Epidemic SIR Model Applied to Delay-Tolerant Networks

Gabriela Moutinho de Souza Dias and Ronaldo Moreira Salles

Abstract— **This paper presents an application and adaptation of the epidemiological model known as Susceptible-Infected-Recovered to the mathematical modeling of the process of message forwarding in a Delay-Tolerant Network scenario characterized by epidemic routing. Simulation results are compared to the mathematical model indicating a good fit of the approach.**

Keywords— **Delay-Tolerant Networks, mathematical modeling, epidemic routing.**

I. Introduction

Delay-Tolerant Network (DTN) is a wireless network with no infrastructure, whose overlay architecture was defined to perform interoperability between and among *Challenged Networks*, working above existing protocol stacks in various network architectures [8], [6]. Challenged Networks, also known as Intermittently Connected Networks, operate in challenging environments, where frequent link disruptions make end-to-end paths between source and destination uncertain. Such environments are characterized specially by: intermittent connectivity, long delays and high error rates. DTN nodes are capable to overcome these obstacles through a message switching technique known as *store-carry-forward*, in which nodes keep copies of messages stored for future forwarding, once a connection is established [13], [19].

Mathematical modeling of DTN embraces the construction of analytical models and performance evaluation of various routing protocols and different simulation scenarios. This area is the one with the most need of research work. The authors of the most recent survey about DTN [13], which consolidates publications between mid 2007 and June 2010, discussed the research areas, the latest advances and the biggest challenges still remaining. They also noticed analytical modeling and performance evaluation to be one of the most important challenging problems, specially due to the lack of a generic analytical model for DTN.

The main motivation to the study of mathematical modeling of DTN presented in this paper, besides the lack of the area, is inspired by engineering, in the sense of network designs. The main purpose of this study is to search for solutions which can contribute to the construction of an analytical model for DTN, that may help designers and engineers to define important project parameters with certain degree of precision. Moreover, a well adjusted model is a powerful instrument to analyze the behavior of the real system.

Gabriela Moutinho de Souza Dias and Ronaldo Moreira Salles, Computer System Engineering Department, Instituto Militar de Engenharia, Rio de Janeiro-RJ, Brazil, E-mails: gabriela.msdias@gmail.com, salles@ime.eb.br.

There are several proposals of routing protocols to different varieties of DTN. In [21] different types of protocols are presented, according to scenario, deterministic or stochastic, and to the available knowledge about network topology. In stochastic scenario, where the network behavior is aleatory and practically unknown, the simplest protocol corresponds to *epidemic routing* [18], and a DTN of this type is called *Epidemic DTN*.

The process of forwarding messages using epidemic routing in DTN is similar to what happens with the propagation of a virus among humans. In a population, the spread of a infectious disease, like influenza, occurs by transmission of a virus through contact between individuals. An infective individual transmits the virus to healthy individuals (susceptible), whom then become new infected themselves, and also capable of transmitting the disease forward.

In a similar way, a source node in an Epidemic DTN transmits copies of a message to all nodes it encounters (makes contact), and this procedure is also followed by the intermediary nodes, until destination. This can be considered the simplest behavior of a DTN, once there is no information about the network topology, and other protocols are particular cases of it. For this reason, our modeling will be conducted to focus on the behavior of Epidemic DTN.

One of the main references to this work is [20]. In that paper, the authors presented an analytical model for epidemic routing based on an epidemiological model for the spread of infectious diseases, with the purpose of showing how that model "can be advantageously employed to study the performance of various epidemic style routing schemes". Despite of not being explicitly written in [20], the epidemiological model used as a base for that study was the Susceptible-Infected-Recovered (SIR) model [7], [15], [16], which employs an ODE (ordinary differential equation) system.

As it will be presented next, the SIR model has two different approaches to treat an infectious process, deterministic and stochastic. This work adopted the stochastic approach (Stochastic SIR model, Section II-B), while in [20] it was given more attention and detail to the deterministic approach.

This work proposes an application of Stochastic SIR model to Epidemic DTN. The main contribution of this article is the development of a simple and direct adaptation of the Stochastic SIR model to an Epidemic DTN scenario, showing in a clear and objective way how the parameters related to the epidemiological model can be adjusted to a DTN scenario.

The model predictions are compared to simulation results and to predictions obtained from the Deterministic SIR model (Section II-A) adaptation, using an expression presented in [20], in order to evaluate the performance of both determinist and stochastic models.

