Bittella: Optimizing the download time and the search procedure on unstructured p2p networks based on Small Worlds

Rubén Cuevas, Carmen Guerrero, Carlos Navarro, Isaías Martinez-Yelmo
Departamento de Ingeniería Telemática. Universidad Carlos III de Madrid (Spain)
{rcuevas,guerrero,cnavarro,imyelmo}@it.uc3m.es

***Abstract*— In this paper we propose a new protocol for unstructured and semantic-searching based p2p networks: *Bittella*. It forms a three level overlay network: the lowest level is the unstructured p2p network (e.g. Gnutella). The medium level is formed by clusters per content, we call the resulting structure *Small World per Content structure*. This level allows the utilization of Bittorrent-like download technique in unstructured p2p networks resulting in a reduction of the download time. Finally, the third level is an overlay based on common interest (the nodes form clusters based on common interests). This level is created by applying a learning algorithm called *Ranking Algorithm* and permits the reduction of the bandwidth used in the search procedure in the unstructured p2p networks. Therefore, Bittella improves the traditional unstructured p2p networks like Gnutella in two main aspects: it reduces the download time and it reduces the bandwidth consumption on the search procedure.**

I. Introduction

Peer-to-Peer (p2p) systems have become one of the most successful technologies in the Internet during the last years supported by file-sharing applications like Gnutella [5], Bittorrent [1] or Kademlia [6].

This paper focuses on fully distributed unstructured p2p systems as Gnutella. It proposes a new protocol which permits the utilization of Bittorrent download strategies (i.e. chunks exchange procedure) within fully distributed unstructured p2p systems. Therefore, it offers the download capacity available in Bittorrent (removing the Trackers) and the robustness of fully distributed unstructured p2p system, as Gnutella, in dynamic environment. This protocol is called *Bittella* (Bittorrent & Gnutella). In order to apply the Bittorrent download technique on unstructured p2p networks as Gnutella, Bittella is based on the Small-World per Content principle which permits the formation of clusters of peers that are sharing the same content. Moreover, Bittella takes advantage of the Small-World structure and defines a simple learning procedure, *The Ranking Algorithm*, that allows the nodes to store local information for future searches. By doing so, the amount of traffic generated in the search process is reduced from traditional unstructured p2p networks (e.g. Gnutella).

Finally, the paper evaluates Bittella by means of simulation in terms of download time and traffic generated during the search procedure and compares it with Gnutella.

The structure of this document is as follows. Section II introduces the new protocol, Bittella. Section III compares the performance of Bittella and Gnutella. Finally, Section IV provides some conclusions and introduces further work.

II. Bittella

II-A. *Three Layer Overlay Structure*

Bittella is a Three Layer Overlay structure. In the lowest layer we find the traditional unstructured p2p scheme (e.g. Gnutella). The medium layer is formed by one cluster per each content being shared. Therefore, this layer is a Small World per Content Structure (also named Content Small World Structure), and is in this level where the Bittorrent-like download technique is applied. Finally, the highest layer is formed by clusters of common interests, that is, each cluster is formed by peers with common interests. Therefore, it is a Common Interest Small World structure. These common interest clusters are formed by applying the Ranking Algorithm locally on each peer. It will be detailed below in Section II-C.

II-B. *Bittella Basic Functionality*

There is at least one Seed (node with the complete file) per each content in the p2p network. Each file is divided in pieces, called chunks, which are shared among the nodes as in Bittorrent.

One node (e.g. *Node A*) looking for the *Content A* searches in its local information in order to discover any node sharing it. If it has not local information about Content A, it launches a search procedure based on flooding (as it is done in Gnutella). When one node sharing Content A is discovered, the Node A obtains from it the information about the chunks ands Seeds of Content A and a list of peers sharing this content. At this point Node A starts to obtain an exchange chunks with the others peers. In addition, Node A obtains from other peers information about other contents which is stored as local information.

By the application of this protocol all the nodes interested in Content A form a cluster in order to share it. This cluster is called *Content A Small World*. The different Content Small Worlds are joined by the neighbours' connections on the Underlying p2p overlay (e.g. Gnutella) and for the presence of one node in different Content Small Worlds (i.e. one node sharing different contents).

