Hernández Orallo, E.; Serrat Olmos, MD.; Cano Escribá, JC.; Tavares De Araujo Cesariny Calafate, CM.; Manzoni, P. (2014). A Fast Model for Evaluating the Detection of Selfish Nodes Using a Collaborative Approach in MANETs. Wireless Personal Communications. 74(3):1099-1116. doi:10.1007/s11277-013-1346-y.

Noname manuscript No.
(will be inserted by the editor)

A Fast Model for Evaluating the Detection of Selfish Nodes Using a Collaborative Approach in MANETs

Enrique Hernández-Orallo , Manuel D. Serrat Olmos, Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni

the date of receipt and acceptance should be inserted later

Abstract Mobile ad-hoc networks (MANETs) rely on network cooperation schemes to work properly. Nevertheless, if nodes have a selfish behaviour and are unwilling to cooperate, the overall network performance could be seriously affected. The use of watchdogs is a well-known mechanism to detect selfish nodes.

In this paper we propose a collaborative watchdog approach, which is based on the fast diffusion of selfish nodes awareness. Then, we introduce an analytical model to evaluate the time of detection and the overhead (number of messages) of our collaborative watchdog approach for detecting one selfish node. This model is extended for the case of several selfish nodes, including a mean-max approximation for a feasible computation when the number of selfish nodes is high. The results show that a collaborative watchdog is a very efficient approach since the detection time of selfish nodes is reduced, and the overall overhead is very low.

Keywords Wireless Networks · MANETs · Performance Evaluation · Selfish nodes

1 Introduction

A Mobile ad-hoc network (MANET) is a network of mobile nodes connected by wireless links without using any pre-existent infrastructure. Each node is free to move independently in any direction and can directly communicate with

Enrique Hernández-Orallo (corresponding author), Manuel D. Serrat Olmos, Juan-Carlos Cano, Carlos T. Calafate, Pietro Manzoni
Departamento de Informática de Sistemas y Computadores.
Universidad Politécnica de Valencia. 46022 Valencia, Spain.
E-mail: ehernandez@disca.upv.es
Tel.: +34-96-3877000 Fax: +34 96 3877579

each other if a contact occurs (that is, if they are within communication range). Opportunistic and Delay Tolerant Networks (DTNs) constitute an emerging subclass of MANETs where there are only intermittent connectivity and opportunistic contacts. Opportunistic nodes collectively form dynamic networks that are built from short unpredictable contact times as nodes move in and out of connectivity [15]. Unlike mobile ad hoc networks, which aim at offering a frequently available connected path through a dynamic network, opportunistic networks only offer a store and forward service in a mostly disconnected network comprised of infrequent contact times between nodes. Applications of such networks include vehicular ad hoc networks (VANETs) or mobile social networks.

MANETs and DTNs nodes must forward traffic unrelated to their own use. That is, these networks rely on network cooperation schemes to work properly. Nevertheless, in the real world, most nodes have a selfish behaviour and are unwilling to forward packets for others. Additionally, some nodes can exhibit malicious behaviour trying to disturb the normal network behaviour, and others can be faulty nodes. In all the cases these *misbehaving* nodes will not cooperate in the transmission of packets. Therefore, detecting such nodes is essential for the overall network performance. Watchdogs are appropriate mechanisms to detect misbehaving and selfish nodes in computer networks. Essentially, watchdog systems overhear wireless traffic and analyse it to decide if the neighbours nodes are behaving in a selfish manner [7].

Several works studied the impact of node selfishness in MANETs. A first study of misbehaving nodes and the proposal to use watchdogs to detect them was introduced in [14]. This work proposed a Watchdog and Pathrater over the DSR protocol to detect non-forwarding nodes, maintaining a rating for every node. In [16] another scheme for detecting selfish nodes based on context aware information was proposed. The CONFINDENT protocol was proposed in [2], which combines a watchdog, reputation systems and bayesian filters from the node and its neighbours to securely detect misbehaving nodes. A Mobile Intrusion Detection System is described in [11] as an advanced watchdog. In [5] an analytical selfish model (which is tied specifically to a routing protocol) is proposed. A recent survey [17] shows the impact of selfish nodes on the performance on ad-hoc mobile networks and reviews some of the proposed solutions presented. Recent papers have focused on DTNs. In [10], the author introduces a model for DTN data relaying schemes under the impact of node selfishness. A similar approach is presented in [13] that shows the effect of socially selfish behaviour. Social selfishness is an extension of classical selfishness (also called *individual selfishness*). A social selfish node can cooperate with other nodes of the same group and it does not cooperate with other nodes outside the group. Social selfishness in DTNs has been studied in [12].

Although some of the aforementioned papers (such as [2, 16]) introduced some degree of collaboration on their watchdog schemes, the diffusion is very costly, since they are based on sending periodic messages. This paper presents an efficient approach to reduce the detection time of selfish nodes using *collaborative watchdogs* based on contact dissemination. If one node has previously

detected a selfish node it can transmit this information to other nodes when a contact occurs. This way, nodes have second hand information about the selfish nodes in the network, thereby reducing the detection time.

In order to evaluate the performance of our collaborative watchdog we introduce an analytical performance model. The problem of using network simulators (such as ns-2 or ns-3) is that evaluating different combinations can be a very time consuming process. If the goal is to evaluate the influence of the number of nodes, the parameters of the watchdog, or the degree of collaboration, we must repeat simulations several times in order to obtain confident results. Using Markov Chain based models, such as the ones presented in this paper, the results are obtained in a faster way. Recent works has shown that the occurrence of contacts between two mobile nodes follows a Poisson distribution λ [3, 4, 9, 18]. This has been shown valid for both human and vehicles mobility patterns. Assuming that the contact occurrence follows and exponential distribution enables an analytical model based on Markov chains. Therefore, the network model only assumes a given contact rate between nodes and, therefore, it is suited for both MANETs and DTNs. The main difference is the rate of contact, that is higher in MANETs and very low in DTNs.

The network is modelled as a set of wireless mobile nodes including collaborative nodes and selfish nodes. The collaborative nodes have a watchdog that can detect a selfish node with a given probability (of detection). When a contact occurs between two collaborative nodes, the information about the selfish(s) node(s) is transmitted with a given degree of collaboration (from no collaboration to full collaboration). Using λ as the contact rate, we model the network as a continuous time Markov chain (CTMC). Using this CTMC we derive expressions to obtain the time of detection (that is, the time when nodes know about the selfish(s) node(s)) and the cost (the number of messages transmitted between collaborative nodes). In this paper we present two models, a first model of a network with only one selfish node, and a second model for the case of several selfish nodes.