The remain of this paper is structured as follows. Section II presents the SIR model. Section III shows the applicability of this model to Epidemic DTN. The mathematical modeling itself is presented in Section IV and the results achieved in Section V. Finally, the conclusions and future work are discussed in Section VI.

II. Epidemic SIR Model

Due to similarities between basic forwarding process in an Epidemic DTN and the spread of infectious diseases in a population, the idea of using epidemiological models as base to create a model for DTN seems quite reasonable.

In the area of epidemiology, there is a large amount of references to mathematical models for infectious diseases, such as [1], [2], [3], [4], [7], [12], [17]. Among those models, the renowned SIR model was chosen, as done by [20]. The SIR model has two different approaches, deterministic and stochastic, both presented next based mostly in [1], [7].

A. Deterministic SIR Model

The classical deterministic continuous time epidemic model for the spread of infectious diseases defined by Kermack and McKendrick dates back 1927 and it is still widely used until today [15], [16].

Let N be the total number of individuals in a population. Consider that the duration of an epidemic is much shorter than the lifetime of the population, i. e., the size of the population can be considered fixed, and, births and deaths not related to the disease can be ignored. At any moment of time, each individual find himself in a specific state according to his disease status, dividing the population in three different compartments:

- S - Susceptible: Individuals who are susceptible to be infected;
- I - Infected: Individuals who have been infected with the disease and are capable of spreading it to susceptible individuals;
- R - Recovered: Individuals who have been recovered, becoming immune. In cases where death or isolation due to the disease may occur, these will also be included in this compartment, that will be called Removed.

An infection is represented by the transition from state Susceptible to state Infected, and a recovery by the transition from Infected to Recovered. Only these two transitions are considered possible.

A Susceptible individual becomes infective right after having contact with an Infected individual. The incubation period is considered negligible and all contacts between Susceptible and Infected result in infection. The model considers, also, that, after recovery, individuals become immune, never returning to Susceptible compartment.

Let $x(t)$, $y(t)$, $z(t)$ be the number of susceptible, infected and recovered individuals, respectively, at any moment of continuous time $t \geq 0$. Therefore, for all $t \geq 0$:

$$x(t) + y(t) + z(t) = N \tag{1}$$

Considering that individuals are homogeneously mixed in the population, according to the Law of Mass Action [7], the evolution of the epidemic is defined by the following equations:

$$\frac{dx(t)}{dt} = -\beta x(t)y(t) \tag{2}$$

$$\frac{dy(t)}{dt} = \beta x(t)y(t) - \gamma y(t) \tag{3}$$

$$\frac{dz(t)}{dt} = \gamma y(t) \tag{4}$$

Initial condition: $(x(0), y(0), z(0)) = (x_0, y_0, 0)$

The parameter β is the *infection parameter* and represents the rate of contacts between two individuals. The parameter γ is the *recovery rate* and represents the rate in which Infected individuals are recovered.

B. Stochastic SIR Model

Considering the same assumptions described in the deterministic analysis, let $N + I$ be the fixed number of individuals in a population subdivided in $X(t)$ susceptible, $Y(t)$ infected and $Z(t)$ recovered individuals, with initial condition $(X(0), Y(0), Z(0)) = (N, I, 0)$, and for all $t \geq 0$:

$$X(t) + Y(t) + Z(t) = N + I \tag{5}$$

The stochastic model considers $\{(X(t), Y(t)) : t \geq 0\}$ a Markov process with finite state space, dependent of two random variables, X and Y. Each state is represented by the pair $(X(t), Y(t))$.

Considering, again, the homogeneous distribution of individuals, let δt be a time interval small enough that only one of the three possible events of state transition in the Markov chain may occur within it, with respective infinitesimal transition probabilities:

- Infection - Probability of a new infection event in δt:
$$Pr\{(X,Y)(t+\delta t) = (i-1, j+1) | (X,Y)(t) = (i,j)\} = \beta ij\delta t + o(\delta t) \tag{6}$$
- Recovery - Probability of a new recovery event in δt:
$$Pr\{(X,Y)(t+\delta t) = (i, j-1) | (X,Y)(t) = (i,j)\} = \gamma j\delta t + o(\delta t) \tag{7}$$
- No state transition - Probability of no occurrence of infection or recovery in δt:
$$Pr\{(X,Y)(t+\delta t) = (i, j) | (X,Y)(t) = (i,j)\} = 1 - (\beta i + \gamma)j\delta t - o(\delta t) \tag{8}$$

