Towards a Clustering Based Data Diffusion Protocol In Delay Tolerant Networks

Sana Tmar
CITI - INSA de Lyon
sana.tmar@insa-lyon.fr

Eric Fleury
ENS Lyon / ARES
eric.fleury@inria.fr

ABSTRACT

In Delay Tolerant Networks (DTN), diffusion protocols can benefit from the users' mobility in order to reach some distant nodes. However, existing protocols like flooding present a rather high overhead because of redundant transmissions. The main contribution of this paper is to introduce a new diffusion protocol, called CBD (Clustering Based data Diffusion). Our solution provides to each node the possibility to adapt its diffusion strategy on-the-fly for a more efficient diffusion by evaluating its local density. We show through the use of a real measurements mobility that CBD brings an improvement in forwarding performances, in terms of delivery ratio, delay and the amount of generated messages.

1. INTRODUCTION

It was shown by previous work that the nodes' mobility in DTN can improve the delivery ratio and the delay [2]. However, taking profit of any contact opportunity without exploring any knowledge about relay nodes or local density can generate an important amount of traffic when forwarding messages. Some authors proposed the use of an historic of contacts, knowledge about location or labels to identify communities. But such informations cannot be offered to all mobiles and require usually an acquisition phase.

In this work, we use real measurements of human mobility to study the characteristics of the earliest diffusion tree. We exploit this results to define a new diffusion scheme which is based on two heuristics taking into account the mobile neighborhood. Finally, we compare our approach to five other diffusion protocols.

2. CONTRIBUTION

$CoNEXT$'07, December 10-13, 2007, New York, NY, U.S.A.

In this section, we present our results using data from an experiment organised in the *Infocom* conference [1]. Similar results have been observed for two other data sets but are not reported here. In this experiment, the mobile devices have been carried by a group of students during 3 days. We focus on the contacts between mobiles to compute the earliest diffusion tree for all nodes and at different times. The edges of this tree form the minimum delay paths between different peers.

First, we measure the number of triangles existing over edges which determines the number of common neighbors between two peers. We found that the edges forming the optimal delay paths are characterised by a low number of triangles, although the extremities of edges have in some cases an important degree. So, the probability to have a lot of neighbors in common between two given peers of the tree is very low.

Algorithm 1 The Neighborhood density heuristic

```
Entry: a vertex v, d (vertex degree)
dense ⇐ false
if d(v) < 3 then
    dense ⇐ false
else if d(v) ≥ 3andclustering.coefficient(v) > 0.7 then
    dense ⇐ true
end if
return dense
```

Second, we propose to use the clustering coefficient (cc) defined as a measure of the local density to characterise the relay nodes of the tree. The cc is the ratio of connected pairs among the neighbors of a vertex, averaged either over all vertexes or all pairs of neighbors [3]. We evaluate the cc for relay nodes. We observed that they are characterised by a low clustering coefficient. This means that the neighborhood of a relay is generally sparse at this moment.

Based on this two characteristics, we define two heuristics allowing mobiles to control the diffusion strategy dynamically.

The principle of CBD is based on two heuristics :

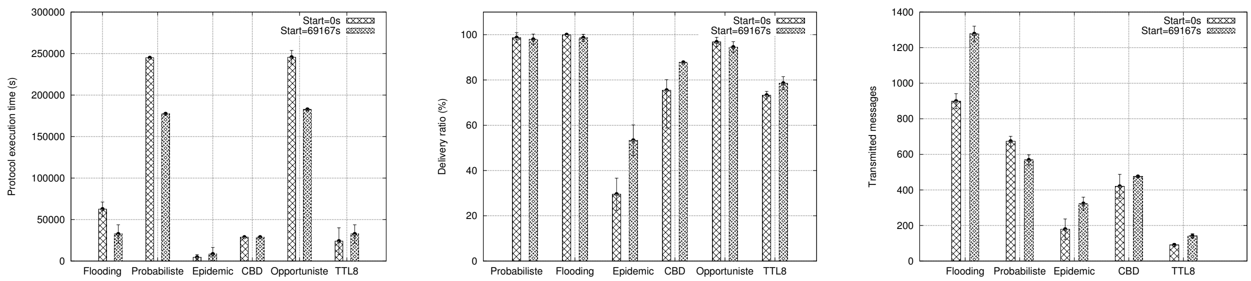

Figure 1: Comparison of (a)delay, (b)delivery ratio and (c)the amount of traffic of different algorithms

neighborhood density and *detection of a new community*. We assume that all mobiles are aware of their 2-hop neighbors through the use of hello messages.

With CBD, messages are transferred along a path in a store and forward manner. When a mobile have a message to forward, it proceed first by testing its local density using the neighborhood density heuristic as described by the algorithm 1. A message is forwarded only if the mobile neighborhood is not dense to avoid the broadcast storm problem.

Algorithm 2 The community detection heuristic

```
Entry: a neighbor v, a node u
newgroup ⇐ false
if triangles(u, v) = 0 and d(v) > 3 then
    newgroup ⇐ true
else if triangles(u, v) + 1 < d(v)/2 then
    newgroup ⇐ true
end if
return newgroup
```

To detect the local density, we use the clustering coefficient (cc). The maximum value of cc is fixed taking into account the results of the earliest diffusion tree. If the local zone density verify our defined condition (algorithm 1) then the mobile proceeds to detect if one of its neighbors is connected to another group using the second heuristic as described by the algorithm 2. A mobile decides if its neighbor relies to a new community in function of triangles number. In other words, the mobile evaluates the common mobiles with its neighbors. In the case where the mobile local topology isn't very dense and if one of its neighbors have more inter-communities links then intra-communities, the mobile forwards the message. The message is then flooded within the community. The mobiles receiving this packet will test the two heuristics described below and will continue the diffusion process.

3. SIMULATION RESULTS

We have implemented CBD and some other forwarding algorithms : based flooding, controlled flooding (using probability or a maximum hop number), the opportunistic routing (the source wait to reach all others mobiles) and the epidemic protocol (two nodes in contact compare their received messages and diffuse only when one have a missed bundle). We show the results about the *Infocom05* data set [1]. The diffusion is limited during the day and the maximum time to live for a stored message is fixed to 8 hours.

We have tested our solution in different times. We show our results in the beginning of the experiment and when the graph is very dense. In figure 1(a), we can see that CBD have a delay between simple flooding and controlled flooding using TTL. Moreover, we can observe in the figure 1(b) that CBD have a satisfactory delivery ratio which may reaches more than 80% of the total mobiles. In term of transmitted messages, we can see in figure 1(c) that CBD is close to the epidemic routing and the probabilistic protocol. Note that the epidemic routing has the lower delay, delivery ratio and the amount of created traffic because it has a fixed number of hops and the messages are in the proximity of the source : we can say that epidemic routing doesn't adaptable in diffusion case. Finally, CBD presents the best trade-off among the evaluated approaches.

4. CONCLUSIONS

The main contribution of this work has been the definition of a new diffusion scheme to avoid flooding at any contact opportunity. Our solution offers the possibility to make more intelligent the diffusion decisions by choosing the better mobiles to reach a new community without using external knowledge like labels, while ensuring a good performance in terms of delivery ratio, delay and generated traffic.

5. REFERENCES

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

[2] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions Networking*, 10(4):477–486, 2002.

[3] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684), June 1998.