

Document downloaded from:

<http://hdl.handle.net/10251/78059>

This paper must be cited as:

Hernández Orallo, E.; Cano Escribá, JC.; Tavares De Araujo Cesariny Calafate, CM.; Manzoni, P. (2013). A Representative and Accurate Characterization of Inter-contact Times in Mobile Opportunistic Networks. ACM. doi:10.1145/2507924.2507936.



The final publication is available at

<http://dx.doi.org/10.1145/2507924.2507936>

Copyright ACM

Additional Information

© ACM 2013. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in MSWiM '13 Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems; <http://dx.doi.org/10.1145/2507924.2507936>.

A Representative and Accurate Characterization of Inter-contact Times in Mobile Opportunistic Networks

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

Departamento de Informática de Sistemas y Computadores.
Universidad Politécnica de Valencia. Valencia. Spain.

ehernandez@disca.upv.es, jucano@disca.upv.es, calafate@disca.upv.es,
pmanzoni@disca.upv.es

ABSTRACT

A representative characterisation of inter-contact times between nodes is essential for the performance evaluation of mobile networks. The most common characterization of inter-contact times is based on the study of the aggregate distribution of contacts between individual pairs of nodes. The problem with this aggregate distribution is that it is not always representative of the individual pair distributions, especially in the short term and when the number of nodes in the network is high. Thus, deriving results from this characterisation, can lead to inaccurate performance evaluations results.

In this paper, we propose and evaluate two new methods for characterising the inter-contact times distribution in mobile networks. We prove that these characterizations have a higher representativeness, thereby improving the accuracy of the derived performance results. For evaluating the precision of the different characterizations we use the epidemic routing protocol, which has an analytical performance expression that is based on a contact rate λ . We derive from each of the characterizations the corresponding λ values. Then, we compare the results obtained using the analytical expression with simulation results using both synthetic and real contact traces. It is shown that the new characterization methods are very accurate, even for reduced contact traces and high number of nodes.

Categories and Subject Descriptors

C.2.1 [Network Architecture and Design]: Wireless communication; I.6 [Simulation and modelling]: Model validation and Analysis

Keywords

Mobile Networking, Performance Evaluation, Inter-contact times

1. INTRODUCTION

Mobile opportunistic networks collectively create dynamic networks that are built from short unpredictable contact times as nodes move in and out of connectivity. Applications of such networks include Mobile Ad-Hoc Networks (MANETs), Vehicular Ad-Hoc Networks (VANETs) and Mobile Social Networks.

Characterizing inter-contact times (or inter-meeting times) between nodes is essential for analyzing the performance of contact based protocols in cooperative or opportunistic networks. The established approach is to characterize the inter-contact times distribution between pairs of nodes using an *aggregate distribution* [2,7,9,13,15]. This distribution is obtained by *aggregating* the *individual pair distribution* of all combinations of pairs of nodes in the network. The *individual pair distribution* is defined as the distribution of the time elapsed between two consecutive contacts between the same pair of nodes [12].

Another form of characterizing the inter-contact times is to consider the time elapsed between contacts for any pairs of nodes in a group (known as *inter-any-contact times*). This characterization was briefly studied in [7] using human mobility traces. The conclusions were that inter-any-contact times are longer than individual pairs inter-contact times (as expected), but with a similar distribution shape. This paper also shows a time dependence in the contact distribution, with different pattern distributions for the diurnal and night periods.

The goal of a correct characterization of the inter-contact times distribution is to improve the precision of the performance evaluations derived from these characterizations. For example, there are many analytical performance models that assume that the inter-contact times distribution between pairs of nodes are exponentially distributed with a given rate λ . For example, using a contact rate λ we can obtain the transmission delay and cost of several mobile protocols, such as epidemic routing protocols [4,14], and the impact of node selfishness on mobile networks [5,6,10]. The precision of the previous models clearly depends on how accurate is the estimation of the contact rate λ , which depends directly on the representativeness of the characterized distribution.

Previous works have studied the distribution of the inter-contact times by collecting data from real mobile network environments [2–4,9,15]. Some works have shown that the aggregate inter-contact times distribution is exponential with rate λ for both human and vehicle mobility scenarios [4,11,15]. The work in [3] analyzed some popular mobility traces

and found that over 85% of the individual pair distributions fit an exponential distribution. Nevertheless, there is some controversy about whether this exponential distribution relates to real mobility patterns. Some empirical results have shown that the aggregate inter-contact times distribution follows a power-law distribution and has a long tail [2], meaning that some pairs of nodes barely experience any contact. In [1], it is shown that in a bounded domain, the inter-contact distribution is exponential, but in an unbounded domain the distribution is power-law. The dichotomy of this distribution is described in [9], which shows a truncated power law with exponential decay appearing in its tail after some cutoff point. A recent paper [12], presents the dependence between the *individual pair distributions* and the *aggregate distribution*. It is stated that, starting from exponentially *individual pair distributions*, the *aggregate distribution* is distributed according to a Pareto law. It also verifies the dichotomy property of the aggregate distribution analytically.

Summing up, most of the literature is based on the aggregate distribution, assuming that it is representative of the individual pair distribution [12]. The practical reason of using the aggregate distribution is that is a more convenient approach, since obtaining enough samples for each individual pair of contacts is not feasible. The practical approach for obtaining the contact rate λ is to assume that all individual pair distributions are similar so the contact rate of the aggregate distribution is similar to the individual pair distributions contact rates. Nevertheless, as shown in [12], these contact rates are only similar when the length of the contact trace is large. Therefore, we need another method to estimate this individual pair distribution (and the contact rate), feasible to measure and more statistically representative.