From the state probabilities

$$p_{ij}(t) = Pr\{(X,Y)(t) = (i,j) | (X,Y)(0) = (N,I)\}, \tag{9}$$

the Kolmogorov forward equations for this process are given by:

$$\begin{aligned} \frac{dp_{NI}(t)}{dt} &= -I(\beta N + \gamma)p_{NI} \\ \frac{dp_{ij}(t)}{dt} &= \beta(i+1)(j-1)p_{i+1,j-1} - j(\beta i + \gamma)p_{ij} \\ &+ \gamma(j+1)p_{i,j+1} \end{aligned} \tag{10}$$

with $0 \leq i + j \leq N + I$, $0 \leq i \leq N$, $0 \leq j \leq N + I$, and initial conditions $p_{NI}(0) = 1$, $p_{ij}(0) = 0$ otherwise.

III. SIR Model Applied to DTN

With the purpose of applying the SIR model to the Epidemic DTN modeling, an analogy is made by defining three states for the network nodes towards a message propagation:

- Available (Susceptible): Nodes that are available to receive a copy of the message;
- Transmitting (Infected): Nodes that have a copy of the message stored and are capable of forwarding it to available nodes;
- Unavailable (Recovered): Nodes that are unavailable to receive or transmit a copy of the message.

At any moment of time, each node find itself in one of these states. An Available node becomes "infected", i. e., receives a copy of the message, right after contact with a Transmitting node, which configures an "infection".

The time period necessary for connection and message transmission between two nodes is considered negligible, i. e., all contacts between Available and Transmitting nodes result in a successful transmission of a message copy, being only necessary for the Available node to be at range of a Transmitting node. The same assumption was made in [20]. It's important to notice that nodes receive one copy of the same message only once.

"Recovery" situations are represented by the Unavailable state, which covers cases of unavailability, both at forwarding and receiving of a message copy. Unavailability at receiving is considered when a Transmitting node, after discarding a message copy due to time limitation (characterized by TTL - time to live), saves that information to reject another copy of that same message, becoming "immune". Unavailability at forwarding is considered when a destination node receives a copy of the message. When it happens, the destination node makes an automatic transition Available $\rightarrow$ Transmitting $\rightarrow$ Unavailable, since, despite it stores a copy of the message, it will not transmit the copy forward. This state could also cover situations of physical destruction of nodes, causing its "death" or permanent "removal" from the network.

IV. Mathematical Modeling

To begin the construction of Epidemic DTN mathematical modeling, the stochastic treatment of SIR model was chosen to be used as a base. This choice comes from the fact that stochastic models are more appropriate to represent random phenomena, such as the spread of infectious diseases and the process of forwarding messages in Epidemic DTN. Besides, in epidemiology literature, deterministic models are usually employed to large populations, such as cities, countries and even at a global level.

According to [7], "When the number of individuals is very large, it is customary to represent the infection process deterministically (...). However, deterministic models are unsuitable for small populations, while in larger populations, the mean number of infectives in a stochastic model may not always be approximated satisfactorily by the equivalent deterministic model."

Moreover, an observation that reinforces this choice is that the result obtained in [20] with the stochastic model, to which the authors referred as "ODE system involving second moments", gives a better adjustment than with the deterministic model.

The model obtained with the adaptation defined in Section III of the Stochastic SIR model, presented in Section II-B, will be denominated *DTN Model*.

DTN Model will then be represented by Eq. (5) to (10), where $X(t)$, $Y(t)$, $Z(t)$ correspond to random variables of Available, Transmitting and Unavailable nodes, respectively. It will also be considered that $(X(0), Y(0), Z(0)) = (N, I, 0)$, and, $(X, Y)(t) = (i, j)$. The parameter β represents the average number of contacts between two nodes per unit time, and will be called *contact rate*. The parameter γ describes the number of Transmitting nodes that become Unavailable per unit time, and will be called *unavailability rate*.

This preliminary version of DTN Model is suitable for Epidemic DTN modeling when the following information is known: mobility model, speed of motion, message TTL, transmission range and area of a limited region where nodes move. Problems or errors at physical and data link layers are not taken into account, and nodes are assumed to have enough storage capacity so buffer overflow is not an issue.

A. Adaptation of Parameters

The parameters related to Stochastic SIR model, as presented, are: number N of Susceptible individuals at $t = 0$, number I of Infected individuals at $t = 0$, time interval δt, infection parameter β and recovery rate γ.

