Self-organizing Contents

Agostino Forestiero
Institute for High Performance Computing and Networking, ICAR-CNR,
Via Pietro Bucci, 41C 87036 Rende (CS), Italy

Keywords: Content Delivery Networks, Self-organizing Agents.

Abstract: To delivery static contents, Content Delivery Networks (CDNs) are an effective solution, but that shows its limits in dynamic and large systems with centralized approach. Decentralized algorithms and protocols can be usefully employed to tackle this weakness. A biologically inspired algorithm to organize the contents in a Content Delivery Networks, is proposed in this paper. Mobile and bio-inspired agents move and logically reorganize the metadata that describe the contents to improve discovery operation. Experimental results confirm the efficacy of the self-organizing and decentralized algorithm.

1 INTRODUCTION

Content Delivery Networks are an efficient alternative to deliver static contents, as video on-demand, TV broadcasts, media streaming services, pay-per-use software or pay-per-download music, etc, to the centralized approach. A group of users, can select, simultaneously watch all and share the control of a multimedia session, thanks to the collaborative playback services provided by such networks. The performance, in terms of response time, accessibility and bandwidth, of Internet-based content delivery by coordinated content replication, can be improved exploiting CDNs. A group of clusters of surrogate servers, located at the network edge, are maintained and geographically distributed. Nowadays, many aspects of Content Networks get better in the available content, the number of hosts and servers, the kind and the number of the users and the number of services where real time and bandwidth that can be exploited. The best surrogate servers - that store copies of the content - can satisfy the user requests. Hence, a system and a mechanisms to provide contents and services in a scalable manner, should be offered. With the explosion of social networks and P2P technologies, the content is incredibly increased, such as the exploitation of the Cloud Computing paradigm, in which numerous server hosts located in the ”Clouds” manage the contents and the services. However, to perform retrieval or access operations, current applications that create, modify and manage the content, and actively place it at appropriate locations, are often insufficient. Decentralized algorithms and protocols, such as peer-to-peer (P2P) and multi agent systems, can be useful to deal with new technologies and complex paradigms (Fortino and Mastroianni, 2009). Small- or medium-sized networks can be acceptably tackled with a centralized approach owing to its poor scalability. But, the CDN paradigm shows its limits in large and dynamic systems as the big-sized networks.

In this paper an algorithm that exploits nature-inspired agents to organize the contents in Content Delivery Networks, is presented. The metadata that describe the content are moved and logically organized by the agents to improve information retrieval operations. The relocation of the metadata is driven by behavior of agents. The metadata are indexed by binary strings obtained as result of the application of a locality preserving hash function. The locality preserving hash function is tailored to map similar resources into similar metadata. It can have several different meanings, for example, each bit can represent the absence or presence of a given topic. To guarantee that similar resources are associated to similar binary strings, an hash function locality preserving must be used. Agents move across the network through the peer to peer interconnections moving the metadata. The similar metadata, representing similar resources, are located into the same or in a neighbor host/server. The host and the metadata in not hardly associated, but easily adapts to the different conditions of the network.

The logical reorganization performed thanks to the self organizing and adaptive behavior of the pro-

Forestiero A..
Self-organizing Contents.
DOI: 10.5220/0004972405340538
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 534-538
ISBN: 978-989-758-027-7

posed algorithm allows to exploit the benefits of structured and unstructured approach of the peer to peer systems. The approach here analyzed is basically unstructured and then it is easy to maintain in a dynamic environment where the departures and arrivals of hosts can be frequent events. The logical reorganization of the metadata can improve the rapidity and effectiveness of discovery operations, and moreover, it is possible perform range queries, which is typical feature of structured peer to peer systems. In fact, thanks to the features of hash function, the metadata with few different bits will be located in neighboring regions. To measure the similarity between two metadata the Hamming distance or the cosine of the angle between the related vectors can be used. In section 3 a brief description of the nature-inspired algorithm is given, but further details can be found on (Forestiero and Mastroianni, 2009), in which a similar algorithm was exploited for building a Grid Information System.