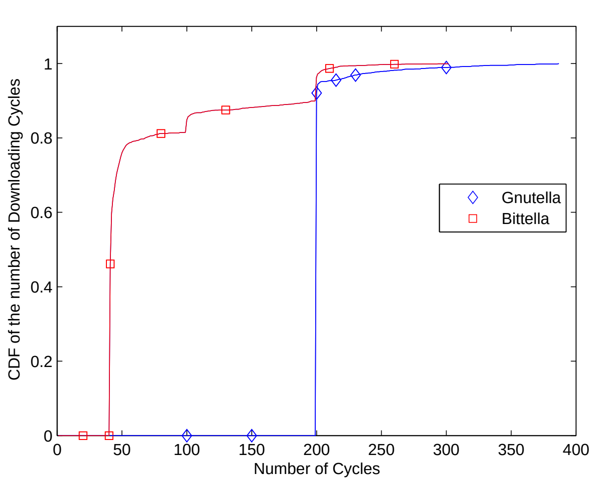

Fig. 1. CDF of the download time: Bittella vs. Gnutella

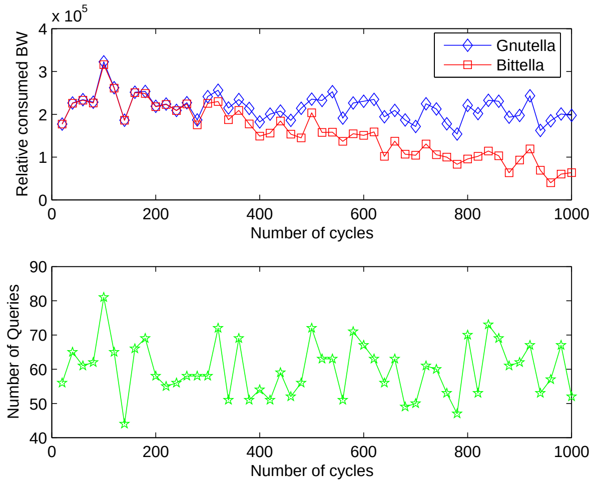

Fig. 2. BW consumption: Gnutella vs. Bittella & Generated Queries

II-C. Management of Local Information: Ranking Algorithm

In Bittella, every peer receives information from many others peers present within the same Content Small-Worlds. It can be potentially a huge amount of information, therefore the peer must discern which information is relevant. This is, the information related to contents which are more likely to be required in the future. Hence, in Bittella each peer elaborates a local Ranking of other Peers. In this ranking, each other peer is ranked by the number of Content Small World where it has been met until the current instant. Then, the top 1 is the one met in more Content Small Worlds. Intuitively, peers with common interests will meet in many Content Small Worlds. Therefore, the Ranking Algorithm produces a Small-World structure based on Common Interests. This structure has been previously analyzed in the literature [4] [2] [3] and all its demonstrated advantages can be considered Bittella advantages as well.

III. Performance Evaluation

This section presents the simulations which demonstrate the effectiveness of the application of the Small World per Content principle in terms of Bandwidth Consumption in the search procedure and Download Rate. The simulations consider only the two lower layer of the Three Layer Overlay structure formed by Bittella.

III-A. Simulation Parameters

The simulation applies the following parameters: N (the number of nodes forming the Gnutella network) equal to 1000; n (the number of neighbours that each node have in the Gnutella network) equal to 5; R (the number of content offered in the p2p system) equal to 62; C (the number of chunks per content) equal to 100; Q (the number of queries -i.e. content solicitation- generated during the simulation) equal to 3000; S (the number of initial Seeds per content) equal to 2; TTL (the radius of the flooding queries) equal to 5; BW (It is the number of chunks which can be simultaneously downloaded or uploaded) equal to 4.

III-B. Simulation Experiments and Analysis of the Results

The experiments was deployed in a static situation, where nodes do not join or leave the network.

The first metric analyzed, is the CDF (Cumulative Distribution Function) of the number of simulation cycles spent in the download process. The results are presented in figure 1. The figure shows that in a time equal to 50 cycles the 80 % of the files had been downloaded using Bittella whereas Gnutella downloads took never less than 200 cycles. Therefore, in the 80 % of the download processes Bittella reduces the time spent in a factor 4. Moreover, in the 100 % of the cases it improves the download time of Gnutella.

The second experiment focuses on evaluating the reduction of bandwidth generation during the searching procedure obtained by Bittella in front of Gnutella. In this experiment, we measure the total number of queries generated every 20 simulation cycles. That is, the original queries but also the replication of them forwarded due to the flooding algorithm. Figure 2 shows the results of the experiment. The upper graphic in figure 2 shows the Bandwidth Consumption of Gnutella and Bittella in terms of relative Bandwidth Units. We can see how at the beginning of the simulation, Bittella and Gnutella present the same Bandwidth Consumption, but according with the simulation advance Gnutella maintains the same Bandwidth Consumption, around $2{,}5 * 10^5$ bandwidth units, whereas Bittella reduces it. At the half of the simulation (around the cycle 500) Bittella offers a Bandwidth Consumption around $1{,}75 * 10^5$ that represent a 30 % of reduction compared with Gnutella. Even more, at the end of the simulation Bittella presents a Bandwidth Consumption below $1 * 10^5$ bandwidth units which means a reduction of 150 % (1.5 times) compared with Gnutella. This reduction in the Bandwidth Consumption is produced by the learning procedure.

IV. Conclusion and Further Work

The paper defines Bittella, a new protocol for unstructured p2p networks. Bittella allows the reduction of the download time due to the use of Bittorrent-like file sharing techniques. In addition it defines a simple learning algorithm, the Ranking Algorithm, which permits to form a Small-World structure based on common interests which has been previously proposed in the literature. The advantage offered by the Ranking Algorithm in front of other proposed solution is its simplicity. The application of the Ranking Algorithm produces the reduction of the amount of traffic generated by the search procedure in unstructured p2p networks like Gnutella. On the other hand, from the structural point of view Bittella defines a Three Layer Overlay network.

In summary, Bittella offers large improvements on Gnutella both in the download time which is reduced by a factor 4 in the 80 % of the cases and in the amount of traffic needed for the search procedure which is reduced up to 150 % compared with Gnutella. In addition, the implementation of Bittella is very simple.

The future work will focus on the simulation of the Three Layer Bittella Overlay model. Moreover, it will be demonstrated that the Three Layer Overlay fulfils the Small-World structure requirements. Finally, the simulation will be extended to consider a dynamic scenario where nodes join and leave the network randomly.

References

[1] Bittorrent. *Web Site-www. bittorrent. org*.
[2] E. Cohen, A. Fiat, and H. Kaplan. Associative search in peer to peer networks: harnessing latent semantics. *Proceeedings of INFOCOM 2003*, 2003.
[3] K. Hui, J. Lui, and D. Yau. Small world overlay P2P networks. *Proceedings of IWQOS 2004*, 2004.
[4] A. Iamnitchi and I. Foster. On Exploiting Small-World Usage Patterns in File-Sharing Communities.
[5] T. Klingberg and R. Manfredi. Gnutella 0.6. *Network Working Group, June*, 2002.
[6] P. Maymounkov and D. Mazieres. Kademlia: A peer-to-peer information system based on the XOR metric. *Proceedings of IPTPS' 02*, 2002.