\[
\begin{align*}
\frac{dP}{dt} &= -\beta DP \\
\frac{dD}{dt} &= \gamma P \\
\end{align*}
\] (3)

From (3) and (2), we can conclude that:

\[
\frac{dP}{dt} = \beta DP - \gamma P \\
\] (4)

If \( \gamma = 0 \) (such consideration will be justified later on), \( \frac{dP}{dt} \) is always greater than zero, except in the end of communication (diffusion process). As no node becomes unavailable, the model can be simplified to:

\[
\left\{ \begin{array}{l}
D + P = N \\
\frac{dD}{dt} + \beta DP = 0 \\
\end{array} \right. \\
\]

or

\[
D' + \beta ND = \beta D^2 \\
\] (5)

This is a Bernoulli differential equation that can be solved for \( D \) as follows. Multiplying (5) by \( D^{-2} \) and making \( S = D^{-1} \), \( S' = -D^{-2}D' \), we have:

\[
S' - \beta NS = -\beta \\
\]

then,

\[
S = \frac{1}{N} + Ce^{\beta Nt} \\
\] (6)

The DTN communication process starts when a node (source) generates the message, then the initial condition is that \( D(0) = N - 1 \Rightarrow S(0) = \frac{1}{N - 1} \). Which applied in (6), gives:

\[
C = \frac{1}{N(N - 1)} \\
\]

then,

\[
S = \frac{1}{N} - \frac{1}{N(N - 1)} e^{\beta Nt} = \frac{N - 1 + e^{\beta Nt}}{N(N - 1)} \\
\] (7)

As \( D = S^{-1} \), the model can be finally expressed by:

\[
\left\{ \begin{array}{l}
D(t) = \frac{N(N - 1 + e^{\beta Nt})}{N(N - 1)} \\
P(t) = \frac{Ne^{\beta Nt}}{N(N - 1 + e^{\beta Nt})} \\
N = D(t) + P(t) \\
\end{array} \right. \\
\] (8)

One of the main differences between the application of the SIR model in a DTN environment and the application in a human epidemic scenario relies on the determination of parameters \( \beta \) and \( \gamma \). In the latter case, parameters are directly obtained from the behavior observed on individuals of an infected population, while in a network they depend on message transmission dynamics.

The work in [3] derived an expression for \( \beta \) under amenable assumptions suitable for DTN environments, such expression will be used in Section 5 and compared to simulation results. The other parameter \( \gamma \) (rate of depletion) enjoys a direct relationship with message TTL as it will be explained next, but first we clarify the role of message TTL in a DTN.

The TTL (Time To Live) is one of the main mechanisms used in an epidemic DTN to discard messages, it can be defined in several different ways, such as: global TTL, local TTL, hop-based TTL and time-based TTL [22]. In most cases, TTL is used as a fixed deadline valid for the original message and all copies forwarded by intermediate nodes [7,9].

In common data networks, the TTL of a packet is usually treated as a hop count, which is decreased by one at each hop a packet goes through. When the number of hops is equal to the original TTL of the packet, it is discarded. Such approach can also be used in DTNs, however given the particular requirements and message transmission dynamics observed in DTNs, we believe the use of message TTL as message lifetime (or literally “time to live”) is more adherent to the architecture.

Typically, in common data networks a message is sliced into packets that are sent through the network until reach the destination, where the slices are remounted. It is also expected that such communication process is reliable and enjoys low packet delays. In this context, package TTL is used more as a network management mechanism to respond to bad routing decisions and avoid packets being indefinitely forwarded inside the network.

DTNs usually operate in a completely different scenario, where end-to-end connections cannot be established, delays are expected to be huge, nodes operate under energy restrictions, messages are not chopped into packets and can be hold in a node’s buffer for a long period of time until a next contact occurs. In most cases, nodes in a real DTN are actually wireless devices and have limited autonomy. In these scenarios message TTL is an important information to be confronted to battery duration in order to drive routing decisions and buffer management procedures. Such information also saves the scarce resources of the network in holding and transmitting a message after it has expired. Therefore, in this work the TTL will be used as a fixed time, literally, the total time-to-live of a message. The simulator TheOne, used to obtain the simulation results presented in this work, also implements TTL in this way.