2 RELATED WORKS

Hybrid approach between CDN and P2P was analyzed by several studies, but while (Kang and Yin, 2010) (Huang et al., 2008) proposed further steps into the use of P2P to deliver multimedia content, (Mulerikkal and Khalil, 2007) (Guomin et al., 2006) exploit the P2P overlay for surrogate cooperation while leaving the clients are regular non-cooperative entities. (Xu et al., 2006) a collaboration between clients is proposed, but clients cannot receive data from different sources, such as from the peering community and CDN entities at the same time. The dynamic nature of the today's networks and the large variety of the resources make the management and discovery operations more troublesome. Administrative bottlenecks and low scalability of centralized systems are becoming unbearable. Innovative approaches need to have properties as self-organization, decentralization and adaptivity. Erdil et al. in (Erdil et al., 2006) outline the requirements and properties of self organizing grids. Reorganization of resources to facilitate discovery operations and adaptive dissemination of information, were introduced and applied in the approach here presented. A class of agent systems which aims to solve very complex problems by imitating the behavior of some in species of ants as introduced in (Bonabeau et al., 1999). In (Forestiero et al., 2008b) and (Forestiero et al., 2008a), the performance of discovery operations are improved through the creation of Grid regions specialized in a particular class of resources. Whereas (Van Dyke Parunak et al., 2005) proposes a decentralized scheme to tune the activity of a single agent. These systems are positioned along a research avenue whose objective is to devise possible applications of ant algorithms (Bonabeau et al., 1999) (Dorigo et al., 2000). A tree-based ant colony algorithm to support large-scale Internet-based live video streaming broadcast in CDNs, was proposed in (Liu et al., 2012). In this paper, differently from the traditional solution to find paths, an algorithm to optimize the multicast tree directly and integrate them into a multicast tree, was introduced.

3 SELF-ORGANIZING ALGORITHM

The work of the nature inspired agents is profitably exploited to logically reorganize the metadata. Agents move among hosts performing simple operations. When an agent arrives to an host and it does'not carry any metadata, it decides whether or not to pick one or more metadata from the current host. While when arrives to an host and the agent is loaded, it decides whether or not to leave one or more metadata in the local host. A couple of probability functions drive agent's decision. The probability functions are based on a similarity function (Lumer and Faieta, 1994), that is:

$$sim(\bar{m}, Reg) = \frac{1}{Num_m} \sum_{m \varepsilon Reg} 1 - \frac{Hamming(m, \bar{m})}{\alpha} \tag{1}$$

The similarity of a metadata m with all the metadata located in the region Reg, is measured through the function sim. The region Reg for each host h, is represented by h and of all host reachable from h with a given number of hops. Here it is set to 1. Num_m is the overall number of metadata located in Reg, while $Hamming(m, \bar{m})$ is the Hamming distance between m and $\bar{m}$. The similarity scale α is set to 2. The value of sim ranges between -1 and 1, but negative values are septated to 0. The probability function of picking a metadata from an host will be inversely proportional to the similarity function sim. Vice versa, the probability function of dropping a metadata will be directly proportional to the similarity function sim. The probability functions of picking a metadata P_1 and the probability function of leaving a metadata P_2, are:

$$P_1 = (\frac{k1}{k1 + sim(\bar{m}, Reg)})^2; \tag{2}$$

$$P_2 = (\frac{sim(\bar{m}, Reg)}{k2 + sim(\bar{m}, Reg)})^2 \tag{3}$$

The degree of similarity among metadata can be refined through the parameter *k1* and *k2*, that have values comprised between 0 and 1 (Bonabeau et al., 1999). The flowchart showed in Figure 1 gives an high-level description of the algorithm performed by mobile agents. Cyclically, the agents perform a given number of hops among servers, and when they get to a server, they decide which probability function compute based on their load. If the agent does not carries metadata it computes P_1, otherwise if the agent carries metadata the P_2 probability is computed.