These models were validated using simulation and it is shown that they are very precise. Finally, the results of the model are evaluated using a contact rate extracted from a human mobility patterns [8]. Using the models we evaluated the performance of the collaborative watchdogs approach. The results show a significant reduction of the detection time of selfish nodes with a reduced overhead. For example, an overall detection time of 442 hours with no collaboration between nodes is reduced to 3.7 hours with full collaboration with an overhead of 210 messages. Furthermore, this reduction is also significant with a moderate degree of collaboration.

Security concerns, such as, malicious nodes that spread false information about selfish nodes, are outside the scope of this paper.

The rest of the paper is organized as follows. We introduce the collaborative watchdog in Section 2. Then, Section 3 presents a performance model for evaluating the collaborative watchdog using a CTMC. First we derive a basic model for one selfish node and then this model is extended to a network with several selfish nodes. The validation of the correctness of the model is described

in Section 4. Section 5 presents the evaluation of our collaborative watchdog. Finally, section 6 presents some concluding remarks.

2 A Collaborative Watchdog Approach

Network monitoring is a common technique to detect selfish and misbehaving nodes. A node's watchdog consists on overhearing the packets transmitted and received by its neighbours in order to detect anomalies, such as the ratio between *packets received* to *packets being retransmitted* [6]. When a watchdog detects a selfish node, it is marked as a *positive*. The diffusion of these *positives* in the network is the foundation of our *collaborative watchdog* approach.

Formally, the network is modeled as a set of N wireless mobile nodes, with C collaborative nodes and S selfish nodes ($N = C + S$). Initially, the collaborative nodes have no information about the selfish nodes (there are no positives). A collaborative node can have a *positive* when a contact occurs in the following way (see figure 1):

- *Selfish contact*: one of the nodes is the selfish node. Then, the collaborative node *can* detect it using its watchdog and have a positive of this selfish node. Nevertheless, a contact does not always implies a detection. To model this fact, we introduce a probability of detection (p_d). This probability depends on the effectiveness of the watchdog and the type of contact (for example if the contact time is very low, the watchdog does not have not enough information to evaluate if a node is selfish).
- *Collaborative contact*: both nodes are collaborative. Then, if one of them has one or more positives, it *can* transmit this information to the other node; so, from that moment, both nodes have these positives. As in the *selfish contact* case, a contact does not always imply a collaboration. We model this with the probability of collaboration (p_c). The degree of collaboration is a global parameter of the network to be evaluated. This value is used to reflect that either a message with the information about the selfish nodes is lost or that a node temporally does not collaborate (for example, due to a failure or simply because it is switched off). In real networks, full collaboration ($p_c = 1$) is almost impossible.

The detection of contacts between nodes it straightforward using the node's watchdog. Notice that the watchdog is overhearing the packets of the *neighbourhood*; thus, when it starts receiving packets from a new node, it is assumed to be a new contact. Concerning the transmission of information about the *positive* states when a collaborative contact occurs, the transmission cost depends on the protocol used. There are two options:

1. *Single protocol*: A node transmits one message for each positive it knows. For example, in a network with three selfish nodes, if a node has a positive of two selfish nodes, it transmits two messages. This *protocol* is very inefficient in terms of number of messages.

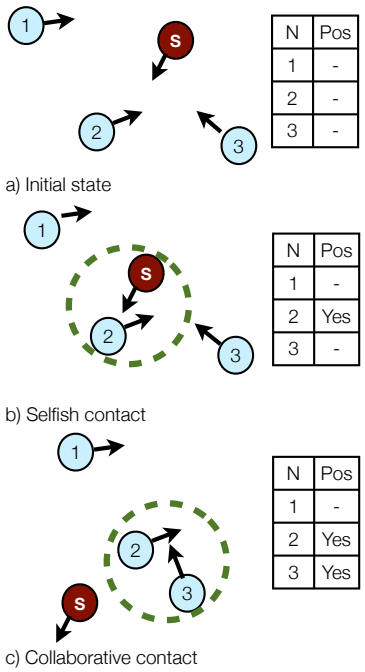


Fig. 1: Detection and transmission of the positives

2. *Group protocol*: A node transmits only one message for all positives it knows. In this case, if a node has a positive of two (or more) selfish nodes, it transmits only one message.

Note that the *group protocol* is only for $S > 1$.

Although defining a reaction scheme when a selfish node is detected is outside the scope of this paper, there are basically two approaches in the literature: isolation and incentivation. Isolation methods are intended to keep the misbehaving nodes outside of the network, excluding them from all kinds of communication. Incentivation methods try to convince the selfish nodes to change their behaviour, and become collaborative instead of selfish, using a virtual payment scheme or a similar mechanism.

3 System Model

The network is modeled as a set of N wireless mobile nodes, with C collaborative nodes and S selfish nodes ($N = C + S$). The collaborative nodes are divided in two sets: a set with D *destination* nodes and a set of $C' = C - D$ *non-destination* nodes. Using this model, our goal is to obtain the time that a set of D destination nodes needs to realize about who is(are) the selfish node(s) in the network. This time value is the detection time for all nodes D.

Therefore, we can evaluate the detection time from one node ($D = 1$) to all the nodes ($D = C - S$). The overhead is the number of messages transmitted up to the detection time.

In all the models, it is assumed that the occurrence of contacts between two nodes follows a Poisson distribution λ. This assumption has been shown to hold in several mobility scenarios of both human and vehicles [3, 4, 18]. For example, in [4] it is shown that for random waypoint and random direction mobility models the parameter λ is related to the mean speed of nodes v, through the following empirical expression $\lambda \approx \frac{8wrv}{\pi l^2}, \quad r \ll l$, where r is the communication range and l is the side of the square network area. The constant w is 1 for the random direction mobility model and 1.3683 for the random waypoint mobility model.

Therefore, the performance model only assumes a given contact rate between nodes and therefore it is suited for both MANETs and DTNs. The main difference is the rate of contact, which is higher in MANETs and lower in DTNs.

In the following subsection, we first derive a simple model for one selfish node ($S = 1$). This model is extended in the following subsection for the case of various selfish nodes ($S > 1$).