The parameters related to DTN Model will then be: number N of Available nodes at $t = 0$, number I of Transmitting nodes at $t = 0$, time interval δt, contact rate β and unavailability rate γ. The challenge is to convert characteristics and functionality of DTN into values for those parameters. In practical terms, it means placing network parameters, e. g., transmission range and speed, into β and γ.

The first step is to analyze the contact rate β. As well noted by [20], an estimate for β was obtained by [10]. In that paper, the authors showed, for Random Waypoint (RW) and Random Direction (RD) mobility models [5], that the event of contact between two nodes can be modeled by a Poisson distribution, being the inter-meeting time distribution exponential, and the contact rate estimate given by

$$\beta \approx \frac{2wrE[V^*]}{L^2}, \tag{11}$$

where $w \approx 1.3683$ is a constant specific to RW, L^2 is the area where nodes move, r is the range of each node ($r \ll L$) and $E[V^*]$ is the average relative speed between two nodes (the exact expression of $E[V^*]$ can be found in [11]). Eq. (11) is the expression that will be used for the rate β on DTN Model, being the probability of transmission of a message copy (probability of "infection") given by Eq. (6).

The next step is to analyze how to define the probability of "recovery", i. e., of transition of a Transmitting node to Unavailable state, which is given by Eq. (7) on Stochastic SIR model. Assuming that the message TTL is long enough to allow delivery to occur before it ends, there will be no copy discards before delivery. In that case, the unavailability situation to be considered is the one of destination node receiving the message copy, which will represent the first "recovery" of the process. So, the probability of "recovery" on DTN Model is the probability of the destination node having contact with some Transmitting node. Therefore, such probability is given by $\beta j \delta t + o(\delta t)$, which gives $\gamma = \beta$.

B. Validation

The DTN Model validation presented here is performed through comparison between simulation results and model predictions.

To work with a representation of a real network, the simulation tool called The ONE (*The Opportunistic Network Environment*) v. 1.4.1 [14] was employed to create and simulate a scenario as described next. The result obtained represents the Epidemic DTN system that one desires to model analytically.

The scenario setting chosen to be used in the simulations has the following characteristics: epidemic routing, transmission speed of 1 Mbps, buffer size of 50 Mb, RW mobility model, node speed between 4 and 10 km/h, total simulation time of 86400 s (24 h), update interval of 1 s, message TTL of 86401 s, range of 100 m and area of 8 x 8 km ($L \gg r$), where nodes are randomly placed at $t = 0$. The RW mobility model was chosen for presenting better results compared to RD in [10]. For each simulation, there is only one source node in the network, at $t = 0$, carrying one message of 10 bytes, which has to be delivered to one particular destination node.

The parameters related to DTN Model should then be configured according to scenario. Starting with the initial condition, the scenario gives $I = 1$. The time interval δt will be the same one used as update interval for the simulation, in that case, $\delta t = 1s$. According to Eq. (11), $\beta \simeq 1.03335 \cdot 10^{-5}$, and finally, $\gamma = \beta$.

Following the approach presented for Stochastic SIR model, a code in C language was created to execute 100000 (one hundred thousand) sample paths on the Markov chain that represents the process of infection and recovery, over time, applying the Monte Carlo method to choose among possible events (infection, recovery or no state transition) according to each transition probability. The parameters for this model were, then, adapted, as explained in Section IV-A, to make this modified SIR model, named DTN Model, suitable to an Epidemic DTN system.

With the purpose of having a metrics of comparison between DTN Model predictions and simulation results, the absolute difference between the number of Available nodes predicted by the Model and the expected value, obtain by simulation, was calculated, for every instant of time. The average value of that difference in relation to the simulation value will be called *relative error*. The same procedure was taken to Transmitting nodes. In addition, the 98% confidence interval for the relative errors was also calculated.

V. RESULTS

The Epidemic DTN scenario described in Section IV-B was simulated with The ONE and modeled with the code developed for DTN Model, varying the total number of nodes from 10 to 200.

