From Pareto inter-contact times to residuals

Chiara Boldrini, Marco Conti, *Member, IEEE* and Andrea Passarella

***Abstract*—Interactions between mobile users are the building blocks of a variety of emerging communication paradigms, among which opportunistic networking is one of the most promising. In opportunistic networks, the information propagates through pair-wise contacts between users, and hence the inter-contact time, i.e., the time between two consecutive interactions between a pair of users, plays a key role in the latency of information propagation. Given that new message availability and actual communication are typically asynchronous, analytical models often rely on the concept of residual inter-contact time, i.e, the time *left* before the next communication opportunity, starting from a random point in time. The statistical properties of the inter-contact times determine those of the associated residual inter-contact time. Of particular interest is the case of inter-contact times featuring a Pareto distribution, due to the great attention this case has received in the literature. In this letter we discuss how to compute the residual inter-contact time when the inter-contact process between a pair of nodes features a Pareto distribution and we show that our exact solution can significantly improve the results commonly used in the literature.**

***Index Terms*—inter-contact times, residuals, renewal process, opportunistic networks, delay tolerant networks**

I. Introduction

INTERACTIONS between mobile users have become a key aspect of future Internet communications. Direct encounters between user devices (which follow the movements of their human owners) are exploited in opportunistic networks [1] in order to deliver messages across multi-hop paths where the user devices themselves act as relays. In this scenario, a model of user interactions is essential in order to characterize the information delivery process, in terms of, e.g., the message delay. Using the term *contact* to generically denote the relevant kind of interaction with respect to the networking scenario considered, the starting point of an interaction model are inter-contact times, defined as the time between two consecutive interactions between the same pair of nodes. However, given that new message availability is an asynchronous process with respect to the contact process between any pair of nodes, and that any forwarding action is asynchronous with respect to the contact process of all pairs but the one involved in that particular action, we have always to wait for the next interaction for communications to take place. Thus, we are generally interested in the *residual* inter-contact time rather than in the inter-contact time itself. Given a random time t_r and a pair i, j of nodes, the residual inter-contact time is defined as the time interval, starting from t_r, before the next interaction between node i and node j takes place.

In the case of opportunistic networks, the interaction model should mimic the encounter process between users. In the recent years, numerous efforts to describe the statistical properties of the human encounter process have been made. Among these, Chaintreau et al. [2] have shown that the Pareto[1] distribution can be considered a good approximation for modelling the contact process between users, and that the presence of a heavy tail might seriously affect the performance of routing protocols in terms of expected delay experienced by messages. Due to the impact such Pareto hypothesis has had in the literature on opportunistic networks, in this letter we discuss how to derive the residual inter-contact times for an interaction model based on Pareto distributed inter-contact times.

II. Pareto distributions

There are two popular versions of the Pareto distribution [3]. The first one, commonly referred to as European Pareto, is described by the following CCDF:

$$\bar{F}_e(t) = \left(\frac{b}{t}\right)^{\alpha}, \quad t > b \tag{1}$$

where $b > 0$ is the lower bound of the Pareto distribution and $\alpha > 0$ is the shape. The American Pareto cumulative distribution function is instead given by:

$$\bar{F}_a(t) = \left(\frac{b}{b+t}\right)^{\alpha}, \quad t > 0. \tag{2}$$

Basically, being X a random variable following a European power law with lower bound b and shape α, then $Y = X - b$ is an American power law random variable. The main difference between the two versions is that the American Pareto allows for t values arbitrarily close to zero, while the European Pareto constrains t to be above a threshold b.

III. Computing the residual

Assuming that inter-contact times are described by the random variable M, in [4] a general formula for the derivation of residual inter-contact time is provided, which we recall in Equation 3.

$$\bar{F}_R(t) = \frac{1}{E[M]} \int_t^{+\infty} \bar{F}_M(u) \mathrm{d}u \tag{3}$$

While everything works smoothly when considering the American Pareto distribution, when inter-contact times feature a European Pareto the application of this formula may be tricky.

This work was partially funded by the European Commission under the SCAMPI (FP7-FIRE 258414) and RECOGNITION (FET-AWARENESS 257756) projects.