3.1 A model for one selfish node

Using λ we can model the network using a 2D Continuous Time Markov chain (2D-CTMC) with states $(d(t), c(t))_{t \geq 0}$, where $c(t)$ represents the number of *non-destination* collaborative nodes with a positive of the selfish node at time t and $d(t)$ represents the number of *destination* collaborative nodes has positives. At the beginning no node has a positive. Then, when a contact occurs, $d(t)$ and $c(t)$ can be increased by one. The final (absorbing) state is when $d(t) = D$. A 2D-CTMC model is used, with an initial state $s_1 = (0, 0)$, $(D - 1)(C' + 1)$ transient states (from $s_1 = (0, 0)$ to $s_\tau = (D - 1, C')$ states) and $C' + 1$ absorbing states (from $s_{\tau+1} = (D, 0)$ to $s_{\tau+\upsilon} = (D, C')$[1]. We define τ as the number of transient states ($\tau = D(C' + 1)$) and the number of absorbing states as υ, that is $C' + 1$. This model can be expressed using the following transition matrix P in canonical form:

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{1}$$

where $\mathbf{I}$ is a $\upsilon \times \upsilon$ identity matrix, $\mathbf{0}$ is a $\upsilon \times \tau$ zero matrix, $\mathbf{Q}$ is a $\tau \times \tau$ matrix with elements p_{ij} denoting the transition rate from transient state s_i to transient state s_j and $\mathbf{R}$ is a $\tau \times \upsilon$ matrix with elements p_{ij} denoting the transition rate from transient state s_i to the absorbing state s_j.

Now, we derive the transition rates p_{ij}. Given the state $s_i = (d, c)$[2], the following transitions can occur:

[1] Note that each state number i can mapped as $i = d(t) \cdot (C + 1) + c(t) + 1$

[2] For simplicity, we omit the time in the states (that is $(d, c) = (d(t), c(t))$

- (d, c) to $(d, c+1)$: This case takes place when a new *non-destination* collaborative node has a positive of the selfish node. The transition probability is $t_c = (\lambda p_d + \lambda p_c(c+d))(C' - c)$. The term λp_d represents the probability of detection of a selfish node (using the watchdog) and $\lambda p_c(c+d)$ the probability of transmission of the information of the selfish node (it depends on $c + d$, so this probability is higher if more nodes have a positive). Finally, the factor $(C' - c)$ represents the number of pending nodes.
- (d, c) to $(d+1, c)$: This case is when a new destination node has a positive and the transition probability is $t_d = (\lambda p_d + \lambda p_c(c + d))(D - d)$.
- (d, c) to (d, c): This is the probability of no changes and is $t_0 = 1 - t_c - t_d$.

For example, for $N = 4$, $S = 1$ and $D = 1$ we have $C' = 2$, so $\tau = 3$ and $\upsilon = 3$, the transition matrix is:

s_i, s_j	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
(0,0)	t_0	t_c	0	t_d	0	0
(0,1)	0	t_0	t_c	0	t_d	0
(0,2)	0	0	t_0	t_c	0	t_d
(1,0)	0	0	0	1	0	0
(1,1)	0	0	0	0	1	0
(1,2)	0	0	0	0	0	1

Using the transition matrix P we can derive two different expressions: one for the detection time T_d and another for the overall overhead (or cost) M_d. We start with the detection time. From the 2D-CTMC we can obtain how long will it take for the process to be absorbed. Using the fundamental matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$, we can obtain a vector $\mathbf{t}$ of the expected time to absorption as $\mathbf{t} = \mathbf{N}\mathbf{v}$, where $\mathbf{v}$ is a column vector of ones ($\mathbf{v} = [1, 1, \ldots, 1]^T$). Each entry t_i of $\mathbf{t}$ represents the expected time to absorption from state s_i. Since we only need the expected time from state $s_1 = (0, 0)$ to absorption (that is, the expected time for all nodes D to have a positive), the detection time T_d, is:

$$T_d = E[T] = \mathbf{v_1}\mathbf{N}\mathbf{v} \tag{2}$$

where T is a random variable denoting the detection time for all nodes D and $\mathbf{v_1} = [1, 0, \ldots, 0]$.

Concerning the overhead we need to obtain the number of transmitted messages for states. During state s_1 no node has a positive: $s_1 = (0, 0)$. In this state, no messages are transmitted and $m_1 = 0$. The second state s_2 starts when a *non-destination* collaborative node has a positive: $s_2 = (0, 1)$ (that is, there is one sender). In this case, this positive can be transmitted to all nodes (except itself) for the duration of this state (denoted as f_2) with a rate λ and probability p_c. Then, the expected number of messages can be obtained as $m_2 = f_2\lambda(C - 1)p_c$, where f_2 is the duration of state 2. For the following states s_i the number of messages depends on the values of D and C'. For example, if $C' = 2$, then $s_3 = (0, 2)$ and the number of nodes with a positive is 2, but for $C' = 1$, then $s_3 = (1, 0)$, meaning that only one node has a positive. Therefore, we need to calculate the number of positives from

each state $s_i = (d, c)$ in order to obtain the number of senders. This value is merely the sum of c and d. Then, the number of messages for each state is $\Phi(s_i) = c + d$.

We can obtain the duration of each state s_i using the fundamental matrix $\mathbf{N}$. By definition, the elements of the first row of $\mathbf{N}$ are the expected times in each state starting from state 0. Then, the duration of state s_i is $f_i = \mathbf{N}(1, i)$. Summing up, the cost of transmission (or the expected number of messages) is:

$$M_d = E[M] = \lambda(C-1)p_c \sum_{i=1}^{\tau} \Phi(s_i)\mathbf{N}(1,i) \tag{3}$$

3.2 A model for various selfish nodes

We have derived expressions for the detection time and cost for $S = 1$. We can now extend the previous model to the case of various selfish nodes ($S > 1$). The solution is based on using a Continuous Time Markov Chain with $2 \times S$ dimensions. We start with $S = 2$, so we have a four-dimensions CTMC (for short, a 4D-CTMC). Each state s_i now has four values $(d_2(t), d_1(t), c_2(t), c_1(t))_{t\geq 0}$, where $c_1(t)$ and $d_1(t)$ represent the number of *non-destination* and *destination* collaborative nodes that have a positive for selfish node 1 and $c_2(t)$ and $d_2(t)$ is the same for selfish node 2. At the beginning no node has a positive. Then, when a contact occurs, $c_1(t)$,$c_2(t)$,$d_1(t)$ and $d_2(t)$ can increase by one. The final (absorbing) state is when $(d_2(t), d_1(t)) = (D, D)$. This 4D-CTMC has $\tau = D^2(C'+1)^2$ transient states and $\upsilon = (C'+1)^2$ absorbing states. This can be expressed using the transition matrix in canonical form (equation 1).