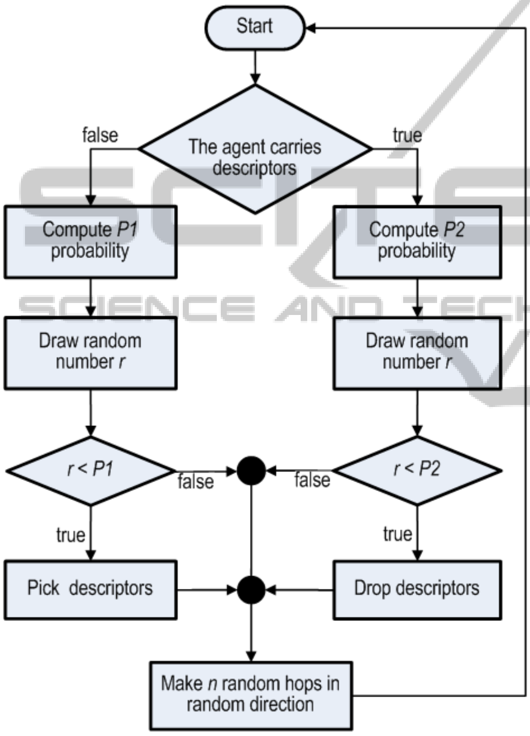

Figure 1: The algorithm performed by the agents.

The processing load *Load*, that is the average number of agents per second that are processed by a server, does not depend neither on the network size nor on the churn rate, but only depends on the number of agents and the frequency of their movements across the network. *Load* can be obtained as:

$$Load = \frac{n.of agent}{n.of server \cdot T_{move}} = \frac{F_{gen}}{T_{move}} \qquad (4)$$

where T_{move} is the average time between two successive movement of an agent and F_{gen} is the mean number of agents generated by a single server. In a typical scenario, T_{move} is set to 60 sec and F_{gen} is set to 0.5, so that each server receives and processes about one agent every 120 seconds, which can be considered an acceptable load. Note that the processing load does not depend on other system parameters such as

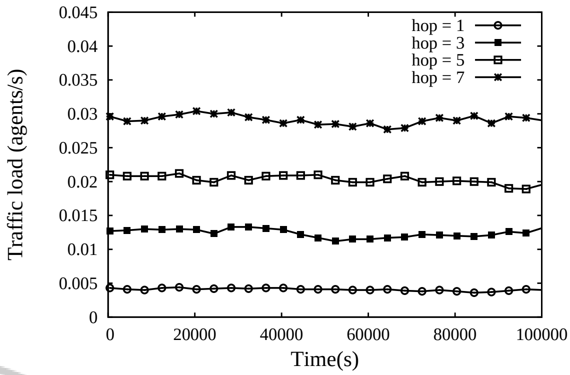

Figure 2: The traffic generated by the algorithm when the number of hops performed within a single agent movement ranges from 1 to 7.

the average number of resources handled by a server or the number of server, which is a further confirmation of the scalability properties of the algorithm.

The traffic load, that is the number of agents that go through a server per unit time, was calculated and shown in Figure 2. As showed, the value of traffic load changes according to the maximum number of hops performed within a single agent movement. During the simulations, it was noted that the reorganization of metadata is accelerated if agent movements are longer, because they can explore the network more quickly. To choice of the number of hops is necessary a compromise between the level of the traffic load tolerable and the rapidity and efficiency of the reorganization process.

The effectiveness of the algorithm has been evaluated exploiting the spatial homogeneity function. By averaging over the whole network the Hamming distance between every couple of metadata, we obtained the average homogeneity **Homogeneity** of the metadata in the region *Reg* related to the host *h*, this is:

$$AVG = Average_{m,\bar{m} \varepsilon Reg}(Hamming(m,\bar{m}))$$

$$Homogeneity_h = n.of bits - AVG \qquad (5)$$

When the similar metadata are collected in the same region, the value of homogeneity function increases. A graphical description of the logical reorganization achieved by agents, in Figure 3 is reported. Here, each metadata is associated to a color. The network is photographed at *Time* = 0 sec, when the process is starting and the metadata are randomly distributed in a random fashion and at *Time* = 50.000 sec when the process is in a steady situation.