In this paper, we evaluate better methods to characterize the inter-contact times distribution presenting higher representativeness and, thus, increasing the precision of the results obtained. Using different characterization methods, three inter-contact time distributions are considered: *Aggregate Pairs* distribution, that is the *established characterization*; the *Aggregate Nodes* distribution, that is obtained as the aggregate of inter-contact distributions between one node and the rest of nodes; and the *Any Contact* distribution, that is the inter-contact distribution between any nodes. First, we study its statistical representativeness showing that, for the same trace length, the *Aggregate Pairs* distribution has a very low representativeness, especially when the number of nodes is high, in contrast with the others that have a good representativeness. Second, we study the relation between these distributions. We prove that, if all individual pair distributions are exponentially distributed, the *Aggregate Nodes* distribution is simply a new exponential distribution. This is not true for the *Aggregate Pairs* distribution, that depends on the distribution of each individual λ [12]. This conclusion is very important, because it allows to obtain in a more precise way the value of λ used in the analytical models.

Finally, we evaluate the precision of the three distributions using both synthetic and real contact traces. The precision is evaluated using a well known analytical model, the epidemic routing, that is based on a given exponential contact rate λ obtained from the three previous characterization methods. Experimental results confirm that the established *Aggregate Pairs* distribution is under-representative

t	n_1, n_2	d	t	n_1, n_2	d	t	n_1, n_2	d
1	(1,2)	2	7	(2,3)	4	14	(2,3)	2
2	(2,3)	3	7	(1,2)	6	17	(1,2)	3
2	(3,4)	2	7	(1,5)	3	20	(1,3)	6
5	(1,3)	5	8	(3,4)	5	20	(2,5)	2
5	(1,2)	1	11	(1,4)	4	-	-	-

Table 1: Contact trace sample. Note that a contact (n_1, n_2) means that both nodes have visibility of each other.

and, consequently, the precision of the results obtained using the epidemic routing model are imprecise. This precision is very low when the number of nodes of the evaluated network is high. Instead, the results using the *Aggregate Nodes* and *Any-Contact* distributions are much more precise, requiring significantly smaller contact traces.

The rest of the paper is organized as follows. We introduce the three methods for characterizing inter-contact times distributions evaluating its representativeness in Section 2. Section 3 study the relation between these distributions. The evaluation of the representativeness and precision of the different characterizations is described in Section 4. Finally, Section 5 presents some concluding remarks.

2. CHARACTERIZING INTER-CONTACT TIMES DISTRIBUTIONS

In this section we describe three possible methods for characterizing the inter-contact times distribution from a contact trace. Then, we study their representativeness by introducing some metrics and evaluating these metrics through some real traces.

A contact trace is obtained measuring the times when contacts between pairs of nodes occur for a given interval time T . The result is a trace of length $C(T)$ (that is, the number of contacts), where each record is a 4-tuple (t_i, a_i, b_i, d_i) , reflecting, that at time $t_i \leq T$, there is a contact between the pair of nodes (a_i, b_i) with a duration of d_i seconds. A *contact* is defined as an opportunity of transmission between a pair of nodes (that is, two nodes are able to communicate between them directly for a given time). Based on this definition, for practical issues, there is only one record for each contact between a pair of nodes (a_i, b_i) , and this contact is reciprocal (in other words, there is no another record with the (b_i, a_i) contact).

Using a contact trace, we can characterize the inter-contact times distribution. This distribution is influenced by the trace's duration and resolution [2]. The resolution is defined as the smallest interval between two successive measurements. Inter-contact times that last more or are close to the duration of the experiment, and inter-contact times that last less than the time resolution, cannot be observed.

For clarity of exposition, we are going to use a simple contact trace as exemplified in Table 1. This trace corresponds to a network of five nodes ($N = 5$), which has a duration T of 20 seconds, resulting in 14 contacts ($C(T) = 14$).

From a contact trace we can characterize three possible inter-contact distributions:

- *Aggregate Pairs* (AP) distribution: the aggregate of inter-contact times distributions between the same pair of nodes. This is the established characterization, usually known as simply *inter-contact times* distribution [2–4, 9, 15]. This distribution is obtained by aggregat-

ing the *individual pair inter-contact times* of all node pair combinations in the network. We need at least 2 contacts between the same pair of nodes in order to obtain an inter-contact time. For example, using our contact trace, for the pair (1,2) we have the following contact times¹: {1,5,7,17} and so the inter-contacts time are {4,2,10}. We can also obtain the inter-contact times for the following pair of nodes: (1,3), (2,3) and (3,4); however, for the rest of pairs, this is not possible. By aggregating all the inter-contact times calculated for the previous pairs, we can obtain the aggregate pairs distribution.

- *Aggregate Nodes* (AN) distribution: the aggregate of inter-contact distributions between one node and the rest of nodes. For example, using the sample trace, we can obtain the contacts of node 1 with the other nodes: (1,x)={1,5,5,7,7,11,17,20}, so the inter-contact times are {4,0,2,0,4,6,3}. For this trace all nodes have contacts with other nodes. By aggregating all the inter-contact times we can obtain the aggregate nodes distribution.
- *Any Contact* (AC) distribution: inter-contact distribution between all nodes. This is known as the *inter-any-contact times* in [7]. In this case there is only one distribution of inter-contact times, which corresponds to the difference between two consecutive contacts. For our sample trace we have 13 inter-contact time values: {1,0,3,0,2,0,0,1,3,3,3,3,0}.