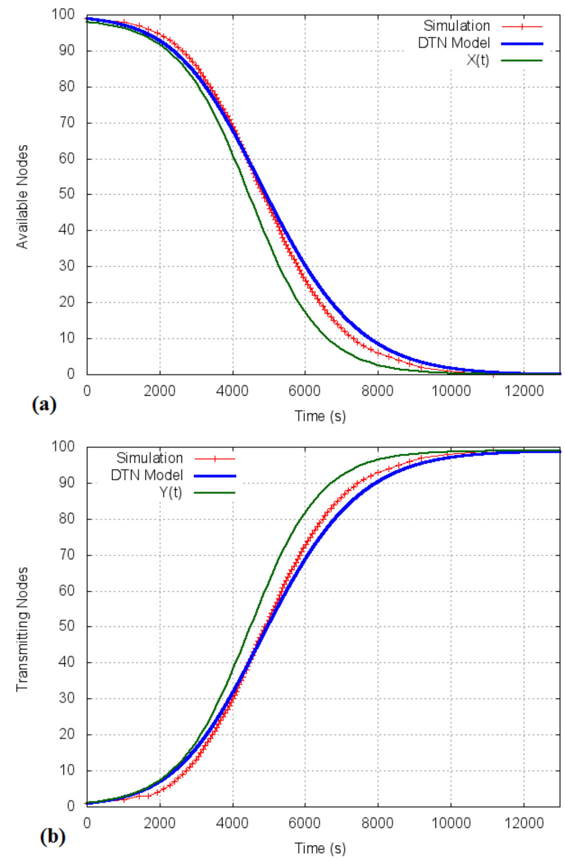

Fig. 1. Results for Epidemic DTN simulation and modeling, for the scenario described in Section IV-B with 100 nodes. (a) Available nodes. (b) Transmitting nodes.

With the purpose of comparing both SIR model approaches, deterministic and stochastic, the expression for the number of Transmitting nodes over time, obtained by [20] from a Deterministic SIR model adaptation, was also implemented in the code. Such expression is given by

$$Y(t) = \frac{N}{1 + (N-1)e^{-\beta N t}}, \tag{12}$$

where N is the total number of nodes excluding the destination. As already mentioned in the beginning, the authors of [20] work with an adaptation of SIR model. From Eq. (12), and knowing that before TTL no Unavailable node will exist among those N nodes, i. e., $X(t)+Y(t) = N$, it is possible to obtain the expression for the number of Available nodes over time, which is given by

$$X(t) = \frac{N(N-1)}{N - 1 + e^{\beta N t}} \tag{13}$$

Fig. 1 shows the results obtained for the modeling of Available and Transmitting nodes, for a total number of 100 nodes, using the Deterministic SIR model adaptation represented by Eq. (12) and (13), and DTN Model. By observing the shape of the curves in Fig. 1, one can notice that the epidemiological

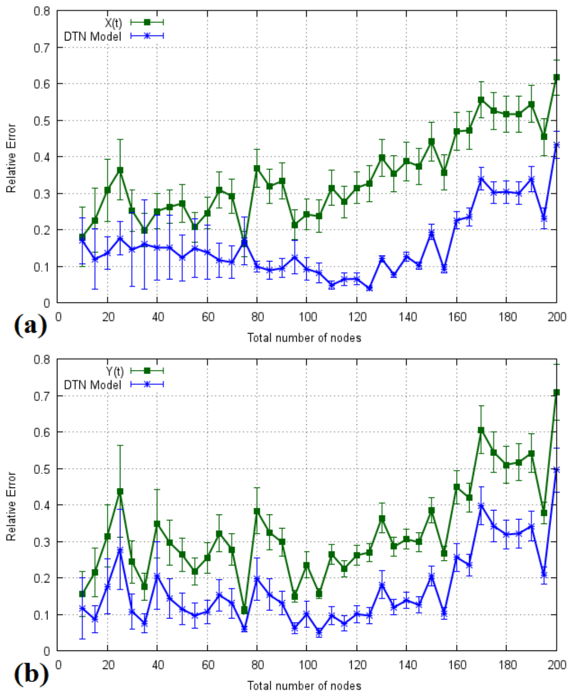

Fig. 2. Relative error for Epidemic DTN modeling for the scenario described in Section IV-B, with 98% confidence intervals. (a) Available nodes. (b) Transmitting nodes.

SIR model represents in a very similar way the behavior of message forwarding process using epidemic routing in DTN.

Fig. 2 shows the comparison between relative errors of deterministic model and DTN Model, with 98% confidence intervals, for the modeling of Available and Transmitting nodes.

The error values in Fig. 2 indicate that, as obtained in [20], the DTN Model (stochastic) presents average relative error smaller than the deterministic model, demonstrating a better adjustment to the system modeled, confirming the expectation. For a smaller number of nodes, however, the confidence intervals intercept themselves, so it is not possible to compare the performances of the models in those cases, specially for Available nodes modeling. On the other hand, from 100 to 200 nodes, it is possible to affirm that the error for DTN Model is smaller than for deterministic model.