Nodes remain transmitting the message until TTL expires and then discard the message before becoming unavailable. However, for a very large TTL (say TTL \( \to \infty \)) upon receiving and storing the message no node will drop it and remains propagating it forever.

Since TTL is an external parameter and can be freely set according to the application, user or network operator, we consider it large enough so that no node becomes unavailable during the time window under study (from message generation until complete diffusion), which justifies the assumption of \( \gamma = 0 \) taken before. Hence, the Epidemical DTN Basic model is reduced to the two state model indicated in the gray area of Fig. 2 and represented by the system of equations in (8).

A similar set of equations could also have been extracted from [7,8] or [9], however we presented a more direct derivation and used some of the intermediate steps in the next sections of this paper.

It remains to be seen if such simplified model indeed approximates the behavior of epidemic DTNs, but before that we use the model to study other important properties.
4. Some important epidemiological DTN model characteristics

4.1. The mean forwarding times and intervals

Based on (8) and keeping in mind that the kth message forwarding will occur when $D = N - (k + 1)$, the average time $t_k$ of the kth message forwarding is given by:

$$N - (k + 1) = \frac{N(N - 1)}{N - 1 + e^{\beta N t_k}}$$
$$N - 1 + e^{\beta N t_k} = \frac{N(N - 1)}{N - k - 1}$$
$$e^{\beta N t_k} = \frac{N(N - 1)}{N - k - 1} - N + 1 = \frac{(N - 1)(k + 1)}{(N - k - 1)}$$

then:

$$t_k = \frac{1}{\beta N} \ln \left( \frac{(N - 1)(k + 1)}{(N - k - 1)} \right)$$

for $k = 1, \ldots, N - 2$ \hspace{1cm} (9)

From (9) and letting $\Delta_k$ be the kth interval between forwarding, i.e., $\Delta_k = t_k - t_{k-1}$, we have:

$$\Delta_k = \frac{1}{\beta N} \ln \left( \frac{(N - 1)(k + 1)}{(N - k - 1)} \right) - \ln \left( \frac{(N - 1)(k)}{(N - k)} \right)$$

$$\Delta_k = \frac{1}{\beta N} \ln \left( \frac{(k + 1)(N - k)}{k(N - (k + 1))} \right)$$

for $k = 1, \ldots, N - 2$ \hspace{1cm} (10)

It is important to observe the conditions in which Eqs. (9) and (10) apply. They are not valid for $k = N - 1$, i.e., for the last message forwarding. Then, to derive an estimation for the time until the whole population has gotten the message, we go further into the analysis in Section 4.2, such estimation will be given as a lower bound for TTL as it will be seen in Eq. (14).

4.2. Lower bound for TTL

To maximize the delivery probability of a message in a DTN, maintaining the message in the network for the shortest time possible, TTL should be large enough so that it is expected that all nodes receive the message before the epidemic process ends, but not much larger than that.

Eq. (9) could not be applied to determine a lower bound for the TTL, because it cannot be used when $D = 0$ (all nodes infected). We are going to explore the symmetry observed in Eq. (10) to estimate such bound.

Expanding (10) for $k = a$ and for $k = N - (a + 1)$ it can be seen that $\Delta_0 = \Delta_N - (a - 1)$, which means that the state change intervals are symmetrical in relation to the point where $D = \frac{N}{2}$, i.e., $\Delta_1 = \Delta_N - 2$, $\Delta_2 = \Delta_N - 3$ and, generalizing:

$$\Delta_k = \Delta_{N-(k+1)}$$

(11)

Back to (9), the average time for the first contact is:

$$t_1 = \frac{1}{\beta N} \ln \left( \frac{2(N - 1)}{N - 2} \right)$$

(12)

If $N$ is large enough, (12) can be reduced to

$$t_1 = \frac{1}{\beta N} \ln \approx 0.69 \frac{203}{\beta N}$$

Because of the symmetry proved in (11), the total time to deliver the message can be approximately calculated as being twice the time to deliver the message to half of the nodes, i.e., $t_{N-1} \approx 2t_2$. Then, applying this to (9):

$$t_{N-1} \approx \frac{2}{\beta N} \ln \left[ \frac{(N - 1)(N/2 + 1)}{(N/2 - 1)} \right]$$

(13)

Then,

$$TTL \geq \frac{2}{\beta N} \ln \left[ \frac{(N - 1)(N + 2)}{(N - 2)} \right]$$

(14)

In average, when $TTL > t_{N-1}$, it is expected that at least about $N - 1$ nodes have received the message, and so to maximize the delivery probability TTL must be greater than $t_{N-1}$. Therefore, the right hand side of (14) can be considered as a practical bound for TTL. As it is not possible to assure that all nodes will receive the message, increasing TTL much above (13) may cause waste of network resources since the message stays in the DTN consuming memory of all nodes for more time than necessary in most cases.