C.Boldrini, M.Conti, and A.Passarella are with the Institute for Informatics and Telematics of the National Research Council of Italy, Via G. Moruzzi 1, 56124 Pisa, Italy.

E-mail: chiara.boldrini, marco.conti, andrea.passarella@iit.cnr.it

[1] In the following we will use the terms "power law" and "Pareto" interchangeably.

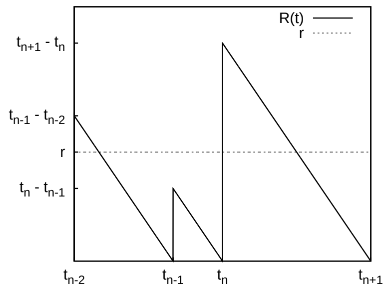

Fig. 1. An Example Realization of the Renewal Process

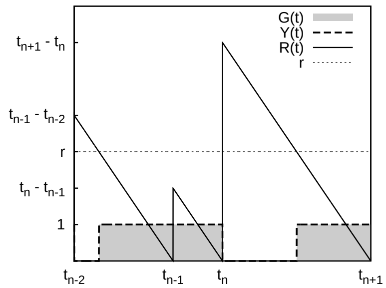

Fig. 2. The Reward Process

Given the widespread use of the European Pareto [5] [2], in the following we show how to deal with this case.

Theorem 1: When inter-contact time M features a European Pareto distribution with shape α and lower bound b ($F_M(t) = \left(\frac{b}{t}\right)^{\alpha}, t > b$), the residual inter-contact time R is distributed as follows:

$$F_R(t) = \begin{cases} \frac{t-\alpha t}{\alpha b} + 1 & t > 0 \wedge t \leq b \\ \frac{1}{\alpha}\left(\frac{b}{t}\right)^{\alpha-1} & t > b \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Proof: We model the contact process as a renewal process [6]. In our case, inter-arrival times $X_1, X_2, \ldots$ follow a European Pareto with shape α and lower bound b. Under this model, the residual inter-contact time corresponds to the residual lifetime before the next renewal event. The residual lifetime is a random function of time, i.e., a stochastic process, and we denote it as $\{R(t), t > 0\}$. A realization of the renewal contact process and its associated residual lifetime process is given in Figure 1, where t_n denotes the arrival time of the n-th event, i.e., the time at which the n-th contact between two nodes takes places. In order to derive the distribution of $R(t)$, we focus on its CDF $P(R(t) < r)$, for any r greater than zero. As can be seen in Figure 1, the residual lifetime $R(t)$ is smaller than r when $R(t)$ lies below threshold r. Thus, let us define a reward function $Y(t)$ that takes value 1 when $R(t)$ is below threshold r, and 0 otherwise (Figure 2). $P(R(t) < r)$ can be interpreted as the fraction of time the process is gaining a reward (i.e., $Y(t) = 1$) with respect to the whole process lifetime. The time interval in which the process is gaining a reward is simply the area below curve $Y(t)$, because by definition we gain a reward every time unit $R(t)$ lies below threshold r. For obtaining $P(R(t) < r)$, we need to divide this area by the duration t of the process lifetime. Thus, Equation 5 holds.

$$P(R(t) < r) = \lim_{t\to\infty} \frac{1}{t}\int_0^t Y(u)\mathrm{d}u \tag{5}$$

Quantity $\int_0^t Y(u)\mathrm{d}u$, which we hereafter denote as $G(t)$, gives the accumulated reward up to time t. $G(t)$ can be expressed in terms of the reward G_n accumulated during each single renewal interval X_n, thus obtaining $G(t) = \sum_{n=1}^{N(t)} G_n$, where $N(t)$ gives the number of events up to time t. It is a well known result from theory on renewal-reward processes that $G(t)$ (the total accumulated reward at time t), $E[G_n]$ (the expectation of the reward G_n accumulated in a generic inter-renewal interval X_n), and $E[X_n]$ (the expected duration of an inter-renewal interval X_n) relate to each other according to $\lim_{t\to\infty}\frac{G(t)}{t} = \frac{E[G_n]}{E[X_n]}$ [6]. Thus, the following relation holds true:

$$P(R < r) = \frac{E[G_n]}{E[X_n]}. \tag{6}$$

If we focus on the reward G_n accumulated in a generic inter-arrival interval X_n, we have that, by definition, $G_n = \min(r, X_n)$, as highlighted in Figure 2. The expectation of G_n can be computed as

$$E[G_n] = \int_0^{\infty} P(\min(X_n, r) > x)\mathrm{d}x, \tag{7}$$

after noting that $\min(r, X_n)$ is by definition a random variable that can only take non-negative values. Please also recall that in a renewal process an inter-renewal interval X_n follows the same distribution for all values of n. In addition, $\min(r, X_n)$ is defined only in the interval $[0, r]$ (i.e., $P(\min(r, X_n) > x \mid x > r) = 0$). In such $[0, r]$ interval, the reward that can be accumulated is, by definition, at most equal to the length of the inter-renewal interval X_n, and thus $P(\min(r, X_n) > x \mid x \leq r) = P(X_n > x \mid x \leq r)$. We can then rewrite $E[G_n]$ as $\int_0^r P(X_n > x)\mathrm{d}x$. However, when X_n follows a European Pareto distribution there can be no inter-renewal interval (or, equivalently, inter-contact time) smaller than b, which means that the accumulated reward increases linearly as long as r is smaller than b. Thus, we have that, if $r \leq b$, $E[G_n] = \int_0^r P(X_n > x)\mathrm{d}x = \int_0^r 1\mathrm{d}x$, while, for $r > b$, $E[G_n] = \int_0^b 1\mathrm{d}x + \int_b^r \left(\frac{b}{x}\right)^{\alpha}\mathrm{d}x$. After integration, we obtain Equation 8.

$$E[G_n] = \begin{cases} r & r > 0 \wedge r \leq b \\ \frac{r^{-\alpha}(\alpha b r^{\alpha} - b^{\alpha} r)}{\alpha - 1} & r > b \end{cases} \tag{8}$$

Equation 4 follows directly after substituting Equation 8 into Equation 6 and recalling that i) $E[X_n]$ is equal to the expectation $\frac{\alpha b}{\alpha-1}$ of the European Pareto distribution and ii) $F_R(t) = 1 - P(R < t)$. ■

Remark 1: Note that the result in Equation 4 differs from that in [4].

In Figures 3 and 4 we plot the CCDF predicted by Theorem 1 and the CCDF provided by [4], and we compare them against the CCDF of the residual inter-contact time obtained from simulations[2]. More specifically, for the latter

[2]To this aim, we have used the statistical software R [7] and the routines for Pareto distributions provided by [5].

we simulated the arrival of a random observer for a renewal process with inter-arrival times that feature a European Pareto distribution with $\alpha = 2$ and $b = 1$. A random observer is one that arrives at a random point in time with respect to the evolution of the underlying renewal process. Then, we measured the residual inter-contact time as the amount of time this observer has to wait until the next renewal (i.e., contact) event. As shown in Figures 3-4, Theorem 1 characterizes exactly the residual inter-contact time distribution both in the head (Figure 3) and in the tail (Figure 4).

