Cache-and-Handover Algorithms in Wireless P2P Networks

Marco Fiore[†], Claudio Casetti[†], Carla-Fabiana Chiasserini[†]
Dipartimento di Elettronica, Politecnico di Torino, Italy
Corso Duca degli Abruzzi 24, I-10129 Torino (Italy)
Fax: +390115644099, Ph.: +390115644183, Email: firstname.lastname@polito.it

[†] EuroFGI member, Partner Number: 28, Partner Name: TNG-POLITO

***Index Terms*—Ad hoc wireless networks, information access, content distribution, mobile networks**

I. ABSTRACT

It is commonly acknowledged that user devices in the not-too-distant future will more and more resemble a communication hub, sporting arrays of GPS navigators, web browsers, videogame consoles and screens flashing the latest news or local sightseeing information. In this context, most pieces of information are likely to be of general use, and therefore a sensible dissemination and caching policy would be desirable.

In this work, we focus on such an environment: few and far between access points, or gateway nodes, in a highly-populated network area where user devices are equipped with a data cache and communicate through the ad hoc networking paradigm. User devices create a cooperative environment where information is exchanged among nodes in a peer-to-peer fashion. In particular, they form a pure peer-to-peer system, whose nodes may simultaneously act as both "clients" and "servers" to the other nodes in the network. Also, we envision a set of information categories that users may be interested in, such as web documents or data files, and associate each information with a desired spatial distribution. By spatial distribution, we mean the distribution according to which the information copies should be placed in the network geographical area. The nodes storing an information copy will act as providers for this content. Clearly, the information spatial distribution will depend on the information popularity as well as on the expected user density on the network area.

Traditional approaches to information caching in communication networks [1] are based on the solution of linear programming problems, which often require global knowledge on the network condition, or lead to quite complex solutions that involve significant communication overhead. Unlike previous approaches, our solution is fully distributed and it comes at a very low cost in terms of communication overhead. Our goal is to achieve the desired content distribution by properly letting the information move across the network.

More specifically, while developing our solution, we identify a number of issues that need to be addressed.

- *Achieving a desired distribution of the information within an area:* regardless of how the information is distributed at the outset, the system should be able to identify if and how frequently the information should be replicated and spread around to neighboring nodes.
- *Fair distribution of information burden:* as mentioned above, a node storing the information acts as provider for that information; of course, this role may exact a high toll from nodal resources in terms of bandwidth or power consumption; it is therefore advisable that the role of content provider be handed over to neighboring nodes quite frequently, without altering the information distribution.
- *Information survival:* regardless of the initial information distribution, and of the density of nodes, information should never be allowed to die out due to node isolation, transmission malfunctions or incorrect distribution. Related to the information survival is the evaluation of the minimum number of copies of a specific information that can satisfy users' needs (i.e., in terms of information retrieval time or response rate).

We focus on the first two issues, namely, achieving the desired distribution while fairly choosing nodes acting as information providers.

Our solution can single out the following advantages:

- it is fully distributed;
- it is content-transparent, i.e., it does not require knowledge on the contents stored by the neighboring users, thus preserving users' privacy;
- it works with minimum overhead.

In this paper, we investigate the applicability of two well-known mobility models, namely, the random walk and the random direction model in letting the information flow across the network. We present some preliminary results detailing the performance obtained by a random walk approach.

We highlight that in the context of sensor networks, approaches based on active queries following a trajectory through the network or agents propagating information on local events have been proposed, respectively, in [2] and [3]. However, both these works focus on the forwarding of these messages through the network, while our scope is to make the the desired information available by letting it move through nodes caches (minimizing cache replication across neighboring nodes). To the best of our knowledge, this is the first time such an approach is proposed.

A. The Information Mobility Models

We consider a tagged information and, as a first step, we take the desired spatial distribution to be uniform. To achieve the target distribution, we let the information move around according to the following mobility models.

- The *random walk mobility model* (RWM) is the discrete-time version of the Brownian Motion [5]. At each time step, mobile entities driven by the RWM movement pick a random direction and speed, and move accordingly until the following time step. It is typically employed, in a number of flavors, in graph theory, when dealing with graph visits. In our case, the graph under study matches the network physical topology. We consider the simplest random walk possible, in which each mobile entity, i.e. information copy, roams the network by moving from one node to a randomly selected neighbor. Each node caches the information for an exponentially distributed amount of time before handing it to the next hop in the information copy visit pattern. This approach requires trivial node operations and introduces minimal overhead, thus representing a lower-bound benchmark for more advanced information mobility models.
- The *random direction mobility model* (RD) [4] is often used to simulate the movement of user nodes in ad hoc networks. According to this model, each mobile entity alternates periods of movement (move phase) to periods during which it pauses (pause phase). Move and pause phases are often assumed to be exponentially distributed. At the beginning of each move phase, an entity independently selects its new direction and speed of movement; speed and direction are kept constant for the whole duration of the entity move phase. In [4], it has been shown that, if at time $t = 0$ the position and the orientation of the mobile entities are independent and uniform over a finite area, they remain uniformly distributed for all time instants $t > 0$, provided the entities move independently of each other.