The above formulas could also be used for network design purposes. Suppose in a given DTN, messages should be delivered before a fixed period of time, say one hour ($TTL = 60$ min.) or one day ($TTL = 1440$ min.). Hence, some DTN configurations would be expected to diffuse the message to all nodes before TTL expires and others not. Those expected to deliver the message are the ones configured with some specific values of $N$ and $\beta$, such that from (14); $t_{N-1} < TTL$.

5. Validation with simulation

In order to validate the model presented in (8) and the formulas derived thereafter, we compare them with the results obtained with a well known DTN simulator: The One [23].

Some simulator modules were modified to facilitate the comparison with the model, specially the generation of specific data reports. In addition to that, given the large amount of experiments that were carried out, it was also necessary to develop scripts to automate the tasks of scenario generation, simulating, processing reports and output files. All the developed framework, including adaptations on the simulator, scripts and results can be found at http://cabreu.vialink.com.br/mestrado/dtn/simulacoes.

In all simulations, the contact rate $\beta$ was obtained and compared to the estimation proposed in [3], which considers the Random Way Point (RWP) and Random Direction (RD) movement models:

$$\beta \approx \frac{203 e V}{L^2}$$

(15)

where:

- $\beta$ is the contact rate between any two nodes, i.e. the mean time between two given nodes meeting each other is $\frac{1}{\beta}$.
• \( \omega \) is a constant related to the specific mobility model used. For random waypoint model (RWP), \( \omega = 1.3683 \).
• \( L \) is the side of the area where nodes are placed and move around.
• \( c \) is the transmission range of each node. It is assumed that only one transmission occurs at each node interfering and that \( c \) is much smaller than \( L \). Furthermore, if \( c \) is near \( L \), the end-to-end direct communication is much more probable, degenerating the DTN to a directly connected network.
• \( E[V] \) is the average relative speed between two nodes, and can be calculated by a numerical integration given by Hyytia et al. [24].

Simulations were set for different values of \( v_{\text{min}} \) (minimum node speed), \( v_{\text{max}} \) (maximum node speed), transmission ranges, number of hosts and TTls. We used 1000 different random seeds for each simulated scenario. The RWP movement model was adopted to all scenarios. In total, we ran over 7000 simulations for about 100 different scenarios.

Considering a given scenario, the number of available nodes \( D \) over time was obtained for different scenario settings (different random seeds). Then, taking all settings, the average number of available nodes over time and the corresponding standard deviation were computed. The evolution of the average and standard deviations on the number of available nodes \( D \) over time were plotted and compared to the expected values given by the mathematical model, see Figs. 3 and 4.

In other words, for each Scenario \( SC_i \) of a set of \( i \) scenarios, \( S_i \) simulations were carried out \( (S_i \) scenario settings), \( SC_{i_k} \) \( k = 1,2,\ldots,S_i \), each one with a different random seed. Then, a sufficiently small time interval, \( \Delta t \), was used to evaluate the amount of available nodes over time \( D_k(t) \): for \( t = 0, \Delta t, 2\Delta t, 3\Delta t, \ldots \). The end of each simulation is reached when \( D = 0 \), producing a sequence of values: \( D_k(0), D_k(\Delta t), D_k(2\Delta t), D_k(3\Delta t), \ldots \) Thus, the average of \( D \) over time for scenario \( SC_i \), \( D_i(0), D_i(\Delta t), D_i(2\Delta t), \ldots \), is obtained from

\[
D_i(s\Delta t) = \frac{\sum_{k=1}^{S_i} D_k(s\Delta t)}{S_i} \quad \text{for } s = 0, 1, 2, \ldots
\]

The curve of average results for scenario \( SC_i \) as plotted in Figs. 3 and 4, is given by the pairs \( \{s\Delta t, D_i(s\Delta t)\}, s = 0, 1, 2, \ldots \). The standard deviation in relation to the number of available nodes \( D \) of each set \( \{D_k(\Delta t), k = 0, 1, \ldots, S_i\} \) was also calculated and used to evaluate the deviations between the model and the simulated scenarios.