VI. Conclusion

This paper presented the development of a mathematical model for Epidemic DTN based on the stochastic approach of epidemiological SIR model, which was named DTN Model. The parameters of SIR model were adapted to consider characteristics of an Epidemic DTN. Results obtained through simulation and modeling in a specific scenario were presented, along with the comparison between deterministic and stochastic analysis for the modeling.

The results presented give a prosperous indication that the epidemiological Stochastic SIR model can be adapted and adjusted to the construction of an analytical model to Epidemic DTN.

The DTN Model presented here is a preliminary version that may suffer future adjustments to treat some assumptions made, such as the inclusion of other unavailability situations for nodes state transition. One interesting direction for future work is to adjust the model parameters, or even to create an extension to it, in order to aggregate more information about the network and possible scenarios.

Acknowledgments

We thank the Instituto Militar de Engenharia (IME) for giving the necessary conditions for the accomplishment of this work.

References

[1] L. J. S. Allen and A. M. Burgin, "Comparison of deterministic and stochastic SIS and SIR models in discrete time", *Mathematical Biosciences*, Elsevier, v. 163, pp. 1–33, 2000.
[2] L. J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology*. Pearson Prentice Hall, Population English Edition, 2003.
[3] N. T. J. Bailey, "A simple stochastic epidemic", *Biometrika*, v. 37, pp. 193–202, 1950.
[4] N. T. J. Bailey, "The Total Size of a General Stochastic Epidemic", *Biometrika*, v. 40, pp. 177–185, 1953.
[5] C. Bettstetter, "Mobility modeling in wireless networks: categorization, smooth movement and border effects", *ACM SIGMOBILE Mobile Computing and Communications Review*, ACM Press, v. 5, pp. 55–66, 2001.
[6] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall and H. Weiss, "Delay-Tolerant Networking Architecture", IETF, RFC4838, v. 54, p. 258, 2007.
[7] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge University Press, 2001.
[8] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets", *Proceedings of the 2003 Conference on Applications Technologies Architectures and Protocols for Computer Communications SIGCOMM 03*, ACM Press, v. 25, p. 27, 2003.
[9] N. C. Grassly and C. Fraser, "Mathematical models of infectious disease transmission", *Nature Reviews Microbiology*, v. 6, p. 477–487, 2008.
[10] R. Groenevelt, P. Nain and G. Koole, "The message delay in mobile ad hoc networks", *Proceedings of the 24th International Symposium on Computer Performance, Modeling, Measurements and Evaluation*, v. 62, pp. 210–228, 2005.
[11] R. Groenevelt, *Stochastic Models for Mobile Ad Hoc Networks*, Universit de Nice - Sophia Antipolis, 2005.
[12] A. Kaddar, A. Abta and H. T. Alaoui, "A comparison of delayed SIR and SEIR epidemic models", *Nonlinear Analysis: Modelling and Control*, v. 16, pp. 181–190, 2011.
[13] M. J. Khabbaz, C. M. Assi and W. F. Fawaz, "Disruption-Tolerant Networking: A Comprehensive Survey on Recent Developments and Persisting Challenges", *IEEE Communications Surveys & Tutorials*, v. 14, pp. 607–640, 2012.
[14] A. Keranen, J. Ott and T. Karkkainen, "The ONE Simulator for DTN Protocol Evaluation", *SIMUTools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, 2009.
[15] W. O. Kermack and A. G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics", *Proceedings of the Royal Society of London Series A*, v. 115, pp. 700–721, 1927.
[16] W. O. Kermack and A. G. McKendrick, "Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity", *Proceedings of the Royal Society of London Series A*, v. 138, pp. 55–83, 1932.
[17] M. G. Roberts and J. A. P. Heesterbeek, "Mathematical models in epidemiology", *Mathematical models*, UNESCO-EOLSS, 2004.
[18] A. Vahdat and D. Becker, "Epidemic Routing for Partially-Connected Ad Hoc Networks", *Technical Report CS-2000-06*, Duke University, 2000.
[19] F. Warthman, "Delay-Tolerant Networks (DTNs): A Tutorial", *Technical Report*, May 2003.
[20] X. Zhang, G. Neglia, J. Kurose and D. Towsley, "Performance modeling of epidemic routing", *Computer Networks*, v. 51, pp. 2867–2891, 2007.
[21] Z. Zhang, "Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges", *IEEE Communications Surveys & Tutorials*, v. 8, pp. 24–37, 2006.