  In our context, a mobile entity is a copy of the tagged information which “travels” from a user node to another according to the random direction model. The pause phase corresponds to the time period during which the entity is stored at a user node, while the move phase starts at the time instant when the current information provider hands over the content to one of its neighbors and drops it from its cache. As the current provider hands the information over, it will include in the transmitted message either the location of the final destination or the selected direction and travel time. In the first case, the user nodes have to be able to estimate their position (i.e., through GPS), which however is a fair assumption in most practical scenarios.

  Clearly, the information moves across user nodes, thus it may be transmitted along a direction that just approximates the selected one, or it may be stored at a node that is nearby (but not exactly at) the selected geographical destination. We therefore need to investigate the actual spatial distribution of the information that is obtained through our approach, how far from the uniform distribution is and how the proposed mechanism can be implemented in a simple but effective manner. Furthermore, we will study the impact of the parameters of the mobility model on the information spatial distribution.

B. Performance Evaluation Methodology

Since the goal of our study is to understand to which extent the information distribution achieved by a mobility model resembles the desired content diffusion, we need to define an ad-hoc metric. To this end, we leverage the distribution of information copies inter-distance. As a matter of fact, we can compare the measured inter-distance distribution against the theoretical distribution of the distance between two points, whose position is a random variable following the objective spatial distribution over the area within which the nodes are deployed.

The comparison metric we employ is the Kullback-Leibler (KL) divergence, defined as

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} P(x) log \frac{P(x)}{Q(x)} dx,$$

where $P(x)$ is the measured copies inter-distance distribution, and $Q(x)$ is the desired distribution.

Since in our analysis we consider a square area where the nodes are deployed and we seek a uniform dissemination of content, the target distribution is the solution to the bidimensional case of the hypercube line picking problem [6], which is known to be:

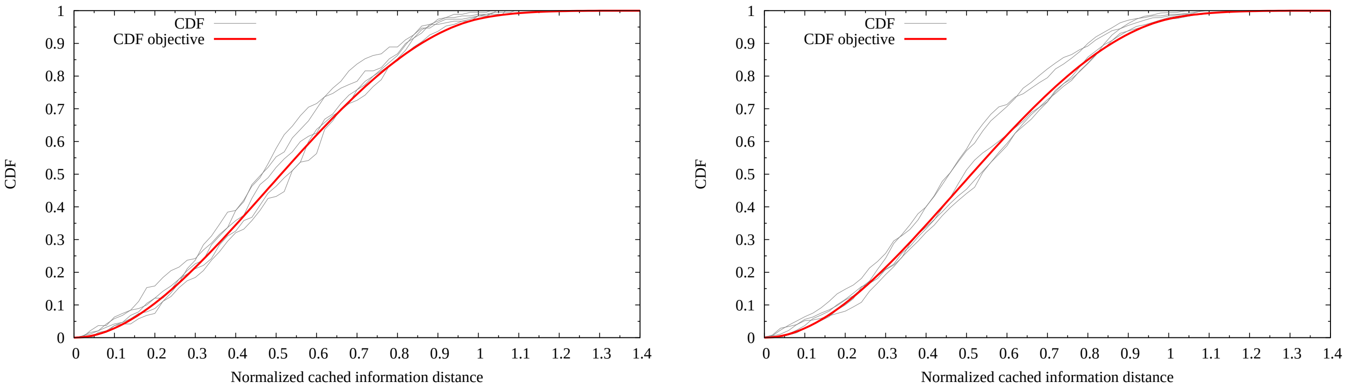
$$Q(x) = \begin{cases} \frac{1}{2}x^4 - \frac{8}{3}x^3 + \pi x^2 & \text{if } 0 \leq x < 1, \\ \left(\pi - 2 - 4\tan^{-1}\gamma + \frac{8}{3}\gamma\right) x^2 - \frac{1}{2}x^4 + \frac{4}{3}\gamma + \frac{1}{3} & \text{if } 1 \leq x < \sqrt{2}, \end{cases}$$

with $\gamma = \sqrt{x^2 - 1}$.