Now, we derive the transition rates p_{ij}. Given the state $s_i = (d_2, d_1, c_2, c_1)$, the following transitions can occur with transition rates:

- (d_2, d_1, c_2, c_1) to $(d_2, d_1, c_2, c_1 + 1)$: $t_{c_1} = (\lambda p_d + \lambda p_c(c_1 + d_1))(C' - c_1)$.
- (d_2, d_1, c_2, c_1) to $(d_2, d_1, c_2 + 1, c_1)$: $t_{c_2} = (\lambda p_d + \lambda p_c(c_2 + d_2))(C' - c_2)$.
- (d_2, d_1, c_2, c_1) to $(d_2, d_1 + 1, c_1, c_1)$: $t_{d_1} = (\lambda p_d + \lambda p_c(c_1 + d_1))(D - d_1)$.
- (d_2, d_1, c_2, c_1) to $(d_2 + 1, d_1, c_1, c_1)$: $t_{d_2} = (\lambda p_d + \lambda p_c(c_2 + d_2))(D - d_2)$.
- (d_2, d_1, c_2, c_1) to (d_2, d_1, c_2, c_1): $t_0 = 1 - t_{c_1} - t_{c_2} - t_{d_1} - t_{d_2}$.

and using equation 2 we can obtain the detection time, T_d.

This model can be extended to the case of $S > 2$. We will have $\tau = D^S(C'+1)^S$ transient states and $\upsilon = (C'+1)^S$ absorbing states. For each state $s_i = (d_S, d_{S-1}, \ldots d_2, d_1, c_S, c_{S-1}, \ldots c_2, c_1)$, the transition rate from c_j to $c_j + 1$ is $t_{c_j} = (\lambda p_d + \lambda p_c(d_j + c_j))(C' - c_j)$ and d_j to $d_j + 1$ is $t_{d_j} = (\lambda p_d + \lambda p_c(d_j + c_j))(D - d_j)$.

For the cost, we derive two different expressions for the single and group protocols. For the *single protocol* it is easy to obtain the number of messages. For each state $s_i = (d_S, d_{S-1}, \ldots d_2, d_1, c_S, c_{S-1}, \ldots c_2, c_1)$ the number of possible senders is simply the sum of all the pair values (d_j, c_j). Then, the number

of messages for each state is:

$$\Phi(s_i) = \sum_{j=1}^{S} (c_j + d_j) \tag{4}$$

and using equation 3 where $\mathbf{N}$ is the fundamental matrix of the S dimensions CTMC obtained for $S > 1$ we can obtain the cost for the *single protocol* (M_{ds}).

For the *group protocol* we must obtain all the possible combinations. For example, for a network with $S = 2$, $C = 3$, and $D = 1$ in the state $s_7 = (1, 2)$ there are two combinations:

1. two nodes with a positive about selfish node one, and one node with a positive about selfish node two, so in this combination there are three senders.
2. one node with a positive about selfish node one and two; one node with a positive about selfish node one, and one node with no positives. In this combination there are two senders.

For this state, the first combination has a probability of $2/3$ and the second one of $1/3$. So, the mean number of senders is $2/3 \cdot 3 + 1/3 \cdot 2 = 8/3$. Obtaining all the combinations when S is high can be very complex. A simple approximation is based on bounding the value of senders. It is easy to see that the number of senders in each state is between the maximum of $c_j + d_j$ and the minimum between the sum of $c_j + d_j$ and C. So, the possible number of senders in each state is between:

$$\max(s_i) \leq \Phi(s_i) \leq \min(\text{sum}(s_i), C) \tag{5}$$

where $\max(s_i) = \max_{j=1}^{j=S}(c_j, d_j)$ and $\text{sum}(s_i) = \sum_{j=1}^{S}(c_j + d_j)$ Then, we approximate the number of messages $\Phi(s_i)$ by calculating the average of the lower and the upper bound. Finally, the number of messages (M_d) is obtained using equation 3.

The problem with the CTMC model for $S > 1$ is that the number of states increases exponentially with S, so it can be computationally intractable for $S > 3$. For example with $C = 30$ and $S = 3$ we have $30^3 = 27000$ states, and so we need a 27000×27000 matrix. Even using the sparse matrix functionality of Matlab it can take a lot of time (and memory) to obtain the result. Therefore, for large values of N and S, we need another solution.

3.3 Mean max approximation

In this subsection we present approximations for the time and cost of detection that are computationally efficient. The solution is based on the partition of the network. Detecting S selfish nodes in a network is the same as obtaining the maximum of the detection times of S networks with C collaborative nodes and 1 selfish node. Statistically, using the expected time $E[T]$ (equation 2) for a network with C collaborative nodes and $S = 1$, the expected time for knowing

about all the selfish nodes is the expected value of the maximum of a set of S random variables, $T_{max} = \max\{T_1, T_2, \ldots, T_S\}$ with the same distribution. Each random variable T_i has, by definition, a *phase-type distribution*. The problem of finding the mean value of the maximum of phase-type distribution is also a complex problem [1]. The distribution of the maximum of two phase-type distributions can be described by the time to absorption of the Kronecker sum of the respective absorbing Markov chains, and which, again, describes an absorbing Markov chain. Unfortunately, this computation leads to a state-space explosion, so we have the same problem of the CTMC model for $S > 1$.

An approximation to the mean value of the maximum is to assume that all randoms variables T_i has an uniform distribution with mean $E[T]$. Then, the expected value of the maximum is simply:

$$\hat{T}_d \approx E[T_{max}] = \frac{2S}{S+1} E[T] \tag{6}$$

For the transmission cost, we have an approximation for both protocols:

1. *Single protocol*: It has a simple solution. Using equation 6 we need to calculate the number of messages transmitted in each partition for the new expected time of detection $\hat{T}_d$. This is the same as assuming that each phase duration f_i is the mean value of the maximum, and so its value is increased by:

$$\hat{f}_i \approx \frac{2S}{S+1} f_i = \frac{2S}{S+1} \mathbf{N}(1, i) \tag{7}$$

   By grouping terms in the sum of equation 3, we have that in each partition the number of messages is $\hat{M}_d^S \approx \frac{2S}{S+1} M_d$. Then, this value is multiplied by the number of partitions S, and so we have:

$$\hat{M}_{ds} \approx S \frac{2S}{S+1} M_d \tag{8}$$

2. *Group protocol*: We can make a similar approximation for each phase. First, the duration of each phase is increased by expression 7. Then, we need to estimate the number of senders in each phase. A simple approximation is to use the upper bound of expression 5. If for a network with $S = 1$ we have $(i-1)$ senders in phase i, then, for S networks, we have that the sum is $(i-1)S$. Thus, the number of senders is:

$$\Phi(s_i) = \min((i-1)S, C) \tag{9}$$

   Finally, the expected number of messages is:

$$\hat{M}_{dg} \approx \frac{2S}{S+1} \lambda (C-1) p_c \sum_{i=1}^{\tau} \Phi(s_i) \mathbf{N}(1, i) \tag{10}$$