Table 1 illustrates two of these scenarios, SC1 and SC2. SC1 is the same scenario used in [3], while in SC2 several simulation parameters were changed (area size, transmission range, number of nodes, node speeds).

Figs. 3 and 4 present a plot of the amount of available nodes \( D \) over time. It can be seen that the theoretical curve given by the proposed model in Eq. (8) is very close to the simulation curve obtained from the average of 1000 runs in the DTN simulator for both scenarios. The coefficient of determination obtained by the method of least squares, \( R^2 = 0.99561 \), indicates a good fit of the model to predict average system behavior. That is, even a continuous model, which is essentially a fluid approximation, provides a good fit for the expected number of available nodes \( D \) in a discrete scenario with a finite number of nodes.

Although the average of simulation results for \( D \) provided a close approximation to the proposed model, it is important to observe that the standard deviation of \( D \) on all simulations is very high.

### Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SC1</th>
<th>SC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area shape</td>
<td>Square</td>
<td>Square</td>
</tr>
<tr>
<td>Side</td>
<td>2 km</td>
<td>6 km</td>
</tr>
<tr>
<td>Range</td>
<td>30 m</td>
<td>100 m</td>
</tr>
<tr>
<td>Movement model</td>
<td>RWP</td>
<td>RWP</td>
</tr>
<tr>
<td>Nodes</td>
<td>40</td>
<td>140</td>
</tr>
<tr>
<td>Node speeds</td>
<td>4–10 km/h</td>
<td>5–50 km/h</td>
</tr>
</tbody>
</table>
It can be seen in Figs. 3 and 4 that the standard deviation is larger than the expected value after 3000 s of simulation for SC1 and 1000 s for SC2. The model cannot capture such deviations since it only provides estimation for expected values based on the continuity (system is continuous) and uniformity (all nodes are equal and present the same behavior) assumptions and on fluid approximation.

Fig. 5 shows two extreme examples of scattering, one much lower than the average (for simulation seed = 53) and the other far above (for simulation seed = 90). The scenario generated with seed = 53 illustrates a fast epidemic spread where message diffusion was completed just after 3000 s. On the other hand, for scenario with seed = 90 the epidemic process was delayed until approximately 2000 s when first contacts start to be established (the origin node was isolated from the others).

Another situation can also be observed from the curves in Fig. 5, there is also a difference on the behavior of the epidemic near the end of the process at each case. Although SC1 with seed = 90 suffered a long initial delay, last contacts were faster (last available nodes were closer to the rest of propagating nodes) than for SC1 with seed = 53.

Such differences on the curves of Fig. 5 may explain why standard deviation on D over time tends to be high.

The effects of the initial and final phases of the epidemic (“border effect”) plays an important role on the deviations observed in the process.

The model as it is cannot predict deviations, such issues are discussed in Section 6.

6. Dealing with model deviations

No mathematical model for DTNs is really useful if there is no way to predict and minimize large deviations that occur in relation to the average case, as already showed in the last section. Such behavior was also observed in [25], but was not predicted nor treated in previous works, like [3,4].

We start investigating the relationship between model deviations and the border effect just mentioned. First of all, the model is itself symmetric as shown in Eq. (11). A system with x propagating nodes and y available nodes has the same behavior of a system with y propagating nodes and x available nodes regarding to message forwardings. From the law of mass action, the product xy is the same in both cases. Hence, model behavior in the initial phase is similar to the final phase.

Another important aspect of the model is that the rate of change in D is lower at the borders (either initial or final phase) than in middle points. The largest decrease rate on D is obtained when D’ is maximum, i.e. when D” = 0. This can be solved looking again to Eq. (5), observing that D” = 2βD – βN and making D” = 0:

\[
2\beta D - \beta N = 0 \Rightarrow D = \frac{N}{2}
\]

Then the largest transmission rates, or the fastest rates of epidemic spread (“epidemic speed”), are obtained when \( D = \frac{N}{2} \) whereas lowest rates occur at the beginning and end of the process. At the borders either \( x \) (the number of propagating nodes) or \( y \) (the number of available nodes) is small.

