Popularity of Nodes in Pocket Switched Networks

Abderrahmen Mtibaa
Thomson
Paris, France
abderrahmen.mtibaa@thomson.net

Augustin Chaintreau
Thomson
Paris, France
augustin.chaintreau@thomson.net

Christophe Diot
Thomson
Paris, France
christophe.diot@thomson.net

1. INTRODUCTION

Pocket Switched Networking (PSN) is a new communication paradigm for mobile devices [1]. It takes advantage of every communication opportunity, and the physical mobility of the devices, in order to transport data to destinations. Efficient forwarding in this context today remains challenging.

When mobile devices are carried by a group of people in the same area, one should expect each device to connect multiple subgroups together, and a fraction of devices to connect many of them. In other words, the "popularity" of a device as a relay between sources and destinations may vary a lot.

In this work, we use real measurements of human mobility to study node's popularity, as a first step towards understanding the impact of popularity on forwarding algorithms.

2. METHODOLOGY

We present results using data from an experiment organized during *Infocom05*. Similar results have been observed for three others data sets but could not be reported here. In this experiment, 41 students attending the Infocom conference carried iMotes for 3 days. The iMotes collected MAC addresses of others Bluetooth devices that came into range, together with a time-stamp of this contact [1].

Connectivity, defined using the total number of contacts per node, may not characterize popularity, as a node could be highly connected with a small set of neighbors. We introduce a new metric of node popularity: the occurrence (the number of time a node appears as a relay) in delay-optimal paths. There may be several variants of this metric depending on the choice of the sources and the destinations (see below).

We represent the data set as a graph where each edge represents a contact and includes a time value (this graph is also called a temporal network). We computed offline and for all departure times the set of delay-optimal paths in this network, where a path may use either a single edge, that is a direct contact, or a sequence of contacts in a time-respecting manner. We refer to all delay-optimal paths starting from a source i, to all possible destinations, as all $(i,*)$ paths. More generally, we define the set of paths $(i,*)$, $(*,j)$, $(*,*)$ and (i,j). We define the popularity of a node as the occurrence of this node as an intermediate relay in one of these sets.

3. MEASURING POPULARITY

First, we measure the popularity of nodes among all the paths starting from source 1. In Figure 1 (a) relays were ordered along the x-axis in decreasing order of their occurrences, we plot the corresponding values. We also present the fraction of total occurrences that

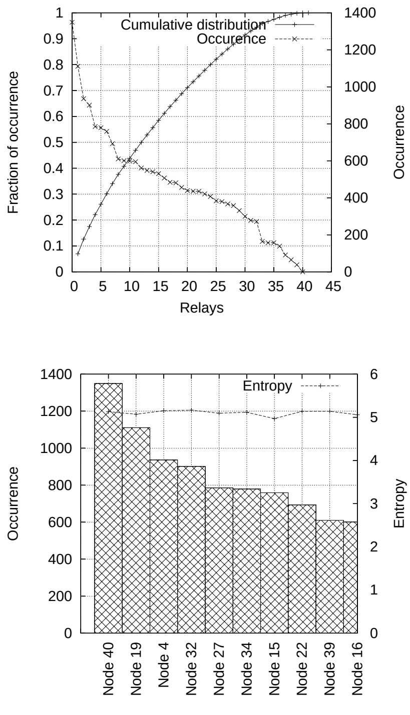

Figure 1: Popular relays for (1,*) paths, (a) Relay distributions. (b) 10^{th} popular relays

are found among the first n relays, or "cumulative distribution". Note that the 12 best relays represent half of the occurrences found in the paths that forward packets coming from a source 1. This result indicates that with only 30% of nodes serving as relays, node 1 may be able to reach a destination with an optimal path roughly 50% of the time. We observe also that the most popular relay is used 1350 times to forward packets created by node 1, twice more than the 10^{th} most popular relay. We present in Figure 1 (b) a zoom for the 10 most popular relays. We show that nodes 40, 19, 4 and 32 are the four most frequently used. They account for almost 25% of the total occurrence. These relays are also the 4 most popular relays for $(*, 1)$ paths, although not exactly in the same order. The symmetry we observe means that these popular nodes are preferentially used by node 1, independently of the direction. Similar results have been obtained for all others source nodes (from 2 to 41).