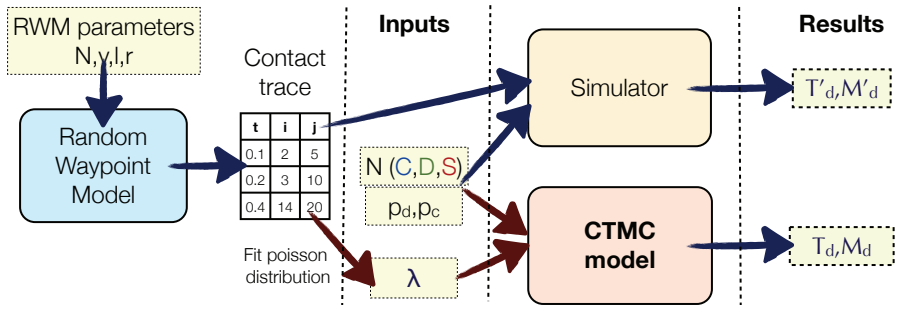


Fig. 2: Validation of the model

4 Model validation

In this section we describe the validation process of the models presented in section 3. In order to validate these models, the results obtained with the models were compared with the simulation results. We implemented all the models and the simulator in Matlab. The simulator is a simple event driven simulator. The network model of this simulator has C collaborative nodes, D destination nodes and S selfish nodes. This simulator generates contact events with a given λ rate. All the nodes have a vector of size S that stores the information about each selfish node. This vector is initialized with no state info and it can change to a positive state. When a contact event occurs, it implements the behavior of the different models, using the probabilities of detection (p_d) and collaboration (p_c) to change the state of a node. The simulation finishes when all the destination nodes have a positive for all the selfish nodes.

The model obtains the time and overhead (T_d,M_d) from a set of inputs: the rate of contacts (λ), the network (N,C,D,S) and the watchdog (p_c,p_d). The correctness of the model was validated by comparing the results obtained from the model with simulation results (see figure 2). We used a random waypoint model (RWP) generator to create a contact trace, which is used, on the one hand to fit the λ value that is used in our performance model and on the other hand to simulate the contacts to obtain the simulation results. The tests have different parameter values that are randomly generated within a pre-defined range. Each simulation was repeated 1000 times in order to obtain a reliable mean value for the detection time and cost ($\bar{T}_d^s$,$\bar{M}_d^s$, ...). For example, for the detection time, the relative error is $\epsilon = \frac{T_d - \bar{T}_d^s}{T_d} \cdot 100$.

The validation of the models was based on a set of 100 repeated random tests. For each test, a relative error ϵ_i of the detection time and cost were obtained. The final result of the validation is the mean and the 95% confidence intervals. For example, in the first validation, the values p_c and p_d were randomly distributed between 0.1 and 1, the number of nodes N between 5 and 100, and finally the λ value has a random distribution of 0.1^n with n from 1 to 5. In order to evaluate the accuracy of the mean max approximations for

		Error %
$S = 1$	T_d	0.60 [0.14, 2.5])
	M_d	1.40 [2.32, 5.2])
$S > 1$	T_d	5.09 [2.84, 12.4])
	$\hat{M}_{ds}$	9.31 [3.42, 13.53]
	$\hat{M}_{dg}$	13.46 [4.1, 20.55]

Table 1: Validation results for $R = 100$. The values presented are the mean error (and the 95% confidence intervals in brackets).

Model	Parameters
All models	$\{p_d, p_c\} \sim \mathcal{U}(0.1, 1), \lambda = 0.1^{\mathcal{U}(1,5)}$
$S = 1$	$N \sim \mathcal{I}(5, 100), D \sim \mathcal{I}(1, N-1)$
$S > 1$	$S \sim \mathcal{I}(1, 5), N \sim \mathcal{I}(S+5, 100), D \sim \mathcal{I}(1, N-S)$

Table 2: Validation scenarios. $\mathcal{U}(a, b)$ stands for the uniform distribution over interval (a, b) and $\mathcal{I}(a, b)$ for an uniform integer distribution between a and b.

$S > 1$, we performed different test for $S = 1$ and for $S > 1$. The results are shown in table 1 and the range of the validation parameters is shown in table 2.

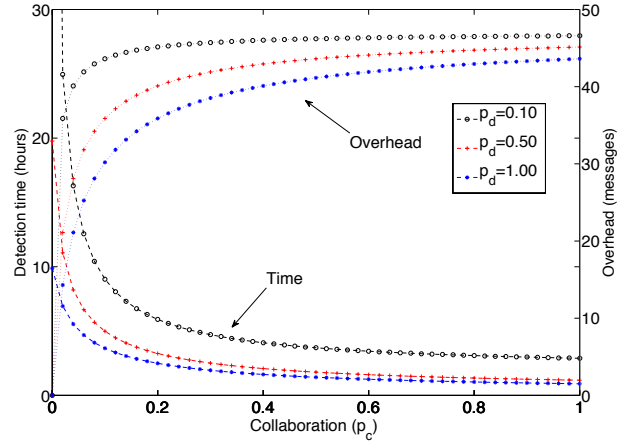
We can see that the differences between the models and the simulation results are low. For $S = 1$ the results are very accurate for all the models. For $S > 1$ the results show that the model is accurate. The greatest error values take place for higher values of S and N, since the number of mathematical operations is huge, and so the precision is reduced. For the approximation we can see that expressions 6 and 10 obtain a fairly good approximation to the simulation results.

5 Evaluation

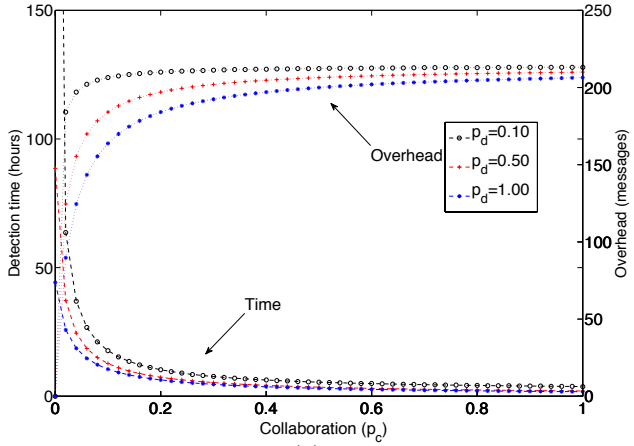
For the following evaluations we used a contact rate of 0.101 contacts/h, $\lambda_h = 2.81 \times 10^{-5} s^{-1}$, obtained from human mobility traces. This value was calculated in [13] using the Cambridge trace date set [8], that was gathered from a set of students of undergraduate years from the University of Cambridge.