One important assumption applied so far is that the analysis does not take in account the spatial dimension. At any time instant, all nodes have the same probability to meet each other. That is probably a good assumption for large systems where one can work with averages (fluid flow approximation), but can cause serious deviations in small finite systems specially at the borders where the epidemic process depends on a single (few) node(s).

Hence, at the borders the particular situation of a (a few) node(s) may cause a great impact and dependency on the message diffusion process as a whole, we call such phenomenon as the “border effect”. Fig. 6 can be used to illustrate different situations at initial and final phases of the process.

Assuming node A the source of a message, the DTN communication process will be delayed until node A eventually encounters another node and forwards the message to it. On the other hand, if B is the source node, epidemic spread will be very fast with very low delays in this initial phase. The probability of an encounter and contact rates for node A are much lower than for node B in the beginning of the process. Such situations resemble the simulation scenario with seed = 90 and seed = 53 as shown in Fig. 5, respectively. In both cases, deviations from what was
expected by the model were observed and suffered the influence of the border effect.

The border effect is specially important because, if the epidemical process is delayed by a slow starting node, or anticipated by a fast one, all subsequent forwardings are affected.

To evaluate how much the first contact affect the overall simulation, we studied the importance (weight) of the first contact times to the variance observed in the overall simulation. To do so, we considered all the initial contact times for each scenario as new starting times for the simulation. Fig. 7 illustrates such approach, where curve “STDEV N-1” considers the standard deviation on the number of available nodes \( D \) over time obtained for the results just after the first contact time (new simulation starting point), “STDEV N-2” for the second contact time, and so on. Thus, any possible effect of first contact times is eliminated, and the resulting variance is due only to the rest of simulations.

It can be seen from the simulation results that the contribution of first contact times is relevant to the variance of the process as a whole. In fact, the first contact contributed approximately with \( \frac{1}{2} \) of the total standard deviation. Fig. 7 for SC1 shows the first, second and third contacts influence on the standard deviation.

Given the importance of first contact to the dynamics of message forwarding in an Epidemic DTN, we rewrite our model in (8) as a function of \( t_1 \), as follows. From eq. (12), \( \beta N = \frac{1}{k} \ln \left( \frac{2N-2}{N-2} \right) \), then applying it to (8), we have:

\[
D(t) = \frac{N(N-1)}{(N-1) + \left( \frac{2N-2}{N-2} \right)^{t}}
\]

Assuming that first contact time is exponentially distributed with mean \( t_1 \) [3], it is possible from (17) to derive a simple Monte Carlo method to estimate average and standard deviation results without resorting to specific DTN simulation codes and packages.

Let \( \tau_1, \tau_2, \ldots, \tau_m \) be a sequence of samples from an exponentially distributed random variable of mean \( t_1 \), representing samples of first contact times. For a given sample \( \tau_k \), (17) can be expressed as:

\[
D(t)_k = \frac{N(N-1)}{(N-1) + \left( \frac{2N-2}{N-2} \right)^{\tau_k}}
\]

which gives the number of available nodes as a function of time assuming first contact occurred at \( \tau_k \). Hence, using all samples one can plot estimations for average and standard deviations results for the number of available nodes \( D \) as shown in Fig. 8 (for SC1) and Fig. 9 (for SC2).

It can be seen that the maximum standard deviation results obtained directly from the model (using simple Monte Carlo calculations)\(^2\) and from DTN simulations (using The ONE simulator) are very close. The advantage of the former method is that it is much faster than the latter one, besides there is no need to set up a simulation model nor to configure and run specific DTN simulation packages.

The whole process is indeed very fast, and so the use of (18) becomes an interesting way to predict the behavior of an epidemical DTN with \( N \) nodes and expected first contact time \( t_1 \).