A key factor of a contact trace characterization is its representativeness. This representativeness depends on the trace length $C(T)$, the number of nodes N in the network, and the method. Two metrics are defined to measure this representativeness: the contact ratio R_C and the *average* number of measures per individual distribution M_C . These metrics are detailed below for each distribution.

For the aggregate pairs (AP) distribution, in a network of N nodes, the number of possible combinations of pair of nodes is $V = N \cdot (N - 1) / 2$. If we denote with $\mathcal{P}(t)$ the set of pairs of nodes that has at least two contacts up to time t , the contact ratio is $R_C = |\mathcal{P}(T)| / V$ (where $|\mathcal{P}(t)|$ is the size of $\mathcal{P}(t)$). For our sample trace, this ratio is $R_C = 4/10 = 0.4$. The average number of measures M_C is computed as the sum of all inter-contact times divided by the number of node pair combinations, that is,

$$M_C = \frac{1}{V} \sum_{p \in \mathcal{P}(T)} (C_p(T) - 1) \quad (1)$$

where $C_p(T)$ is the number of contacts between pairs of nodes $p \in \mathcal{P}(T)$. So, in our example, we have $M_C = (3 + 2 + 1 + 1) / 10 = 0.7$.

In the aggregate nodes (AN) distribution, the contact ratio is: $R_C = |\mathcal{N}(T)| / N$, where $\mathcal{N}(t)$ is the set of nodes that has at least two contacts up to time t . For our sample trace, this ratio is $R_C = 5/5 = 1$. The average number of measures

¹In this paper, the inter-contact time is computed as the difference between the starting times of two consecutive contacts. Another way to compute this time (see [2]) is to obtain the difference between the end of a contact and the start of the next one. In this case, the duration of the contact is used to obtain the inter-contact times.

	Cambridge	Shanghai
Type	Human	Vehicle
Device	iMote	GPS+GPRS
Network Type	Bluetooth	WiFi
Duration (hours)	274	24
Resolution (seconds)	120	60
Nodes	36	2288
Contacts (C)	21200	1262498
<hr/>		
Repr. AP distribution		
Contact ratio (R_C)	0.81	0.04
Aver. measures (M_C)	16	0.08
<hr/>		
Repr. AN distribution		
Contact ratio (R_C)	1	0.99
Aver. measures (M_C)	459	474
<hr/>		
Repr. AC distribution		
Contact ratio (R_C)	1	1
Aver. measures (M_C)	21199	1262497

Table 2: Description of contact traces and representativeness.

M_C is computed as:

$$M_C = \frac{1}{N} \sum_{n \in \mathcal{N}(T)} (C_n(T) - 1) \quad (2)$$

where $C_n(T)$ is the number of contacts for each node $n \in \mathcal{N}(T)$. So, in our example, we have $M_C = (7 + 6 + 6 + 2 + 1) / 5 = 4.2$.

Finally, in the any contact distribution (AC) the ratio is 1 (if $C(T) > 1$) and the number of measures is $C(T) - 1$.

As expected, we can see that the representativeness of the AP distribution is very low: a contact ratio of only 0.4 and less than one measure per pair of nodes. In the AN distribution the representativeness is greater: full contact ratio and a mean of 4.2 measures for each node. The greatest representativeness corresponds to the AC distribution which has a full ratio and the highest number of measures.

In general, for the same trace time T , the lower representativeness is for the AP distribution, and this representativeness (the values of R_C and M_C) is inversely proportional to V , that exhibits quadratic growth with N . Therefore, when the number of nodes is high, the representativeness is very low. For the AN distribution it representativeness is inversely proportional to the number of nodes N . And finally the best representativeness is for the AC distribution.

We study the representativeness of the previous distributions using some well known real contact traces (see Table 2). The *Cambridge* mobility set trace [8] was gathered from a set of undergraduate students from the University of Cambridge carrying small devices (iMotes) in 2005. This trace has a duration of 274 hours and has 36 mobile nodes (students)². The *Shanghai* Taxis GPS Trace [15] was collected from 2100 taxis in Shanghai city during February of 2007³. This trace does not contains the contacts (it contains GPS locations), so a pre-process for obtaining the contact trace is needed. Following the method used in [15] we assume that a contact occurs if both vehicles are in WiFi range (100 meters).

²Although this trace has also static nodes we only evaluate the mobile nodes.

³The trace used in [15] is for 4 weeks. In this paper we use only the trace for one day, that it is the only available online.

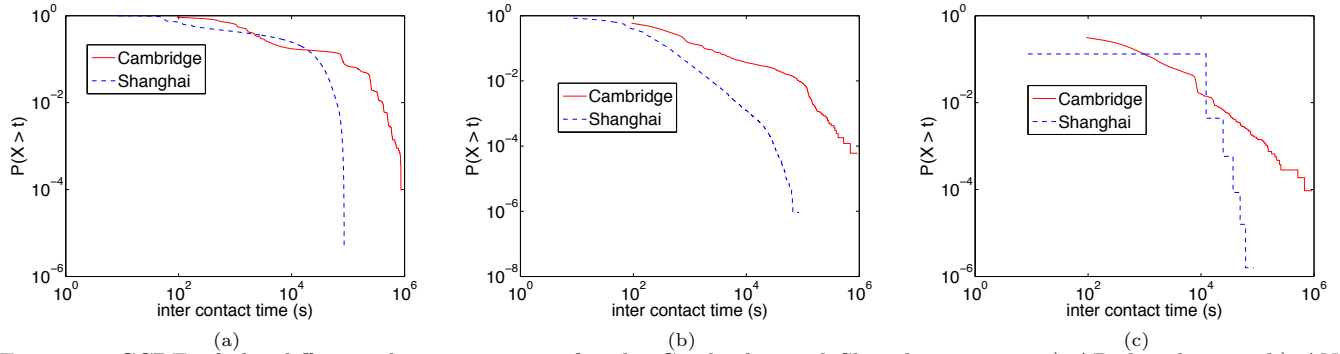


Figure 1: CCDF of the different characterizations for the Cambridge and Shanghai traces. a) AP distribution, b) AN distribution, c) AC distribution