The first evaluation shows the influence of the degree of collaboration (p_c) in a network with 50 nodes ($N = 50$), one selfish node ($S = 1$) with different detection probabilities values (p_d). Figure 3a shows the detection time and the overall overhead when we consider $D = 1$ (that is, only one destination node), and figure 3b when all collaborative nodes are destination ($D = N - S = 49$). We observed that increasing the degree of collaboration from 0 to 0.2 reduces the detection time exponentially. As expected, the detection time is greater considering that all the collaborative nodes are destinations ($D = 49$) since the information about the positive has to get to all these nodes. In this case, the reduction of the detection time is quite significant for low detection probabilities ($p_d = 0.1$). For $p_c = 0$ (no collaboration), the detection time

is $15.9 \cdot 10^6 s$ (about 442 hours). This value can be greatly reduced by using our collaborative watchdog. Thus, if all nodes implement the collaborative approach ($p_c = 1$) the detection time is reduced to 3.7 hours. Even for a low collaboration rate $p_c = 0.2$ the time is reduced to 10 hours. For both cases, the overhead is always under 215 messages, which is a very reduced overhead.



(a)

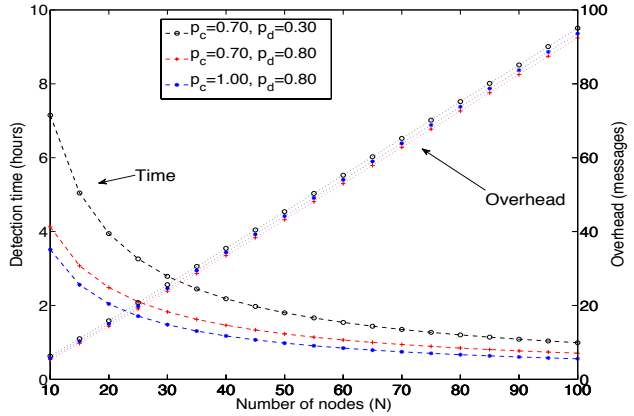


(b)

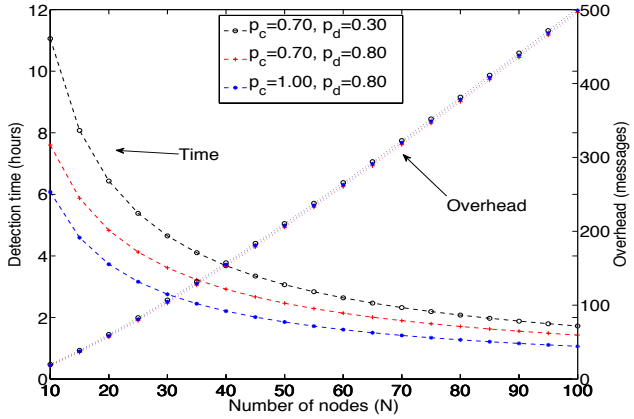
Fig. 3: Detection time and cost evaluation depending on collaboration. a) for $S = 1$ and $N = 50$ for one destination node, b) the same for all nodes ($D = N - 1$)

The goal of second evaluation is to evaluate the impact of the number of nodes, ranging from 10 to 100 (see figures 4a and 4b). Three different sets of values for p_c and p_d were used. The first set (1, 0.8) is a fully collaborative net-

work with a high probability of detection, the second set has a reduced degree of collaboration (0.7); finally, the last set has a low probability of detection (0.3). We observe that, in general, the greater the number of nodes, the lesser the detection time and the greater the number of messages. The main reason is that when the number of nodes is greater, the number of contacts is increased and so the information about the positive detection is disseminated quicker. Reduced values for the collaboration and detection probabilities imply greater detection times (as expected). Nevertheless, the cost only depends on these values (only on N).



(a)



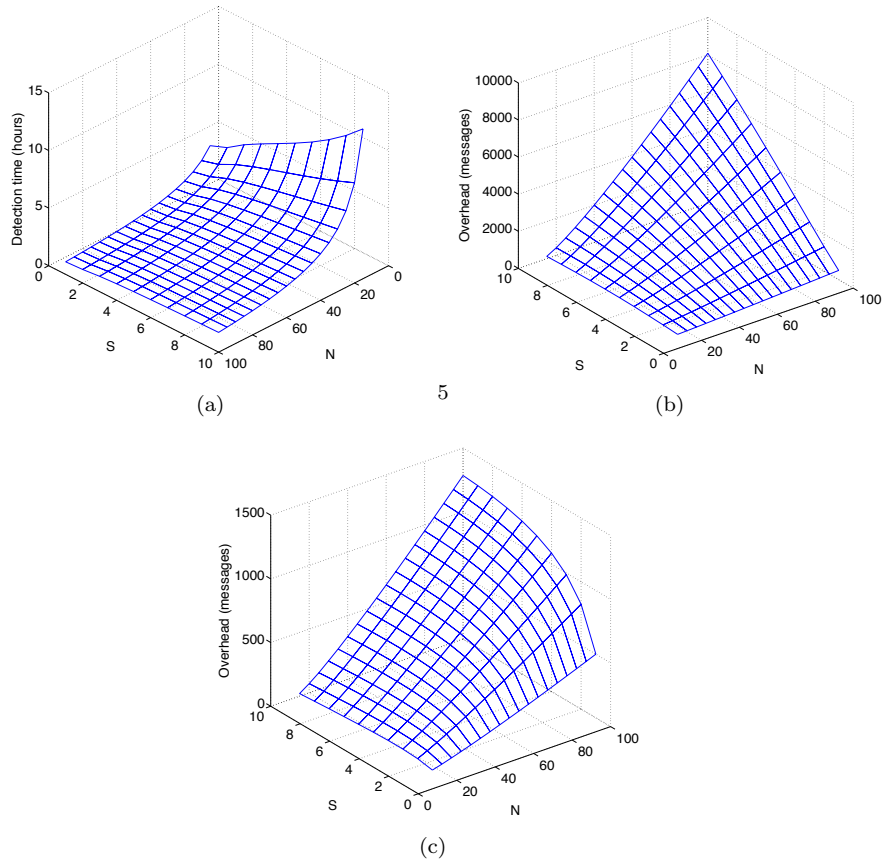
(b)