It may also be important to have an approximate estimation of the overall elapsed time to deliver a message in an Epidemic DTN (message diffusion time). When \( N \) is large enough, we have \( N(N-1) \approx N^2 \) and \( \frac{2N-2}{N-2} \approx 2 \), and then from (17):

\[
D(t) \approx \frac{N^2}{N + 2^t}
\]

As the model is continuous, when \( 0 < D(t) < 1 \) the last available node has already received the message. In this way, it will be considered that all nodes received the message when \( t = t_e \) and \( D(t) = \frac{1}{k} \), \( k > 1 \), in (19):

\[
kN^2 = N + 2^t_k \Rightarrow 2^t_k = kN^2 - N \approx kN^2 \Rightarrow \frac{t_e}{t_1}
\]

\[
= \log_2 kN^2 \approx \log_2 k + 2 \log_2 N \Rightarrow t_e
\]

\[
\approx t_1 (\log_2 k + 2 \log_2 N)
\]

\(^2\) The code used to implement this procedure is dvar2.py available in http://cabreu.vialink.com.br/mestrado/dtn/simulaes.
From simulation results, the best fit for $t_e$ was obtained making $k = 2$. Then, (20) can be reduced to:

$$t_e \approx t_1(1 + 2\log_2 N)$$  \hspace{1cm} (21)

where $t_1$ is the mean time to the first contact, which is a constant given by (12).

For instance, in a DTN with 40 nodes, the expected message diffusion time (time to deliver a message to almost all nodes) can be calculated as a function of its own first contact $t_{1_{40}}$ as $t_e \approx 11t_{1_{40}}$. For 90 nodes, $t_e \approx 12t_{1_{90}}$ and so forth. Eq. (21) was also compared to simulation results providing again a good fit of the model.

Looking at (21), one may observe that the total time to deliver a message is much more sensible to the first contact time than to the number of nodes. This means that, besides having a fundamental influence on the deviations of the epidemic process, the first contact also has great influence on the total time to deliver a message (time to infect all nodes in an epidemic DTN).

With such analysis, we may conclude that all actions that anticipate first contact times bring improvement to the epidemic process as a whole, and reduce its deviations.

It is important to note that in an Epidemic DTN, as opposed to epidemiological studies in populations, the main goal is to speed up message (virus) spreading in order to infect all nodes in the shortest period of time. Then some possible procedures can be envisaged in the effort to reduce the border effect and the variance on message diffusion of the entire DTN: starting the DTN process with a fast source node, because it is likely to meet more nodes in a shorter period of time; starting the DTN process with a group of $k$ propagating nodes, because this has a similar effect of a DTN after $k$ message forwardings; and starting the DTN process with a group of $k$ fast nodes for the same reasons before.

The above mentioned procedures were evaluated through simulations in order to verify their efficacy. First, we run scenario 1 again for one fast source node with speeds of 15, 20 and 30 km/h. Fig. 10 shows the standard deviation of $D$ for the three cases, it can be seen a significant reduction on those values as long as speed is increased. The reduction achieved up to 35% in the standard deviation for the best case.

In Fig. 11 we repeated the same simulation scenario but now considering two fast source nodes. The general behavior was the same, but a larger reduction in the standard deviation was achieved, about 60%. Such results confirmed...
our expectations on what can be done to enhance Epidemi-
cical DTN performance.

7. Summary of the main DTN model results

It follows below the main DTN model parameters de-
ved so far (see Table 2) taking into account all assump-
tions and simplifications made in this work.

These results are valid under the assumptions that:

- The number of nodes \( N \) is constant during message
  propagation, and is big enough for a good approxima-
tion with the differentiable model equations.
- When a contact occurs between a node in \( P \) and a node
  in \( D \), the message transferring probability is 1.
- When a message is received, the node is immediately
  able to transfer it to all reachable available nodes during
  message lifetime.
- A node makes only one message forwarding at each
time.
- A node makes an average of \( \beta N \) contacts per unit time,
  where \( \beta \) is the contact rate of the network.
- No nodes becomes unavailable \( (\gamma = 0) \).
- The transmission range (coverage) of each node is much
  smaller than the scenario area.

8. Conclusion and future work

This work investigated modeling and analysis of mes-
sage diffusion in Epidemical DTNs. As stated in a recent
survey [18], analytical modeling and performance evalua-
tion for DTN are major research challenges in the field.