Table 2 shows that in the Cambridge trace, which has only 36 nodes, the contact ratio for the AP distribution is high (although not 1) and the average number of measurements is relatively high. We can say that it has a moderate representativeness. Nevertheless, for the Shanghai trace, that has a high number of nodes, the AP distribution has a very low representativeness, so this characterization is not useful for obtaining information about the contact pattern on this network. For the AN distribution, both traces have a good representativeness (in terms of contact ratio and average measures). And finally, for the AC distribution, both traces have excellent representativeness (full ratio and the highest number of measures).

3. RELATION BETWEEN DISTRIBUTIONS

In this section we study the relation between the different distribution characterizations (AP, AN, AC). In order to study the inter-contact times distribution, the CCDF (Complementary Cumulative Distribution Function) is widely used [2, 9, 12]. The CCDF (also known as the *tail distribution*) is useful to study how often a random variable is above a particular level, that is, $\bar{F}(x) = P(X > x)$. So, first, we are going to introduce the relation between the different CCDFs. As we will show, if the individual pair distributions are not exponential we cannot obtain a simple relation between these distributions. Thus, we also study the relation between the average contact rates (λ) that are valid for any distribution.

3.1 Relation between CCDF

We can see in Figure 1 the CCDF of the different distributions for the Cambridge and Shanghai traces. In the CCDF for the AP characterization, we can clearly observe the dichotomy of the distribution: the inter-contact time follows a power-law decay up to a characteristic time (about $10^5 s \approx 12$ hours for the Cambridge set and $20000s \approx 6$ hours for the Shanghai set) and beyond this time, the decay is exponential. This confirms earlier studies [9, 15]. For the AN distribution, the previous dichotomy is not so evident, and it is clearer that it follows an exponential decay. Finally, for the AC distribution we can see an exponential decay. Note also the resolution issues in the Shanghai traces, where the inter-contact times are below the resolution time, and so its values are very discretized.

The relation between the CCDF of individual pair p ($\bar{F}_p(x)$) and the aggregate pairs CCDF ($\bar{F}_{AP}(x)$) for the AP distribution was established in [9] as:

$$\bar{F}_{AP}(x) = \sum_{p \in \mathcal{P}(T)} \frac{C_p(T) - 1}{C(T) - |\mathcal{P}(T)|} \bar{F}_p(x) \quad (3)$$

and when T tends to be large,

$$\bar{F}_{AP}(x) = \sum_{p \in \mathcal{P}} \frac{\lambda_p}{\Sigma \lambda_p} \bar{F}_p(x) \quad (4)$$

where $1/\lambda_p$ is the average of inter-contact times for the pair p (in other words, λ_p is the average contact rate of the pair), and $\Sigma \lambda_p = \sum_{p \in \mathcal{P}} \lambda_p$. Expression 4 shows that the aggregate CCDF is equal to the weighted sum of individual CCDF with weight proportional to the rate of contacts λ_p . Furthermore, if all individual CCDF ($\bar{F}_p(x)$) are identical and the ratio of contacts (R_C) is assumed to be 1 (note that if $T \rightarrow \infty$ then $R_C \rightarrow 1$), then the aggregate CCDF and the individual CCDF are the same. Based on equation 4, when the individual distributions are different and exponentially distributed, we cannot obtain a simple relation between these distributions and the aggregate one. More information is needed about the distribution of the individual rates λ_p and, depending on this distribution, the aggregate distribution can follow an Exponential, Pareto, or Power Law with Exponential Decay distribution (see paper [12] for a detailed study about these relations).

In a similar way, we can establish a relation between the individual node n CCDF ($\bar{F}_n(x)$) and the aggregate nodes CCDF for the AN characterization:

$$\bar{F}_{AN}(x) = \sum_{n \in \mathcal{N}(T)} \frac{C_n(T) - 1}{C(T) - |\mathcal{N}(T)|} \bar{F}_n(x) = \sum_{n \in \mathcal{N}} \frac{\lambda_n}{\Sigma \lambda_n} \bar{F}_n(x) \quad (5)$$

where $1/\lambda_n$ is the mean value for the inter-contact times for node n , and $\Sigma \lambda_n = \sum_{n \in \mathcal{N}} \lambda_n$. If all individual CCDF ($\bar{F}_n(x)$) are identical and $R_C = 1$, then the aggregate CCDF is equal to the individual CCDF.

Now, we are going to study the relation between the three different distributions. Expressions 4 and 5 establish the relation between the aggregate distribution and the individual distributions. These expressions are derived on the basis that the aggregate distribution is obtained using the aggregation of the individual contact times.