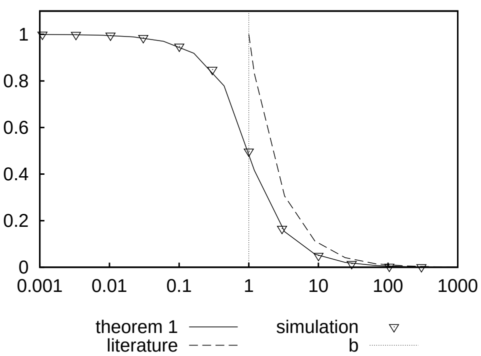

Fig. 3. Residual Inter-contact Time - Lin-Log Scale

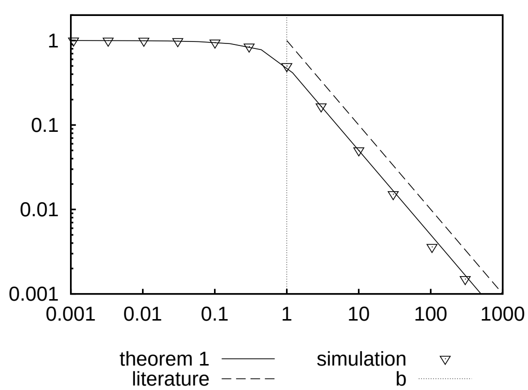

Fig. 4. Residual Inter-contact Time - Log-Log Scale

IV. A SIMPLE CASE STUDY

In this section we provide a practical example in which knowing the exact distribution of the residual inter-contact time drastically improves the knowledge on how the delivery process behaves. More specifically, we discuss the simple case of an opportunistic network in which the Direct Transmission [8] routing protocol is in use. According to the Direct Transmission protocol, the source of a message can only hand over the message directly to the destination, if ever encountered. In this case, the probability distribution of the message delay follows exactly that of the residual inter-contact time between the source node and the destination node, because the time it takes for the source to meet the destination, starting from the random point in time at which the message was generated, is the time it takes for the message to be delivered. We assume that inter-contact times feature a European Pareto distribution with shape $\alpha = 2$ and lower bound $b = 1$, which is the same configuration we used for plotting the figures in Section III. Given that the delay and the residual inter-contact time feature the same distribution, Figures 3 and 4 also provide the CCDF of the delay experienced by messages delivered according to the Direct Transmission policy. Let us focus on $P(R > 1)$, i.e., the probability that messages have to wait more than 1 second to be delivered. Using the result in [4], we obtain that $P(R > 1) = 1$, while $P(R > 1) = 0.5$ if we apply Theorem 1. This example shows that, relying on the formula commonly used in the literature, we could seriously underestimate the presence of small delay values, and thus inaccurately characterize the delivery process, even in the simplest scenarios.

V. CONCLUSION

In this letter we have discussed how to compute the *residual* inter-contact time starting from Pareto distributed inter-contact times, referring to opportunistic networks as the most immediate field of application. However, the applicability of our results goes well beyond opportunistic networking. In fact, inter-contact times, and thus their residuals, emerge naturally as one the most important metrics when considering models of interactions between users. Be it the exchange of messages via direct encounters in an opportunistic network, or Web 2.0 communications through virtual contacts in an Online Social Network (such as Facebook and Twitter), an information delivery model cannot leave aside an appropriate interaction model, and the omnipresence of the Pareto distribution [9] [10] [11] [12] makes it essential to be able to accurately characterize analytically also the case of power law distributions. Even if the power law predominance is being questioned by many (e.g., [5]), surely the Pareto distribution is to be confronted with when studying networks of users and their interactions.

REFERENCES

[1] M. Conti and M. Kumar, “Opportunities in opportunistic computing,” *Computer*, vol. 43, no. 1, pp. 42–50, 2010.
[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*, vol. 6, no. 6, pp. 606 –620, 2007.
[3] M. Rytgaard, “Estimation in the Pareto distribution,” *ASTIN Bulletin*, vol. 20, no. 2, pp. 201 – 216, 1990.
[4] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, “Power law and exponential decay of intercontact times between mobile devices,” *IEEE Trans. on Mobile Computing*, vol. 9, no. 10, pp. 1377 –1390, 2010.
[5] A. Clauset, C. Shalizi, and M. Newman, “Power-law distributions in empirical data,” *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.
[6] S. Ross, *Introduction to probability models*. Academic Press, 2007.
[7] The R Project for Statistical Computing. [Online]. Available: http://www.r-project.org/
[8] M. Grossglauser and D. Tse, “Mobility increases the capacity of ad hoc wireless networks,” *IEEE/ACM Trans. on Networking*, vol. 10, no. 4, pp. 477–486, 2002.
[9] W. Aiello, F. Chung, and L. Lu, “A random graph model for massive graphs,” in *Proceedings of the 32nd annual ACM symposium on Theory of computing - STOC'00*, 2000, pp. 171–180.
[10] S. Redner, “How popular is your paper? An empirical study of the citation distribution,” *The European Physical Journal B*, vol. 4, no. 2, pp. 131–134, 1998.
[11] A. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, “Evolution of the social network of scientific collaborations,” *Physica A: Statistical Mechanics and its Applications*, vol. 311, no. 3-4, pp. 590–614, 2002.
[12] C. Shirky, *Power Laws, Weblogs and Inequality*, 2003.