C. Preliminary results for the random walk case

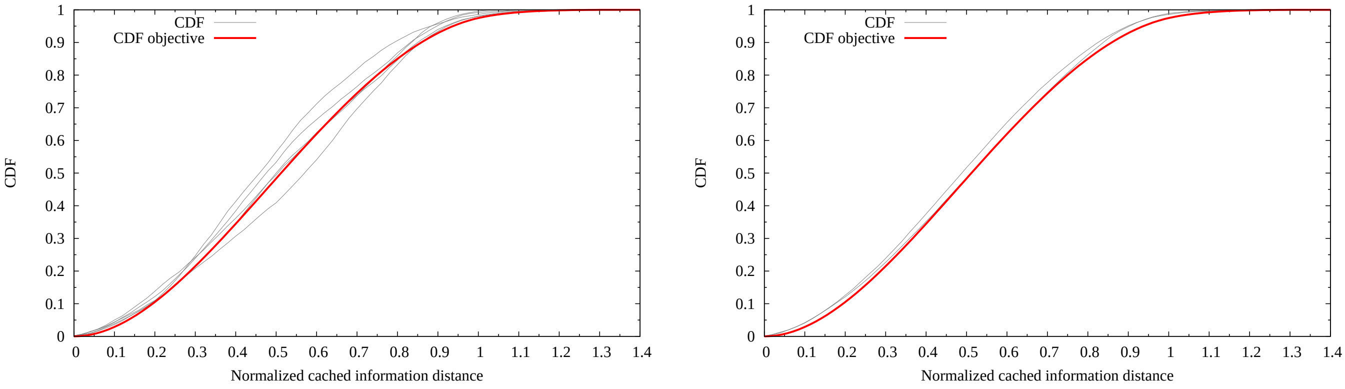
We consider a static network of 2000 nodes over a square area of 500 m side. Nodes are located over the simulated area according to an instance of random distribution, as shown in Fig. 1. Each node has a transmission range of 20m, resulting into 9 neighbors for each node on average. The number of information copies concurrently moving through the network sums to 20, and the caching time at each node is exponentially distributed around a mean of 10 s.

Fig. 2 shows the target cumulative distribution function (CDF) of the distance between information copies, corresponding to a uniform distribution of contents, versus realizations



(a) 10 s time granularity

(b) 50 s time granularity



(c) 500 s time granularity

(d) 5000 s time granularity

Fig. 2. CDF of the information copies inter-distance for different observation interval granularities

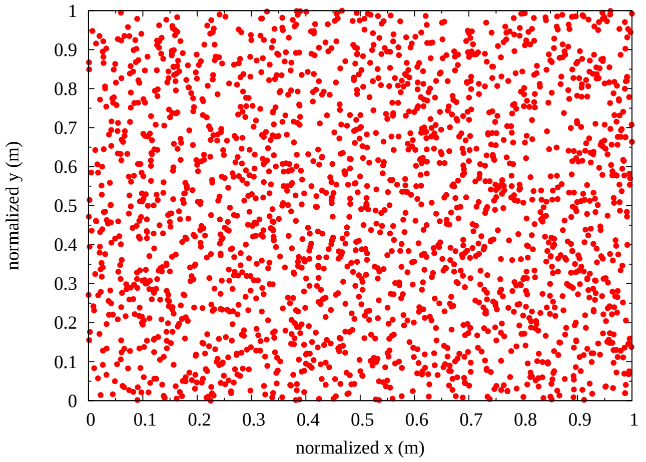


Fig. 1. Simulated nodes distribution

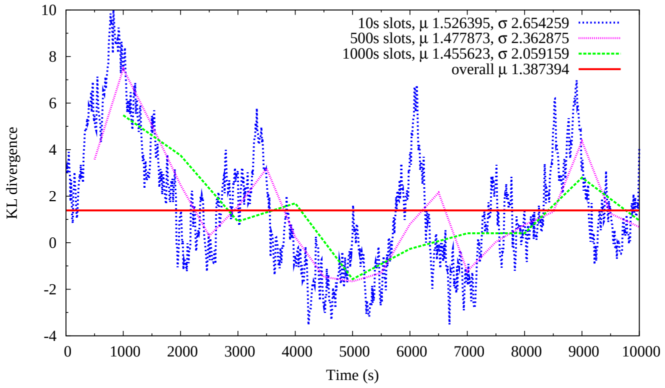


Fig. 3. Time evolution of KL divergence, for different observation interval granularities and 20 copies

of the observed equivalent CDF. In particular, each plot corresponds to a different observation interval: e.g., in the case of 10 s time granularity, each CDF is obtained from a 10 s observation, which is the mean caching time. These results show how uniform the distribution of the content appears at a given instant. On the other hand, the case of 5000 s time granularity shows CDFs as measured over 5000 s, i.e., circa 500 information movements. The other plots represent intermediate cases. We can notice that the obtained CDF closely approximates the target CDF and, as the observation interval granularity increases, the approximation becomes more and more accurate. This also gives us an idea of the amount of time we have to wait in order to achieve a distribution which can be considered uniform on average.