First, we establish a relation between the AC distribution and the AP distribution, through the individual pair CCDF of the AP distribution ($\bar{F}_p(x)$). We can see that if a contact occurs at time t , the next contact time will be the minimum time of all possible contacts between pairs. That is, the minimum of the random variables associated with the individual distributions $F_p(x)$. So, we have:

$$\bar{F}_{AC}(x) = \min_{p \in \mathcal{P}(T)} \{F_p(x)\} \quad (6)$$

where $\min\{\}$ is the CCDF of the minimum of the random variables of the distributions. This minimum depends on the type of distribution. Fortunately, for exponential distributions, this expression is very simple. For n independent exponentially distributed random variables with rate parameters $\lambda_1, \dots, \lambda_n$, the minimum is also exponentially distributed with parameter $\lambda = \lambda_1 + \dots + \lambda_n$. Then, if all distributions are exponentially distributed, the CCDF is also exponentially distributed:

$$\bar{F}_{AC}(x; \lambda_{AC}) = \bar{F}(x; \sum_{p \in \mathcal{P}(T)} \lambda_p) \quad (7)$$

where λ_{AC} is the contact rate of the AC distribution. We can see that this relation is more simple than expressions 4 and 5. Furthermore, it allows to obtain the average contact rate $\bar{\lambda}_p$ from λ_{AC} , when T tends to be large:

$$\lambda_{AC} = \sum_{p \in \mathcal{P}} \lambda_p \rightarrow \bar{\lambda}_p = \frac{\lambda_{AC}}{|\mathcal{P}|} = \frac{\lambda_{AC}}{\frac{1}{2}N(N-1)} \quad (8)$$

This relation is of special interest, as many analytical models based on Markov Chains use λ_p as the contact rate.

The relation between the AC distribution and the AN distribution is obtained in a similar way. If we use expression $\min_{n \in \mathcal{N}(T)} \{\bar{F}_n(x)\}$, we can see that for a given pair (a, b) , $\bar{F}_n(x)$ is repeated twice, one for the first node on the pair $a \in p$, and another one for the second node of the pair $b \in p$. For the AC distribution the contacts are no longer repeated, and so we have:

$$\bar{F}_{AC}(x) = \frac{1}{2} \min_{n \in \mathcal{N}(T)} \{F_n(x)\} \quad (9)$$

Then, if all distributions are exponentially distributed, the CCDF is:

$$\bar{F}_{AC}(x; \lambda_{AC}) = \bar{F}(x; \frac{1}{2} \sum_{p \in \mathcal{N}(T)} \lambda_n) \quad (10)$$

Finally, we can obtain a similar relation between the individual distributions of the AP and AN distributions. For a node n , when a contact occurs at time t , the next contact time will be the minimum time of all contacts between pairs p that have node n . So, we have:

$$\bar{F}_n(x) = \min_{p \in \mathcal{P}(T) \wedge n \in p} \{F_p(x)\} \quad (11)$$

If all distributions are exponential distributed, the CCDF is:

$$\bar{F}_n(x; \lambda_n) = \bar{F}(x; \sum_{p \in \mathcal{P}(T) \wedge n \in p} \lambda_p) \quad (12)$$

In conclusion, we have shown that the AN and AC distributions have a simpler relation with the individual pair distribution than the previous AP distribution. Furthermore, if the individual pair distributions are exponentially

distributed, the AN and AC distributions are also exponentially distributed. This fact confirms the exponential shapes shown in Figure 1.

3.2 Relation between Contact Rates

In this section we study the relation between the contact rates of the different characterizations, based on the individual contact rates. The following relations are valid for all types of distributions.

The relation between the average contact rate of the AC distribution (λ_{AC}) and the average contact rates of individual pair of nodes (λ_p) is obtained as follows⁴. For large values of T , the number of contacts generated up to time T in the AC distribution is $T \cdot \lambda_{AC}$. This number of contacts must be the same as the number of contacts for all pair of nodes p in the AP distribution. These contacts are obtained as the sum of the contacts of all pairs p . Thus, making equal both expressions and removing T on both parts, we have:

$$\lambda_{AC} = \sum_{p \in \mathcal{P}(T)} \lambda_p \quad (13)$$

The relation between λ_{AC} and the individual node contact rate λ_n is obtained in a similar way.

$$\lambda_{AC} = \frac{1}{2} \sum_{n \in \mathcal{N}(T)} \lambda_n \quad (14)$$

Note that in the previous expression we take into account repeated pair of nodes, as detailed for expression 9.

Now, we can obtain the relation between the average contact rate of the AP distribution λ_{AP} and λ_p . The number of contacts generated for each λ_p is $T \cdot \lambda_p$, so the total number of contacts is $T \cdot \sum_{p \in \mathcal{P}(T)} \lambda_p$. Note that this number of contacts must be equal to the number of contacts for the AP aggregate distribution, that depends on the number of pairs and its rate: $T|\mathcal{P}(T)|\lambda_{AP}$. Thus, making equal both expressions and simplifying, we have:

$$\lambda_{AP} = \frac{\sum_{p \in \mathcal{P}(T)} \lambda_p}{|\mathcal{P}(T)|} \quad (15)$$

Finally, the relation between the average contact rate for the AN distribution λ_{AN} and λ_n is obtained in a similar way:

$$\lambda_{AN} = \frac{\sum_{n \in \mathcal{N}(T)} \lambda_n}{|\mathcal{N}(T)|} \quad (16)$$

That is, the contact rate of the aggregate distribution is equal to the mean of the individual contact rates.

Combining expressions 13 and 15 we can obtain the relation between λ_{AP} and λ_{AC} , for T large enough, as:

$$\lambda_{AP} = \frac{\lambda_{AC}}{|\mathcal{P}(T)|} = \frac{\lambda_{AC}}{\frac{1}{2}N(N-1)} \quad (17)$$

and expressions 16 and 14:

$$\lambda_{AN} = \frac{2\lambda_{AC}}{|\mathcal{N}(T)|} = \frac{2\lambda_{AC}}{N} \quad (18)$$

and from expressions 17 and 18 we have:

$$\lambda_{AN} = (N-1)\lambda_{AP} \quad (19)$$

⁴Abusing notation, we denote the average contact rate using λ . This does not imply that the distributions must be exponential.