Some recent and promising DTN models are the ones
derived from well known SIR epidemiological models,
given the similarity between message forwardings in DTNs
and the spread of diseases in populations. This paper also
follows this line of thought and started revisiting the basic
SIR model and using a special case, where the recovery rate
is zero \( (\gamma = 0) \) to better fit a specific DTN environment that
uses fixed message TTL. From Section 4 onwards some spe-
cific formulations, conclusions and findings were reported.

One of our main findings is that SIR models are quite
accurate for the average behavior of Epidemical DTN, even
in a discrete scenario with a finite and not very large num-
ber of nodes (e.g. 40 in SC1). As it was shown in Section 5,
simulation provided a good fit when average results on \( D \)
(number of available nodes over time) were compared to
expected model results. However, simulation results also
showed that the model may deviate with significant stan-
dard deviations for some specific cases, e.g. at some points
in time during simulations, standard deviations on the number
of available nodes were higher than average values.

In Section 6 we studied such deviations and observed
that they are intrinsic to the system under study (Epidemi-
cical DTN) given its strong dependency to first and last con-
tact times (border effect). Such initial transient behavior on
epidemical spread has to be investigated and treated in or-
der to make the model useful for practical DTN networking
design, evaluation and optimization purposes.

We proposed a simple Monte Carlo method to compute
standard deviations without resorting to complex network
scenario simulations. It was also presented some possible
procedures to mitigate the effects of deviations and speed
up the epidemic process, such procedures were tested and
validated by simulations confirming our expectations. This
result has important impact in real life scenarios, because
it helps to choose, if possible, the best node or nodes to first
transmit a message.

As a future work, we intend to evaluate and extend the
model to more practical scenarios where the problems of
channel propagation, node coverage, spatial distribution
and multi-path fading should be considered, as well as
memory and energy consumption.

References

[1] K. Fall, A delay-tolerant network architecture for challenged
internets, in: Proc. of the 2003 Conference on Applications,
Technologies, Architectures, and Protocols for Computer
Communications, SIGCOMM ’03, ACM, New York, NY, USA, 2003,
p. 27–34.
of Computer Science, Duke University, April, 2000.
http://dx.doi.org/10.1098/rspa.1927.0118.
[6] W. Kermack, A. McKendrick, Contributions to the mathematical
of mobility on the performance of relaying in ad hoc networks, in:
evidence of variability and of persistency of type, Lancet 1 I (1906)
733–739.
[12] H.W. Hethcote, Mathematical Understanding of Infectious Disease
Dynamics, Lecture Notes Series, vol. 16, Institute of Mathematical
Sciences, National University of Singapore, 2000.
[13] K. Fall, A Message- Switched Architecture for Challenged Internets,
model delay-tolerant networks, in: Military Communications
delay tolerant networks, in: Proc. IEEE Fifth Int. Networking,
delay/disruption tolerant networks through passive relay points,
[18] M.J. Khabbaz, C.M. Assi, W.F. Fawaz, Disruption-tolerant
networking: a comprehensive survey on recent developments and
607–640.
[19] Q. Wang, Z.J. Haas, Analytical model of epidemic routing for delay-
tolerant networks, in: Proceedings of the 1st ACM Workshop on


Claudio Abreu is president of Vialink Informatica Ltda., Rio de Janeiro, since 1997. He was graduated in electronic engineering in 1987 and received his master degree in Computer Science in 2012 both from the Military Institute of Engineering (IME). His main research interests include performance analysis of computer networks, Internet traffic engineering, network resilience, graphs theory and delay and disruption tolerant networks.

Ronaldo Salles is appointed professor at the Military Institute of Engineering (IME), Rio de Janeiro, and currently serves the Brazilian Army as a lieutenant colonel. Since 2007 he holds the position of postgraduate course coordinator in the Department of Computer Engineering. He received his master degree in Computer Science from IME in 1998, and his Ph.D. degree in 2004 from the Imperial College London, U.K. His main research interests include performance analysis of computer networks, internet traffic engineering, delay and disruption tolerant networks, network resilience and network security. He received best paper awards at the following conferences: ACM LANC’09, IEEE/IFIP LANOMS’09 and SBRC’08. He has served on the technical program committees of many conferences and has published more than 70 papers in international journals and conference proceedings. He is a member of the IEEE and SBC.