The occurrence of a node measures its overall popularity when packets are sent from source 1 but it does not describe how this popularity is distributed among several destination. In particular a node with a large overall occurrence may be highly popular for a few destinations, or it may be moderately popular uniformly among all destinations. To distinguish between these two cases, we define the entropy function for a relay R as

$$E(R) = -\sum_{j=1}^{n} p(j)\log(p(j)) \ , \qquad (1)$$

where $p(j)$ represent the fraction of occurrence of R that are found towards the destination j. Values of entropies for most popular relays can be found in Figure 1 (b). As can be observed, entropies for these popular nodes are very high (close to a maximum entropy value, that is $\log(40) = 5.33$). In other words frequently used relays are popular almost uniformly among all destinations. This means that overall occurrence measures well in that case the importance of a relay at the network level.

We know consider occurrence of nodes as relays for all source-destination pairs (i.e. for $(*, *)$ paths). We show in Figure 2 (a) that the 12 most popular relays account again for half of the total occurrence, they are described in more details in Figure 2 (b). Nodes 40, 4 and 19, and 32 are responsible for more than 22% of total occurrences found in delay-optimal paths for this data set. Note that this is consistent with the previous results found when all paths start from node 1. Using the entropies metric, we have observed that these nodes are those with the highest entropy, although the distribution among source destination pairs is not uniform.

Others results have been obtained, including popularity of links (i.e. occurences of a sequence of two consecutive nodes in a path). We can summarize them

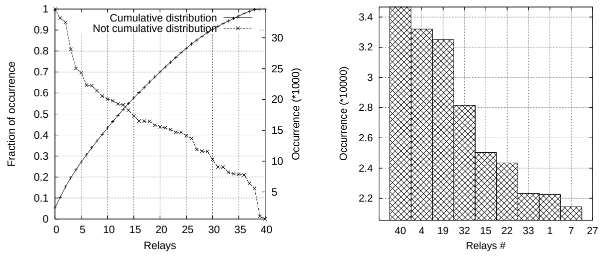

Figure 2: (a) Relay distributions, (b) Popular relays for $(*, *)$ paths

as follows. (1) A few popular links (less than 1%) contribute to more than 50% of the total occurrence of links among delay-optimal paths, and these popular links have the highest entropy values, (2) Popular links almost always contain a popular relay (like the one we extracted above) and some other node: Nodes 40, 19 and 4 appear in most of best popular links. (3) Popular links occur almost the same number of times in the two directions. These links represent solicited links in the network. They are frequently used in the two directions; they represent relevant patterns for efficient paths in this data set.

4. SUMMARY AND DISCUSSIONS

This work presents a new method to characterize popularity in a PSN environment. The occurrence of a node in delay-optimal paths allows to distinguish between nodes using network-wide information. In particular, it improves on a measure of a node's connectivity that could sometimes be misleading.

Using real world measurement, we show that nodes have different levels of popularity. Moreover, the most frequently used nodes from a given source are popular almost uniformly among all destinations. The same results seem to extend when one looks at all sources, and other type of popularity (like the popularity of a link defined by two relays). This indicates that selecting preferential relays or looking for relevant patterns to construct online paths may be done using this metric as an estimate.

It is important nevertheless to show the impact of these popular nodes on the network performance; popular node may be relevant but not critically needed for a network. In particular, it improves upon estimating the importance of a node via its total number of contacts.

5. REFERENCES

[1] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, *Impact of human mobility on the design of opportunistic forwarding algorithms*, in Proceedings of IEEE INFOCOM, 2006.