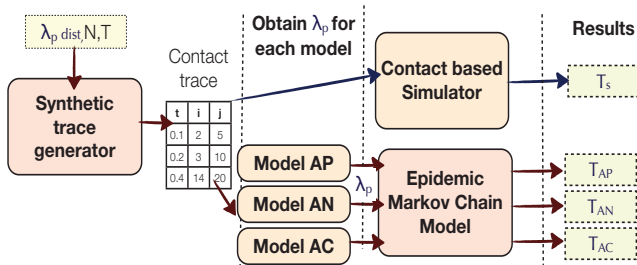


Figure 2: The evaluation process.

As a corollary, we can see, that if all λ_p values are equal, then $\lambda_{AP} = \lambda_p$, $\lambda_n = (N - 1)\lambda_p$, and $\lambda_{AN} = \lambda_n$.

From the previous relations two important conclusions can be drawn. First, if the individual inter-contact times distribution between pairs of nodes is exponentially distributed, we can obtain the average λ in a simple way, using expressions 8 and 10. Since AN and AC distributions are more statistically representative than the AP distribution for the same contact trace time T , the value of λ obtained using these distributions will be more precise. Second, based on the relation between contact rates, we can also obtain the average contact rate between pairs of nodes, using the more representative AN and AC distributions.

4. EXPERIMENTAL EVALUATION

In previous sections we have studied the representativeness of the different distributions and the relation between them. We showed that the average node (AN) distribution, and especially the any contact (AC) distribution has better statistical representativeness than the widely established aggregate pairs (AP) distribution.

Now, we are going to evaluate the precision of these distributions. There are many analytical performance models that assume that the inter-contact times distribution between pairs of nodes is exponentially distributed [4, 10, 14, 15]. This fact allows using the contact rate λ in Markovian models. We use the model for the Epidemic Routing as a way to evaluate the precision of the different distributions. Specifically, the model we use, based on Markov chain models, was introduced in [4] for obtaining the average source-to-destination delivery delay ($T_d = E[T]$). This model assumes a unique contact rate λ between all pair of nodes and full probability of transmission when a contact occurs. So, all individual pair distributions must be equal and exponentially distributed with mean λ_p . As shown in section 3, when assuming the previous condition we have that:

$$\lambda_p = \lambda_{AP} = \frac{\lambda_{AN}}{N-1} = \frac{\lambda_{AC}}{\frac{1}{2}N(N-1)} \quad (20)$$

so λ_p can be estimated using any of the previous distributions.

The process for evaluating the distributions is the following (see Figure 2). We synthetically generate a contact trace for a given time T with inter-contact times between pairs of nodes that are exponentially distributed with a contact rate λ_p . This contact trace is used to estimate three different contact rates $\hat{\lambda}_p$ from the different contact rates of the distributions ($\lambda_{AP}, \lambda_{AN}, \lambda_{AC}$) using expression 20. Then, using the different $\hat{\lambda}_p$ generated for each distribution (AP, AN, AC)

we can obtain the delivery time using the Markov model for epidemic routing. For validation purposes, we also implemented a custom simulator. This simulator reads the contact trace and simulates the transmission of a message between a randomly selected pair of nodes in the network. The simulation finishes when the message reaches the destination node. This evaluation is repeated 1000 times (that is, 1000 different traces are generated) in order to obtain a mean value. We also obtain the ratio of contacts on the simulator, that is similar to the ratio of contacts (R_C) obtained for evaluating the representativeness of the distributions. In case a simulation ends without the message being delivered to the destination node, we count this as a miss, and when the experiments end, we obtain the delivery ratio as $(1000 - \text{misses})/1000$.

In the real trace experiments the process is very similar, except that in this case we used a real traffic trace as the input of the model and simulator. In order to evaluate the precision of the different distributions depending on the duration of the traffic trace, the original traffic trace is trimmed from T ranging from 0 to the maximum duration (that is, only the contacts with time $t \leq T$ are used). Finally, although the traffic trace is always the same, the simulation is repeated 1000 times, varying the pair of selected sender and destination nodes of the network.

4.1 Synthetic traces

The following experiments used synthetic traces as a way to validate the conclusions derived about representativeness and precision in sections 2 and 3. In Figure 3a we show the delivery delay time for the different distributions calculated using a synthetic trace with time T ranging from 1 to 200s that was generated assuming that all inter-contact times distributions between pairs are exponentially distributed with the same contact rate $\lambda_p = 0.1$ (that is, all distributions are homogeneous).

We can see that AN and AC distributions obtains the best results, converging very fast to the simulation results (for example, for $T = 200s$, the delay obtained through simulation is 0.918s and the delays obtained using the AN and AC distributions are the same: 0.905s, a relative error of 1.42%). Instead, the AP converges to the simulation results very slowly. As expected, these results clearly depend on the representativeness of the different distributions as shown in Figure 3b. The AP contact ratio is very low for a reduced trace time (T), reaching 1 for $T = 100$. Nevertheless, for this trace time the average number of measures is still very reduced ($M_C = 9.6$ is for $T = 100$ in the AP distribution compared to $M_C = 765.86$ in the AN distribution), so the precision is also low. Thus, as expected, the conclusions are clear, the most precise characterizations are the Aggregate Nodes (AN) and Any Contact (AC) distributions.