Notice that similar metadata are located in the same region and between near region the color change

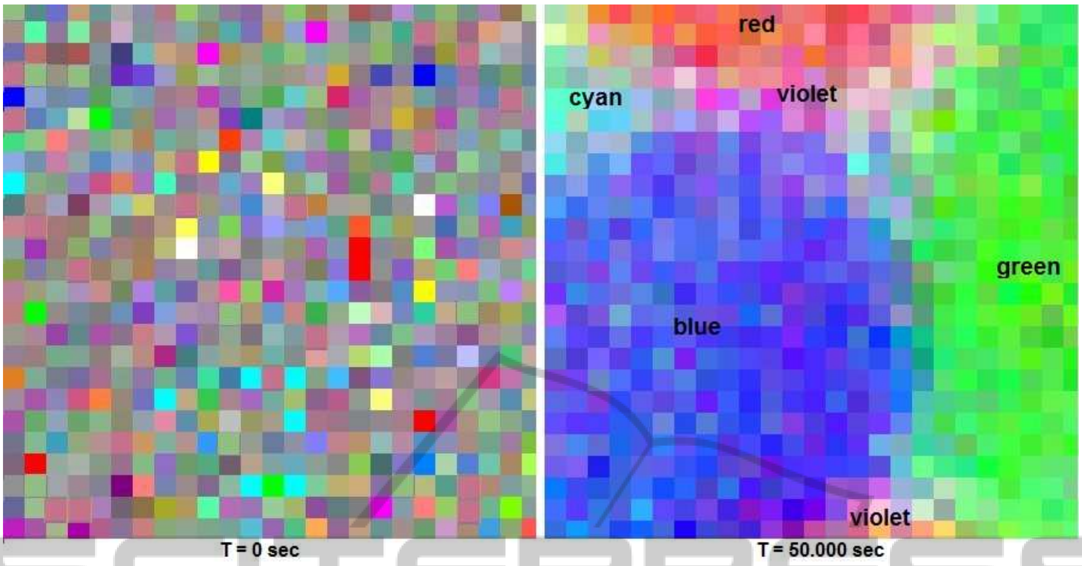

Figure 3: Snapshots of the system when the process is starting, and when the process is in a steady situation.

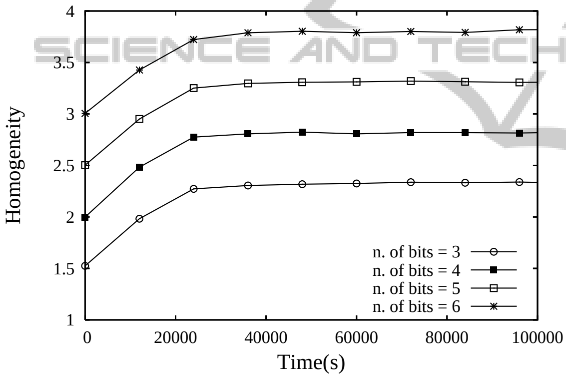

Figure 4: Homogeneity of the whole network when the number of bit of the binary string representing the content ranges from 3 to 6.

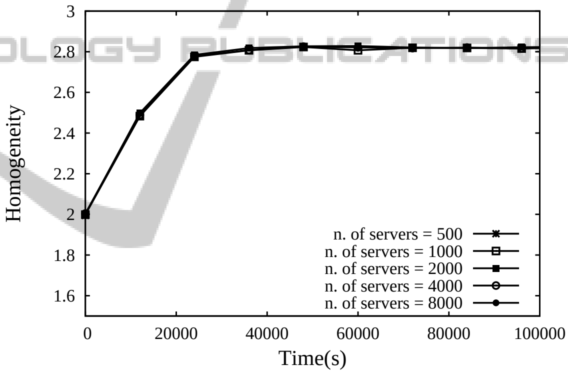

Figure 5: Homogeneity, vs. time, for different values of the number of servers.

gradually, which proves the spatial sorting on the network. In Figure 4 the overall homogeneity when the number of bit of the metadata describing the content, is varied. We can see as the logical reorganization is obtained independently of the number of bits. To confirm the scalability nature of the algorithm, which derives from its decentralized and self-organizing characteristics, its behavior when number of involved servers changes from 1000 to 7000, was analyzed and reported in Figure 5. Notice that the size of the network, that is the number of servers involved in the logical organization of contents, has no detectable effect on the performance, as overall homogeneity index.