In Fig. 3, the evolution of the Kullback-Leibler divergence between the measured and objective CDF is plotted over

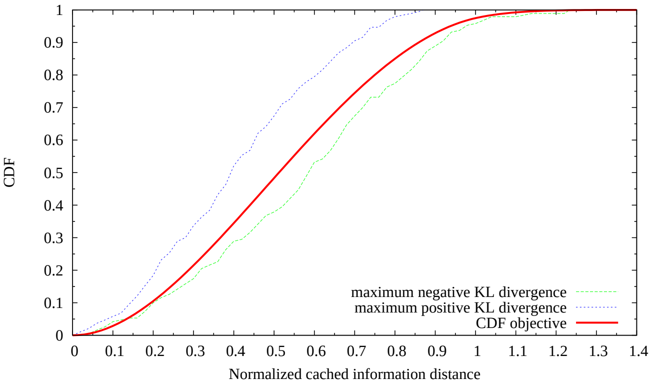


Fig. 4. Measured CDFs corresponding to maximum negative and maximum positive KL divergences with respect to the target CDF

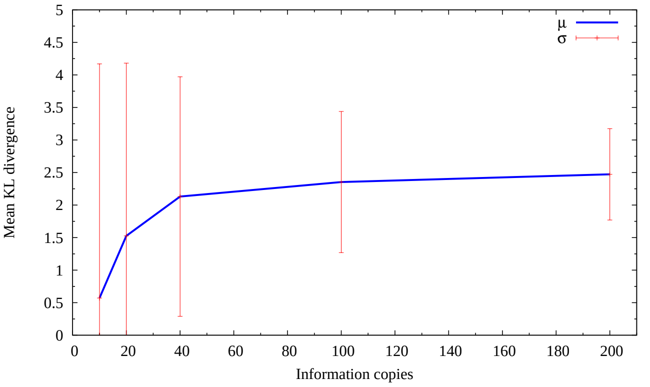


Fig. 5. KL divergence mean (μ) and standard deviation (σ) with 10 s observation intervals, for a different number of copies in the system.

time, as the granularity of the observation interval varies. We can see that the distance is obviously more subject to high oscillations when the observation intervals are short. Also, the mean (μ) and standard deviation (σ), reported in the legend of Fig. 3, decrease with increasing observation time granularity, thus confirming that longer observations lead to more uniform content distribution, on average.

In order to provide a visual representation of the magnitude of the KL divergence reached by the system, in Fig. 4 we compare the target CDF against the CDFs corresponding to the maximum negative and maximum positive values of KL divergence obtained over time. The results refer to the case where the observation time is set to 10 s.

Finally, Fig. 5 shows an interesting effect that takes place when increasing the number of copies. We observe that, as the number of information replicas cached in the network grows, the KL divergence becomes more stable, but the average skewfrom the desired distribution increases as well. The reason for this behavior lies in the location of nodes, which is static and presents different node density over the simulated area. The effect of an uneven density becomes more remarkable as the number of information copies increases, since the distribution of copies tends to resemble more closely that of nodes.

II. CONCLUSIONS AND FUTURE WORK

In this work we addressed the problem of achieving a desired distribution of information in a cooperative environment, using a low-overhead, content-transparent, distributed approach. Preliminary results have shown that the random walk approach is viable as long as the desired distribution is uniform. In the prosecution of this work, we are currently investigating the impact of random walk parameters on system performance, as well as the random direction approach.

REFERENCES

[1] P. Nuggehalli, V. Srinivasan, C.-F. Chiasserini, and R.R. Rao, “Efficient Cache Placement in Multi-hop Wireless Networks,” *IEEE/ACM Transactions on Networking,* Vol. 14, No. 5, pp. 1045–1055, Oct. 2006.

[2] N. Sadagopan, B. Krishnamachari, and A. Helmy, “Active Query Forwarding in Sensor Networks,” *Ad Hoc Networks Journal, Elsevier,* Vol. 3, No. 1, pp. 91-113, Jan. 2005.

[3] D. Braginsky and D. Estrin, “Rumor Routing Algorithm For Sensor Networks,” *First Workshop on Sensor Networks and Applications (WSNA),* Atlanta, GA, Sept. 2002.

[4] P. Nain, D. Towsley, B. Liu, and Z. Liu, “Properties of Random Direction Model,” *IEEE INFOCOM,* Mar. 2005.

[5] A. Einstein, “Investigations on the Theory of the Brownian Motion,” ed. R. Forth, translated by A.D. Cowper (1926, reprinted 1956), Einstein, Collected Papers, Vol. 2, pp. 170–82, pp. 206-22.

[6] M. Trott, “The Mathematica Guidebooks Additional Material: Average Distance Distribution,” 2004.