In the following experiment we study the case of heterogeneous exponential distributions. We generate a synthetic contact trace with the contact times between pair of nodes exponentially distributed, where the contact rate λ_p is also exponentially distributed with mean 0.1. The delay of a message using epidemic transmission is shown in Figure 3c. In this case we can see that the delay obtained using the AC and AN distribution is very precise (a relative error of 2.34% for $T = 200s$). Nevertheless, the delay obtained for the AP distribution converges to the solution very low. The reason is that, for an heterogeneous distribution the representative-

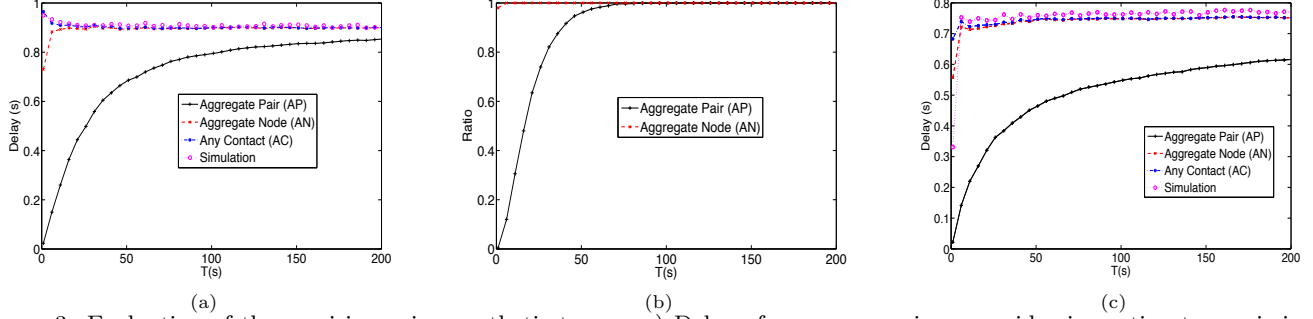


Figure 3: Evaluation of the precision using synthetic traces. a) Delay of a message using an epidemic routing transmission with a homogeneous trace, b) Contact ratio of the homogeneous trace (R_C), c) Delay with a heterogeneous trace

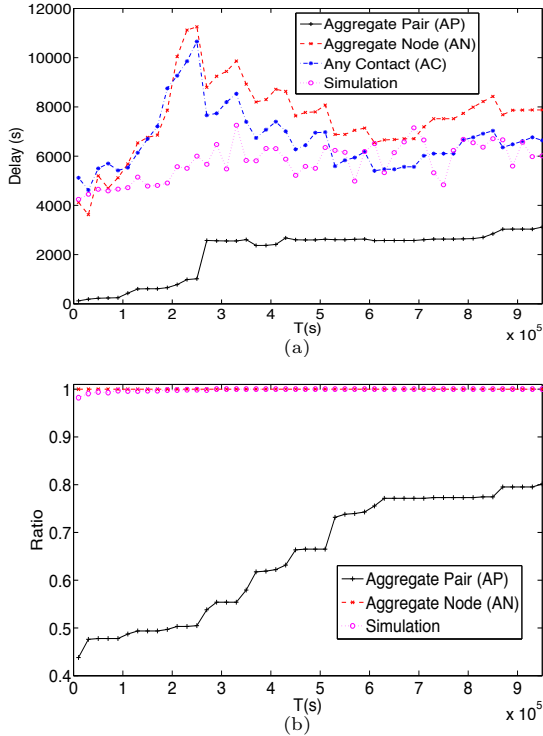


Figure 4: Evaluation of the precision using the Cambridge traces. a) Delay of a message using an epidemic transmission, b) Contact ratio (R_C).

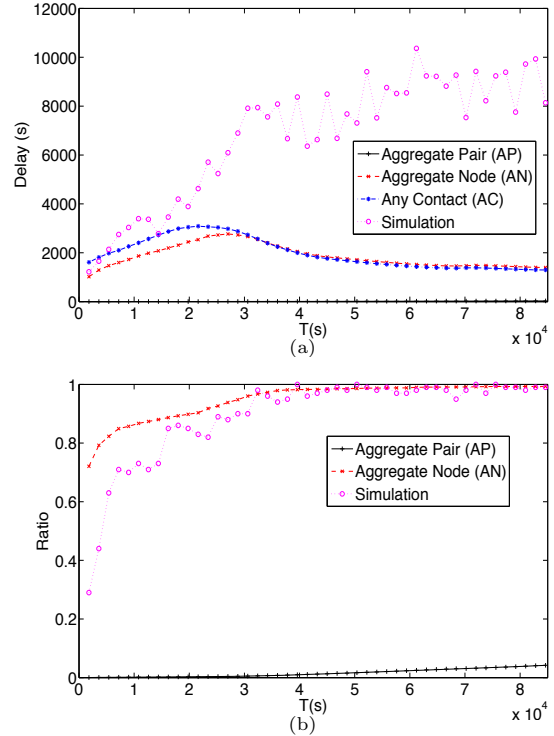


Figure 5: Evaluation of the precision using the Shanghai traces. a) Delay of a message using an epidemic transmission, b) Contact ratio (R_C).

ness of the AP distribution is lower than in the homogeneous case (for example, for $T = 200s$, $R_C = 0.92$ and $M_C = 23$). The rate of contact reaches 1 for approximately $T = 2000s$ with $M_C = 233$, and a relative error of 7.24%.