Fig. 4: Detection time and cost evaluation depending on the number of nodes. a) for $S = 1$ and one destination node, b) for all nodes (except the selfish node)

Figure 5a shows the influence of the number of selfish nodes S on the detection time. The number of selfish nodes has more impact when the number of nodes is low, that is, when the ratio S/N is higher. The reason is obvious: the number of collaborative nodes is also low (for example, for $N = 15$ and $S = 10$, there are only 5 collaborative nodes), and this implies a reduction in the overall number of contacts. For the overhead, we evaluate both distribution protocols. In figure 5b we can see the number of message generated. As expected, the number of messages increases linearly as the number of nodes increases, and also as the number of selfish nodes increases. We can see that the number of messages can be very high when there is more than one selfish node. Thus, the group protocol is the solution when the possible number of selfish nodes is high. The results for the group protocols are shown in figure 5c. We can observe that the number of selfish nodes has a reduced influence on the overhead. This confirms the cost is not increased because only one message is sent for all positives a node has. We repeated this experiment for different values of p_c and p_d revealing the same patterns.

A more detailed evaluation about the influence of the collaboration and detection probabilities is detailed below. Figures 6a and 6b show the influence of p_c and p_d on the detection time for $N = 5$ and $N = 50$, respectively, when all nodes all destination nodes. We observe that for $N = 5$, low values of p_d have a greater impact on the detection time than the collaboration probability. Nevertheless, for $N = 50$, low values of p_c have a greater impact on the detection time than the probability of detection. That is reasonable because, in a network with few nodes a low degree of collaboration has less impact that in a network with more nodes. Figure 6c shows the overhead depending on the values of p_d and p_c. We can see that when the degree of collaboration is low (less than 0.2) the number of messages is reduced drastically (as expected). This is more evident for higher detection probabilities.

Now, we evaluate the dependency on the number of destination nodes D. This value can range from 1, that is the detection time and cost for a single node to detect the selfish(s) node(s); to $N - S$, that is the time and cost that all nodes in the network detect the selfish(s) node(s). Figure 7a shows that the number of destination nodes has a strong influence on the detection time when the number of nodes is low ($N = \{10, 20\}$), and a low influence when the number of nodes is high. The reason is that, when N is low, the number of contacts is also low and so the diffusion of the positives becomes very slow. On the other hand, in a network with more nodes, there are more contacts, meaning that this diffusion is very fast. In terms of overhead and using the group protocol, figure 7b shows that, for D ranging from 1 to 10, the increase is exponential and then mostly lineal. The reason is the same: the number of contacts increases and so the number of messages sent is also greater. Summing up, for a network with a moderate number of nodes ($N > 30$), the number of destination nodes has a strong impact on the overhead.



5

Fig. 5: Detection time and cost evaluation for $S > 1$. a) Detection time; b) Cost for single protocol; c) Cost for group protocol

6 Conclusion

In this paper we have presented a new approach to reduce the detection time of selfish nodes using *collaborative watchdogs*. The network is modelled as a set of wireless mobile nodes that includes both collaborative and selfish nodes. The collaborative nodes have a watchdog that can detect a selfish node with a given probability (of detection). When a contact occurs between two collaborative nodes, the positives are transmitted with a given degree of collaboration (ranging from no collaboration to full collaboration).

We modelled the performance of the collaborative watchdog using a Continuous Time Markov Chain using a contact rate λ. We first introduce a model for evaluating the detection of one selfish node, and then we extended this model for the case of several selfish nodes, including a mean-max approxi-

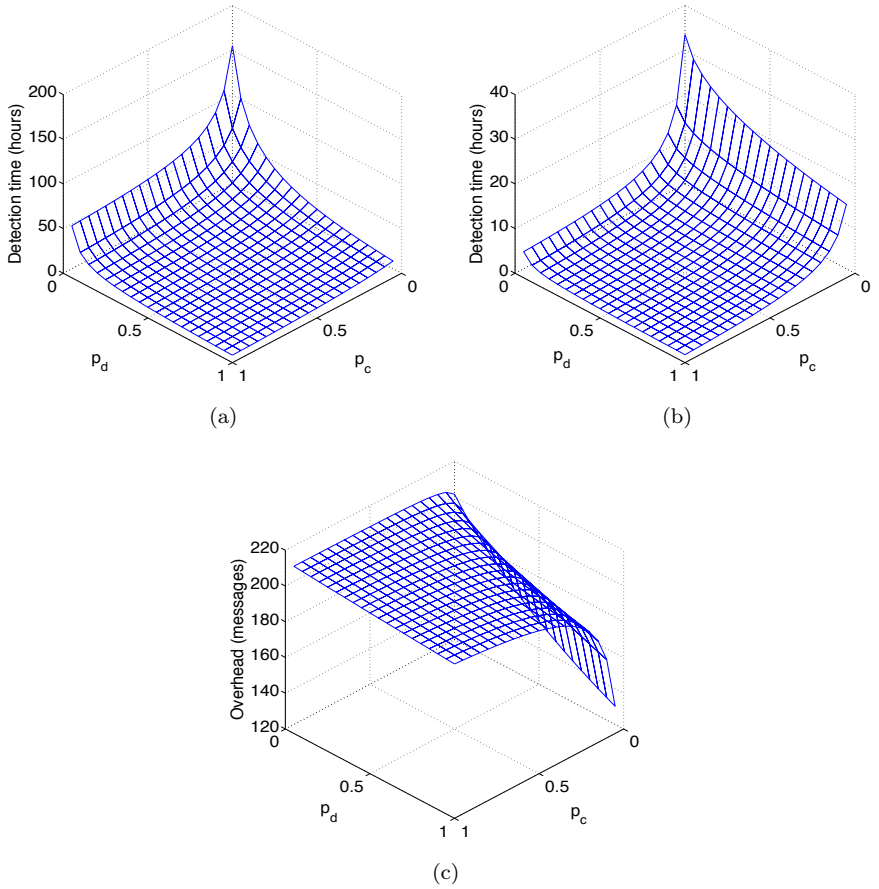


Fig. 6: a)-c) Evaluation depending on collaboration and detection probabilities. a) Overall detection time with a reduced number of nodes ($N = 5$). b) Detection time with a greater number of nodes ($N = 50$) c) Overhead for $N = 50$;

mation for a feasible computation when the number of selfish nodes is high. Numerical results show that our collaborative watchdog can reduce the overall detection time with a reduced overhead (messages cost). This reduction is very significant when the watchdog detection effectiveness is low. Furthermore, this reduction can be obtained even with a moderate degree of collaboration. These two properties are very important for the practical implementation of the collaborative watchdog. Our approach can obtain great results with a moderate precision watchdog, and it can tolerate some degree of no collaboration (for example, when the contact duration is too low to allow transmitting a message with the positives).