4 CONCLUSIONS

In this paper a nature inspired approach to build a P2P information system for CDNs, was proposed. Thanks to its swarm intelligence characteristics, the proposed algorithm features fully decentralization, adaptivity and self-organization. Ant-inspired agents move and logically reorganize the metadata representing the contents. Agent's operations are driven by simple probability functions that are evaluated when the agent gets to a new server. In this way, similar metadata representing similar contents are placed in the same region, that is in neighbor server. Moreover, the reorganization of metadata spontaneously adapts to the ever changing environment as the joins and departs of servers and the changing of the characteristics of the contents. The experimental results proved the effectiveness of the algorithm.

REFERENCES

Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems*, volume 4. Oxford university press New York.

Dorigo, M., Bonabeau, E., and Theraulaz, G. (2000). Ant algorithms and stigmergy. *Future Generation Computer Systems*, 16(8):851–871.

Erdil, D. C., Lewis, M. J., and Abu-Ghazaleh, N. B. (2006). Adaptive approach to information dissemination in self-organizing grids. In *Autonomic and Autonomous Systems, 2006. ICAS'06. 2006 International Conference on*, pages 55–55. IEEE.

Forestiero, A. and Mastroianni, C. (2009). A swarm algorithm for a self-structured p2p information system. *Evolutionary Computation, IEEE Transactions on*, 13(4):681–694.

Forestiero, A., Mastroianni, C., and Spezzano, G. (2008a). Reorganization and discovery of grid information with epidemic tuning. *Future Generation Computer Systems*, 24(8):788–797.

Forestiero, A., Mastroianni, C., and Spezzano, G. (2008b). So-grid: A self-organizing grid featuring bio-inspired algorithms. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 3(2):5.

Fortino, G. and Mastroianni, C. (2009). Next generation content networks. *Journal on Network and Computing Applications*, 32(5):941–942.

Guomin, Z., Changyou, X., and Ming, C. (2006). A distributed multimedia cdn model with p2p architecture. In *Communications and Information Technologies, 2006. ISCIT'06. International Symposium on*, pages 152–156. IEEE.

Huang, C., Wang, A., Li, J., and Ross, K. W. (2008). Understanding hybrid cdn-p2p: why limelight needs its own red swoosh. In *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, pages 75–80. ACM.

Kang, S. and Yin, H. (2010). A hybrid cdn-p2p system for video-on-demand. In *Future Networks, 2010. ICFN'10. Second International Conference on*, pages 309–313. IEEE.

Liu, G., Wang, H., and Zhang, H. (2012). An ant colony optimization algorithm for overlay backbone multicast routing in content delivery networks. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, pages 1878–1882. IEEE.

Lumer, E. D. and Faieta, B. (1994). Diversity and adaptation in populations of clustering ants. In *Proceedings of the Third International Conference on Simulation of Adaptive Behavior : From Animals to Animats 3: From Animals to Animats 3*, SAB94, pages 501–508. MIT Press.

Mulerikkal, J. P. and Khalil, I. (2007). An architecture for distributed content delivery network. In *Networks, 2007. ICON 2007. 15th IEEE International Conference on*, pages 359–364. IEEE.

Van Dyke Parunak, H., Brueckner, S. A., Matthews, R., and Sauter, J. (2005). Pheromone learning for self-organizing agents. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35(3):316–326.

Xu, D., Kulkarni, S. S., Rosenberg, C., and Chai, H.-K. (2006). Analysis of a cdn–p2p hybrid architecture for cost-effective streaming media distribution. *Multimedia Systems*, 11(4):383–399.