4.2 Real traces

We now, evaluate the precision of the different distributions using real traffic traces.

Figure 4 shows the results for the Cambridge trace depending on the duration T (that is, a trimmed traffic trace with duration T). Regarding the delay, the results obtained using the AC and AN distributions are close to the simulation results. There is a strong time dependence of this trace (the sawtooth shape), due to the high difference between contact rates in diurnal and night periods. Regarding the

AP distribution, the precision increases as the contact ratio also increases (see Figure 4b), but it is very far of the results obtained using the other distributions.

Figure 5 shows the results for the Shanghai traffic traces. In this case, the delay obtained using the AN and AC distributions for traffic trace length T less than 20000 (about 5 hours) are close to the simulation ones, although the results for greater values of T show a higher error. The results using the AP distribution are quite poor, due to the low representativeness of this distribution (see Figure 5b). We can also observe how the contact ratio of the simulation is low for values less than 10000s, so the representativeness of the simulation is also low.

The conclusions drawn from the previous experiments for the AN and AC distributions are clear: the greater the con-

tact trace duration, the lesser the precision. This is mainly due to the long-tail behaviour of the inter-contact times distribution between the same pair of nodes. If the duration of the trace is high (days in the case of Cambridge, or hours in the case of Shanghai), the probability of contact between some reduced number of pair of nodes is very low, so its inter-contact time is very high. In other words, the greater the duration of the trace, the greater the tail of the distribution⁵.

We repeated the previous experiments using other performance models (such as the Two Hop multicopy protocol, and a more sophisticated model for evaluating the selfish detection introduced by the authors [6]), obtaining similar results in terms of precision.

5. CONCLUSIONS

In this paper we introduced new approaches to characterize the inter-contact times distribution in order to increase the representativeness, and thus the precision of the analytical performance models based on these distributions.

We have seen that the established characterization, the Aggregate Pairs (AP) distribution, has a very low representativeness especially when the number of nodes is high. This leads to poor results when applied to analytical models (such as the Epidemic routing). Therefore, we introduced and evaluated two alternative methods for characterizing the inter-contact times distribution: the Aggregate Nodes (AN) and Any-Contact (AC). The resulting distributions have an excellent representativeness for shorter trace durations. We also obtained a simple relation between the individual contact rate and the contact rate for the AP and AN distributions. Thus, we can use these distributions in order to obtain the individual pair contact rate used in analytical models.

The experiments confirm, that the representativeness and precision achieved using these new distributions is high. Using synthetic and real traces we showed that the Aggregate Nodes (AN) and Any-Contact (AC) distributions are more precise than the Aggregate Pairs (AP) distribution. Finally, regarding which distribution (AP or AC) choose, we found that the AC distribution is able to provide better results for very low trace sizes. However, when the number of contacts is moderately large, the results for both distributions are very similar.

Acknowledgments

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

6. REFERENCES

[1] H. Cai and D. Y. Eun. Crossing over the bounded domain: From exponential to power-law intermeeting time in mobile ad hoc networks. *IEEE/ACM Trans. on Networking*, 17(5):1578–1591, oct. 2009.

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. on Mobile Computing*, 6:606–620, June 2007.

[3] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '09, pages 299–308, 2009. ACM.

[4] R. Groenevelt, P. Nain, and G. Koole. The message delay in mobile ad hoc networks. *Performance Evaluation*, 62:210–228, October 2005.

[5] E. Hernández-Orallo, M. D. Serrat, J.-C. Cano, C. M. T. Calafate, and P. Manzoni. Improving selfish node detection in MANETs using a collaborative watchdog. *IEEE Communications Letters*, 16(5):642–645, 2012.

[6] E. Hernández-Orallo, M. D. Serrat Olmos, J.-C. Cano, C. T. Calafate, and P. Manzoni. Evaluation of collaborative selfish node detection in MANETS and DTNs. In *Proceedings of the 15th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, MSWiM '12, pages 159–166, 2012. ACM.

[7] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, WDTN '05, pages 244–251, 2005. ACM.

[8] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '08, pages 241–250, 2008. ACM.

[9] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović. Power law and exponential decay of inter contact times between mobile devices. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, MobiCom '07, pages 183–194, 2007. ACM.

[10] M. Karaliopoulos. Assessing the vulnerability of dtn data relaying schemes to node selfishness. *IEEE Communications Letters*, 13(12):923–925, dec 2009.

[11] Y. Li, G. Su, D. Wu, D. Jin, L. Su, and L. Zeng. The impact of node selfishness on multicasting in delay tolerant networks. *IEEE Trans. on Vehicular Technology*, 60(5):2224–2238, jun 2011.

[12] A. Passarella and M. Conti. Characterising aggregate inter-contact times in heterogeneous opportunistic networks. In *Proceedings of the 10th international IFIP TC 6 conference on Networking*, Networking'11, pages 301–313, 2011. Springer-Verlag.

[13] A. Passarella, M. Conti, C. Boldrini, and R. I. Dunbar. Modelling inter-contact times in social pervasive networks. In *Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, MSWiM '11, pages 333–340, 2011. ACM.

[14] X. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. *Computer Networks*, 51(10):2867–2891, 2007.

[15] H. Zhu, L. Fu, G. Xue, Y. Zhu, M. Li, and L. M. Ni. Recognizing exponential inter-contact time in VANETs. In *Proceedings of the 29th conference on Information communications*, INFOCOM'10, pages 101–105, 2010. IEEE Press.

⁵Note that we are studying the average behaviour, thus, if we made a good estimate of the tail distribution, it can lead to wrong average results