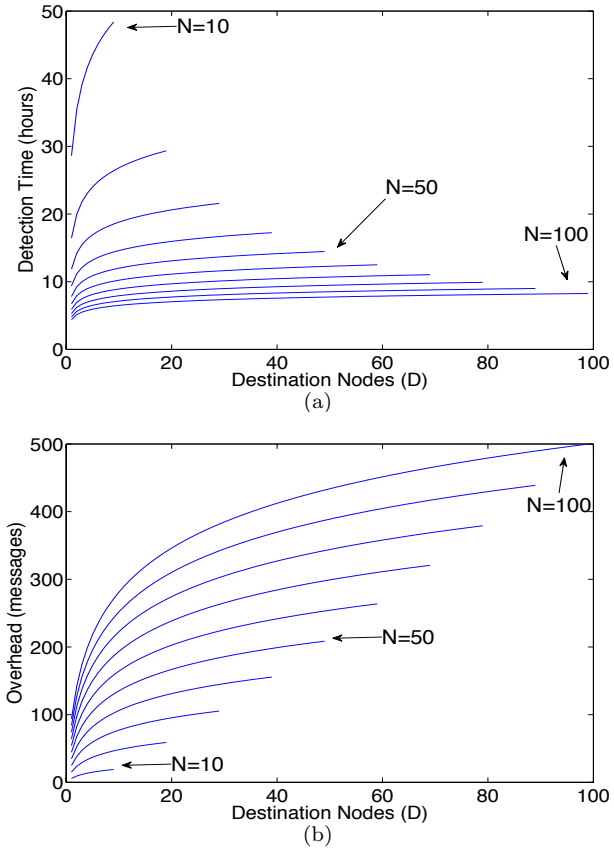


Fig. 7: Evaluation of the impact of destination nodes for $p_c = 1$ and $p_d = 0.7$. a) Detection time; b) Cost for group protocol

As a future work, we want to improve the analytical model to evaluate more complex scenarios. We also plan to implement the collaborative detection mechanism in order to evaluate its performance on a more realistic scenario.

Acknowledgments

This work was partially supported by the *Ministerio de Ciencia e Innovación*, Spain (Grant TIN2011-27543-C03-01).

References

1. Bohnenkamp, H., Haverkort, B.: The mean value of the maximum. In: APM-PROBMIV, *LNCS*, vol. 2399, pp. 37–56. Springer-Verlag (2002)

2. Buchegger, S., Le Boudee, J.Y.: Self-policing mobile ad hoc networks by reputation systems. Communications Magazine, IEEE **43**(7), 101 – 107 (2005)
3. Gao, W., Li, Q., Zhao, B., Cao, G.: Multicasting in delay tolerant networks: a social network perspective. In: Proceedings of MobiHoc '09, pp. 299–308 (2009)
4. Groenevelt, R., Nain, P., Koole, G.: The message delay in mobile ad hoc networks. Performance Evaluation **62**, 210–228 (2005)
5. Hollick, M., Schmitt, J., Seipl, C., Steinmetz, R.: On the effect of node misbehavior in ad hoc networks. In: Proceedings of IEEE International Conference on Communications, ICC'04, pp. 3759–3763. IEEE (2004)
6. Hortelano, J., Cano, J.C., Calafate, C.T., de Leoni, M., Manzoni, P., Mecella., M.: Black hole attacks in p2p mobile networks discovered through bayesian filters. In: P2P Collaborative Distributed Virtual Environments (P2P CDVE 2010) (2010)
7. Hortelano, J., Ruiz, J.C., Manzoni, P.: Evaluating the uselfusness of watchdogs for intrusion detection in vanets. In: ICC'10 Workshop on Vehicular Networking and Applications (2010)
8. Hui, P., Crowcroft, J., Yoneki, E.: Bubble rap: social-based forwarding in delay tolerant networks. In: Proceedings of MobiHoc '08, pp. 241–250. ACM (2008)
9. Karagiannis, T., Le Boudec, J.Y., Vojnović, M.: Power law and exponential decay of inter contact times between mobile devices. In: Proceedings of MobiCom '07, pp. 183–194 (2007)
10. Karaliopoulos, M.: Assessing the vulnerability of dtn data relaying schemes to node selfishness. Communications Letters, IEEE **13**(12), 923 –925 (2009)
11. Kargl, F., Klenk, A., Schlott, S., Weber, M.: Advanced detection of selfish or malicious nodes in ad hoc networks. In: Proceedings of the 1st European on Security in Ad-Hoc and Sensor Networks (ESAS 2004, pp. 152–165. Springer Verlag (2004)
12. Li, Q., Zhu, S., Cao, G.: Routing in socially selfish delay tolerant networks. In: Proceedings of INFOCOM'10, pp. 857–865. IEEE Press, Piscataway, NJ, USA (2010)
13. Li, Y., Su, G., Wu, D., Jin, D., Su, L., Zeng, L.: The impact of node selfishness on multicasting in delay tolerant networks. Vehicular Technology, IEEE Transactions on **60**(5), 2224–2238 (2011)
14. Marti, S., Giuli, T.J., Lai, K., Baker, M.: Mitigating routing misbehavior in mobile ad hoc networks. In: Proceedings of MobiCom '00, pp. 255–265 (2000)
15. Palazzo, S., Campbell, A.T., Dias de Amorim, M.: Opportunistic and delay-tolerant networks. EURASIP Journal on Wireless Communications and Networking **2011** (2011)
16. Paul, K., Westhoff, D.: Context aware detection of selfish nodes in dsr based ad-hoc networks. In: Proceedings of IEEE Globecom (2002)
17. Toh, C., Kim, D., Oh, S., Yoo, H.: The controversy of selfish nodes in ad hoc networks. In: Proceedings of IEEE International Conference on Advanced Communication Technology (IEEE ICACT Conference)", vol. 2, pp. 1087 –1092 (2010)
18. Zhu, H., Fu, L., Xue, G., Zhu, Y., Li, M., Ni, L.M.: Recognizing exponential inter-contact time in vanets. In: Proceedings of INFOCOM'10, pp. 101–105. IEEE Press (2010)